ROBUST ESTIMATION AND HYPOTHESIS TESTING

Moti L. Tiku
Aysen D. Akkaya

NEW AGE

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS

ROBUST ESTIMATION
AND
HYPOTHESIS TESTING

This page intentionally left blank

ROBUST ESTIMATION AND HYPOTHESIS TESTING

Moti L. Tiku
Department of Mathematics and Statistics
McMaster University, ON L9H 4A6
CANADA
and
Department of Statistics
METU, Ankara 06531
TURKEY

Aysen D. Akkaya
Department of Statistics
METU, Ankara 06531
TURKEY

NEW AGE

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
New Delhi • Bangalore • Chennai • Cochin • Guwahati • Hyderabad
Jalandhar • Kolkata • Lucknow • Mumbai • Ranchi

Published by New Age International (P) Ltd., Publishers

ISBN (13) : 978-81-224-2537-6

PUBLISHING FOR ONE WORLD

NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS
4835/24, Ansari Road, Daryaganj, New Delhi - 110002
Visit us at **www.newagepublishers.com**

To – –

Our Children

and

Grand Children

DIYA, IDIL, NISHA, VITASTA, ILGE and ISHAN

This page intentionally left blank

PREFACE

Most classical procedures are essentially based on two assumptions, that the sample observations are independently and identically distributed, and that the underlying distribution is normal. While the first assumption is not unreasonable in most situations, it is the second assumption which is rather unrealistic from a practical point of view. To quote **R. C. Geary (*Biometrika*, 1947)**: " Normality is a myth, there never was, and never will be a normal distribution". This view is corroborated by the findings of **E. S. Pearson (*Biometrika*, 1932)** and **Elveback et al. (*J. American Medical Assoc.*, 1970)** among others. Geary might have overstated the case, but the fact is that non-normal distributions are more prevalent in practice, and to assume normality instead, might lead to erroneous statistical inferences. On the other hand, to assume that nothing is known about the underlying distribution is to deny the information that Q-Q plots and goodness-of-fit tests can provide about the nature of the underlying distribution. It is true that for small to moderate sample sizes, these procedures are not very successful in discriminating between normal and moderately non-normal distributions, but they are quite successful in distinguishing normal from extremely non-normal distributions both symmetric and skew (Cauchy and exponential, for example). For a sample of size 20, for example, the **Shapiro-Wilk** omnibus goodness-of-fit test of normality has, at 10 percent significance level, power values of 0.99 and 0.92 against exponential and Cauchy distributions, respectively; against a logistic distribution, however, this test and other tests of normality have a power value as low as 0.16 for sample size 20 as given in Table I of **Tiku (*Commun. Stat.*, 1974)**. Thus in situations where one finds it expedient to assume normality, what should perhaps be assumed instead is that the underlying distribution is one of a reasonably wide class of distributions consisting of those distributions which constitute plausible alternatives to normality, e.g., symmetric distributions and moderately skew distributions, all having finite mean and finite variance; which one exactly is not known. The normal distribution is, of course, a member of this class, but extremely non-normal symmetric and skew distributions need not be included, since they are distinguishable from a normal and can be considered separately. Of course, whatever has been said above about a normal distribution can be said for any other assumed distribution. What is needed, therefore, is a method of estimation which yields robust estimators. An estimator is said to be robust if it is fully efficient (or nearly so) for an assumed distribution but maintains high efficiency for plausible alternatives. Moreover, the estimator needs to be robust to outliers, inliers and contaminations. This is of enormous importance since, in practice, samples contain discrepant observations so often. We also need robust hypothesis testing procedures. A hypothesis testing procedures is said to have criterion robustness if its Type I error is never substantially higher than a presumed level for any plausible alternative. It is said to have efficiency robustness if its power is high for all plausible alternatives. We also require the procedure to have criterion and efficiency robustness if the sample contains a small number of outliers or contaminated observations (usually not more than 10 percent). We present in this book the methodology of modified likelihood estimation based on complete and censored samples. The former (complete samples) is based on numerous publications which have appeared since 1985. The latter (censored samples) is presented in detail in the book entitled

Robust Inference by **Tiku, Tan and Balakrishnan (1986)** and covers publications prior to 1985. The main purpose of this monograph is to focus on complete samples (avoiding unnecessary censoring of observations) and develop robust procedures in numerous areas, some of them not covered in **Huber (1981)**, and **Tiku, Tan and Balakrishnan (1986)**, and other such books. The method of modified likelihood estimation adopted here gives MML (modified maximum likelihood) estimators which have three very desirable properties: (i) asymptotically, they are BAN (best asymptotically normal), (ii) for small samples, they are almost fully efficient, i.e., they have no or negligible bias and their variances are only marginally bigger than the minimum variance bounds, and (iii) they are explicit functions of sample observations and are, therefore, easy to compute. Moreover, the MML estimates are numerically very close to the **Fisher** ML (maximum likelihood) estimates for all sample sizes. Being explicit functions of sample observations, they are also amenable to analytical studies **(Vaughan, 2002).** A remarkable property of the MML estimator of a scale parameter is that it is always real and positive while the ML estimator does not necessarily have these important properties. Several such examples are given in the book. The MML estimators are also shown to be robust. The hypothesis testing procedures based on them are shown to have both criterion robustness as well as efficiency robustness. Comparisons with **Huber (1981)** and other available robust procedures (based on censored samples) are given. The methodology presented here is shown to have definite theoretical and computational advantages besides providing efficient and robust estimators and hypothesis testing procedures.

In Chapter 1, the robustness of the classical Student's t (and t^2) and ANOVA tests are studied. It is shown that they have criterion robustness but not efficiency robustness. This is primarily due to the fact that the sample mean is normally distributed at any rate for large sample sizes n (by central limit theorem) but is an inefficient estimator of the population mean, unless the underlying distribution is normal or close to it. For testing the equality of two population variances, however, the classical F test is shown not to have even the criterion robustness. **In Chapter 2**, estimation of the location and scale parameters is considered. The method of maximum likelihood is shown to be ideal for certain situations but problematic for others. In general, the likelihood equations are shown not to have explicit solutions. As a result, they have to be solved by iteration which can be problematic for reasons of **(i) multiple roots, (ii) slow convergence, or (iii) convergence to wrong values**. The method of modified likelihood estimation is formulated. This method gives estimators which are explicit functions of sample observations and, hence, they are easy to compute. Besides, these estimators are shown to be fully efficient asymptotically and highly efficient for small n. **In Chapter 3**, linear regression models are considered with errors having normal and non-normal distributions. Three families of distributions are considered, e.g., the Weibull, the Student's t, and the family of short-tailed symmetric distributions recently introduced by **Tiku and Vaughan (1999)**. The MML (modified maximum likelihood) estimators are derived and shown to be considerably more efficient than the commonly used LS (least squares) estimators. In fact, the LS estimators are shown to have a disconcerting feature, namely, their efficiencies relative to the MML estimators decrease as the sample size n increases. **In Chapter 4**, the important topic of binary regression is considered. This has many medical and biological applications. Logistic and non-logistic density functions are considered and the effects of risk factors evaluated. **In Chapter 5**, autoregressive models are considered. They have applications in agricultural, biological and biomedical sciences besides business and economics. Both normal and non-normal error distributions are considered. The supposedly difficult problems of estimation and hypothesis testing are considered and efficient and robust solutions given. **In Chapter 6**, fixed-effects models in experimental design are considered. Solutions are developed for both normal and non-normal error distributions. The methodology is applied to the well known **Box-Cox** biometrical data and shown to give accurate results besides being easy theoretically and computationally. The method is extended to non-identical error distributions. **In Chapter 7**, estimation and hypothesis testing based on censored samples is considered. Efficiency and distributional properties are enunciated and a few important real life examples given. **In Chapter 8**, the very important issue of robustness is considered. The estimators and hypothesis testing procedures developed in **Chapters 2-7**, are shown to be remarkably efficient and robust. They are compared to **Huber** and several other procedures (based on censored samples), and shown to have definite advantages. In particular, they are applicable to both symmetric (short-and long-tailed) as well as skew distributions, whereas, **Huber** and other procedures (based on censored samples) are applicable to only long-tailed symmetric distributions**. In Chapter 9**, Q-Q plots, goodness-of-fit tests and outlier detection procedures are given. They are shown to be very useful in locating a plausible model for a given data set. This enhances the scope of application of robust procedures. **In Chapters 10,** usefulness of the MMLE in the important area of sample survey is discussed. The estimators are shown to be enormously more efficient than the sample mean under both super-population as well as finite population models. **In Chapter 11**, numerous real life applications of the procedures are given and discussed, and appropriate statistical analyses performed.

This book is intended as a **reference for theoreticians and practitioners of statistics** and can be used as an **auxiliary text for a senior-level course on robustness and statistical inference**. This book is also intended to illustrate the appropriateness of various robust methods based on complete samples (avoiding unnecessary censoring of observations) for solving numerous statistical inference problems in a number of important areas. This monograph has provided the authors an opportunity to consolidate the results given in numerous papers published since **1985**. A basic knowledge of sampling distributions, statistical inference, and order statistics is assumed. The book contains numerous examples which illustrate the usefulness of robust methods based on complete as well as censored samples. For readers interested in research on modified likelihood estimation, robust estimation and hypothesis testing, fairly extensive references to recent works are given.

Moti L. Tiku

Ayşen D. Akkaya

This page intentionally left blank

CONTENTS

This page intentionally left blank

CHAPTER 1

Robustness of Some Classical Estimators and Tests

1.1 INTRODUCTION

Statistical methods have been used in almost every applied field to analyse experimental data. But, unfortunately, misuses of statistical methods are very common. One of the major reasons for such misuses is that many practitioners do not realize that statistical methods are derived under certain assumptions; yet, in practice, many of these assumptions do not hold. For example, many of the statistical procedures, such as the classical t test and the variance-ratio tests, are derived under the assumption of normality. However, as noted by Geary (1947) and Scheffé (1959), the assumption of normality very often is not a realistic one. A crucial question then is: what are the effects of non-normality on the normal-theory estimators and hypothesis testing procedures. An estimator is called robust if it is fully efficient (or nearly so) for an assumed distribution but maintains high efficiency for plausible alternatives. A fully efficient estimator is one which is unbiased and its variance is equal to the Cramér-Rao minimum variance bound.

A statistical test has traditionally been called robust if both its type I error (significance level) and its power are not affected much by departures from normality. Box and Tiao (1964a,b) have, in fact, advocated that when the underlying distribution can not be assumed to be exactly normal, a statistical procedure should be derived under a reasonably wide class of distributions. This will undoubtedly enhance the scope of application of statistical procedures. It is, therefore, advantageous and desirable to develop tests that have (i) criterion robustness and (ii) efficiency robustness, for a range of plausible alternatives to a particular distribution (normal, for example). The purpose of this book is to illustrate that robust estimators and hypothesis testing procedures can be obtained in most situations from applications of modified likelihood methodology (Tiku, 1967a; 1968a, b; 1973; Tiku and Suresh, 1992; Vaughan, 1992a). To set the groundwork, we first study in this chapter the robustness of the sample mean and sample variance, and tests based on them.

1.2 ROBUSTNESS OF SAMPLE MEAN AND VARIANCE

Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal population $N(\mu, \sigma^2)$. The likelihood function, i.e. the joint pdf (probability density function) of the n random observations, is

$$L \propto \left(\frac{1}{\sigma}\right)^n e^{-\sum_{i=1}^{n}(y_i-\mu)^2/2\sigma^2} \qquad (1.2.1)$$

The likelihood equations for estimating μ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) = 0 \tag{1.2.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{n}{\sigma^3}\sum_{i=1}^{n}(y_i - \mu)^2 = 0. \tag{1.2.3}$$

The solutions of these equations are the ML estimators $\hat{\hat{\mu}} = \bar{y} = \sum_{i=1}^{n} y_i/n$ and $\hat{\hat{\sigma}} = \sqrt{\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2/n\right\}}$ of μ and σ, respectively. In the latter, one may replace the divisor $\sqrt{n}$ by $\sqrt{n-1}$ and write $s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$; s^2 is the sample variance and is an unbiased estimator of σ^2.

The Cramér-Rao MVB (minimum variance bound) for estimating μ is $1/\{-E(\partial^2 \ln L/\partial\mu^2)\} = \sigma^2/n$. Now, $E(\bar{y}) = \mu$ and $V(\bar{y}) = \sigma^2/n$. Thus, $\bar{y}$ is fully efficient. Moreover, $\bar{y}$ is normally distributed. In other words, $\bar{y}$ is an ideal estimator of μ under the assumption of normality.

To illustrate the robustness (or lack of it) of $\bar{y}$, we consider a range of distributions which supposedly constitute plausible alternatives (confining to long-tailed symmetric distributions) to the assumed normal distribution, namely,

$$f(y) \propto \frac{1}{\sigma}\left\{1 + \frac{(y-\mu)^2}{k\sigma^2}\right\}^{-p}, \; -\infty < y < \infty; \tag{1.2.4}$$

$k = 2p - 3$ and $p \geq 2$; $E(y) = \mu$ and $V(y) = \sigma^2$. It may be noted that $t = \sqrt{(\nu/k)}\,(y - \mu)/\sigma$ has the Student t distribution with $\nu = 2p - 1$ degrees of freedom. Another plausible alternative is that the sample contains a small number (usually, not more than 10 percent) of outliers.

Alongside the sample mean $\bar{y}$, we consider Tukey's estimator (of location) based on a censored sample of size $n - 2r$,

$$\hat{\mu}_{Trim} = \left(\sum_{i=r+1}^{n-r} y_{(i)}\right)/(n - 2r) \tag{1.2.5}$$

where r is usually taken to be the integer value $r = [0.5 + 0.1\,n]$ (Tiku, 1980a; Dunnett, 1982); $y_{(1)} \leq y_{(2)} \leq \ldots \leq y_{(n)}$ are the order statistics obtained by arranging $y_1, y_2, \ldots, y_n$ in ascending order of magnitude. The estimator $\hat{\mu}_{Trim}$ is called 10% trimmed sample mean. For some basic properties of order statistics, see Appendix 1A.

For the family (1.2.4), the MVB for estimating μ is (Tiku, 1980a)

$$MVB(\mu) = 1\Big/\left\{-E\left(\frac{\partial^2 \ln L}{\partial \mu^2}\right)\right\} = \frac{(p+1)(p-3/2)}{np(p-1/2)}\sigma^2. \tag{1.2.6}$$

Given below in Table 1.1 are the exact values of the relative efficiencies

$$E_1 = 100\,\{MVB(\mu)/V(\bar{y})\} \text{ and } E_2 = 100\{MVB(\mu)/V(\hat{\mu}_{Trim})\} \tag{1.2.7}$$

of the two estimators. Like $\bar{y}$, $\hat{\mu}_{Trim}$ is an unbiased estimator of μ for the family (1.2.4); this follows from symmetry. Its variance is

$$V(\hat{\mu}_{Trim}) = (\beta' \Omega \beta)\sigma^2/(n - 2r)^2 \tag{1.2.8}$$

where $\beta' = (1, 1,\ldots, 1)$ is a vector with $n - 2r$ elements and $\Omega = (\sigma_{ij:n})$ is the $(n - 2r) \times (n - 2r)$ variance-covariance matrix of the standardized variates $z_{(i)} = (y_{(i)} - \mu)/\sigma$, $r + 1 \le i \le n - r$; $\sigma_{ij:n}$ are available for $n \le 20$ (Tiku and Kumra, 1981).

Table 1.1: Values of the MVB and relative efficiencies

			p						
			2	2.5	3.5	4.5	5*	10	∞
(n/σ^2) MVB (μ) :			0.500	0.700	0.857	0.917	0.933	0.984	1.000
All n	E_1	:	50	70	86	92	93	98	100
n = 10	E_2	:	83	91	96	97	97	97	95
20			88	94	98	98	98	97	95

* This distribution is almost identical with the Logistic (Tiku and Jones, 1971).

It can be seen that $\hat{\mu}_{Trim}$ is overall considerably more efficient than $\bar{y}$, and its variance is much closer to the MVB. Thus, $\hat{\mu}_{Trim}$ is a robust estimator of μ, the sample mean $\bar{y}$ is not.

Consider now Dixon's outlier model, namely, $(n - j)$ observations come from normal $N(\mu, \sigma^2)$ and j (we do not know which) come from $N(\mu, 9\sigma^2)$. For n = 20, the relative efficiency of the sample mean

$$E = 100\{V(\hat{\mu}_{Trim})/V(\bar{y})\}$$

is only 84 and 72 percent for j = 1 and 2, respectively. Realize that like $\bar{y}$, $\hat{\mu}_{Trim}$ is an unbiased estimator of μ for this outlier model. Another plausible alternative is Tiku's outlier model (Tiku, 1975c; 1977; Hawkins, 1977) which will be considered in later chapters.

We will introduce Huber (1964) M-estimators of μ in Chapter 2 and show that they are remarkably efficient and robust for long-tailed symmetric distributions but not for skew and short-tailed symmetric distributions.

In Chapter 7, we develop another estimator based on a censored normal sample. The estimator is denoted by $\hat{\mu}_c$ and called MML (modified maximum likelihood) estimator; $\hat{\mu}_c$ is also a robust estimator of μ and has a neat distribution (normal, for large n), and its variance is very easy to compute (variances and covariances of order statistics are not required); see also Appendix 2E (Chapter 2).

For the normal distribution $N(\mu, \sigma^2)$, the MVB for estimating σ is $\text{MVB}(\sigma) = 1/\{-E(\partial^2 \ln L/\partial\sigma^2)\} = \sigma^2/2n$. For large n, s is fully efficient. Realize that $\bar{y}$ and s are uncorrelated, in fact, independent. For non-normal distributions, the variance of s is

$$V(s) \cong \left(1 + \frac{1}{2}\lambda_4\right)\sigma^2/2n, \quad \lambda_4 = (\mu_4/\mu_2^2) - 3, \tag{1.2.9}$$

for large n. The result (1.2.9) follows from the fact that for large n (Roy and Tiku, 1962),

$$V(s^2) \cong 2\left(1 + \frac{1}{2}\lambda_4\right)\sigma^4/n, \tag{1.2.10}$$

and if g(x) is a regular function of x then

$$V\{g(X)\} \cong [g'(\mu)]^2 \sigma^2;\ \mu = E(X),\ V(X) = \sigma^2. \tag{1.2.11}$$

The equation (1.2.11) follows from a Taylor series expansion; see, for example, Meyer (1970, p. 139). It is seen that the variance of s (and s^2) heavily depends on the population kurtosis μ_4/μ_2^2 (= 3 for a normal distribution). For the family (1.2.4), λ_4 is positive and decreases with increasing p. It is, in fact, infinite for $p \leq 2.5$. Clearly, s is not a robust estimator.

We now discuss the robustness of hypothesis testing procedures for testing assumed values of the population location and scale parameters.

1.3 ASYMPTOTIC ROBUSTNESS OF THE t TEST

Let $y_1, y_2, \ldots, y_n$ be a random sample from a normal population $N(\mu, \sigma^2)$. To test an assumed value of μ, $H_0: \mu = 0$ (say) against $H_1: \mu > 0$, the t statistic is employed:

$$t = \sqrt{n}\,\bar{y}/s; \tag{1.3.1}$$

$\bar{y}$ and s^2 are the sample mean and variance. Large values of t lead to the rejection of H_0 in favour of H_1. Under the assumption of normality, the test is UMP (uniformly most powerful). The null distribution of t is the Student t with $\nu = n - 1$ d.f. For large n, the null distribution of t is normal since $\bar{y}$ is normal $N(\mu, \sigma^2/n)$ and s converges to σ as n tends to infinity. For assessing the effects of non-normality on the t test, we let the distribution of y be one of a reasonably wide class of distributions with mean μ and variance σ^2, and finite cumulants κ_r, $r \geq 3$ ($\kappa_2 = \sigma^2$). Since the variance of s (equation 1.2.9) tends to zero as n tends to infinity, s/σ converges in probability to 1 (Kendall and Stuart, 1969, p.286). Further, by the Central Limit Theorem, the distribution of $\sqrt{n}(\bar{y} - \mu)/\sigma$ converges to normal N(0, 1) as n tends to infinity. Thus, by Slutsky's theorem (see Rao, 1975, Chapter 2), the distribution of t converges to normal N(0, 1) as n tends to infinity. The asymptotic distribution of t is, therefore, normal N(0, 1) irrespective of the underlying distribution. Thus, the t test has criterion robustness (asymptotically).

The power function of the t test for testing H_0 against H_1 above is asymptotically

$$P[t \geq t_\alpha(\nu) \mid H_1] \cong P(Z \geq z_\alpha - \sqrt{\delta}),\ \nu = n - 1, \tag{1.3.2}$$

where $\delta = n(\mu/\sigma)^2$ is the noncentrality parameter; Z is normal N(0, 1) and z_α is its $100(1-\alpha)\%$ point. Thus, the asymptotic power function of the test is insensitive to the underlying distribution. In the traditional sense, therefore, the t test is robust. However, we will show that the test does not have efficiency robustness in terms of having high power, although it has criterion robustness at any rate for large n.

1.4 COMPARING SEVERAL MEANS

As an extension of the one-sample t test above, we now consider the problem of comparing c means ($c \geq 2$). Let

$$y_{i1}, y_{i2}, \ldots, y_{in_i}\quad (i = 1, 2, \ldots, c) \tag{1.4.1}$$

be independent random samples from c normal populations $N(\mu_i, \sigma^2)$. One wants to test the null hypothesis $H_0 = \mu_1 = \mu_2 = \ldots = \mu_c$ against the alternative $H_1 = \mu_i \neq \mu_j$ (for some i and j). The test is based on the F statistic (Lehmann, 1959, pp. 268, 313),

$$F = s_b^2/s_e^2,\ \bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i,\ \bar{y} = \sum_{i=1}^{c}\sum_{j=1}^{n_i} y_{ij}/N,\ N = \sum_{i=1}^{c} n_i \tag{1.4.2}$$

$(c-1)s_b^2 = S_b = \sum_{i=1}^{c} n_i(\bar{y}_i - \bar{y})^2$ is called "block" sum of squares, and $(N-c)s_e^2 = S_e = \sum_{i=1}^{c}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$ is called error or "within" sum of squares.

Large values of F lead to the rejection of H_0 in favour of H_1. The distribution of F under H_0 (null distribution) is central F with degrees of freedom $\nu_1 = c - 1$ and $\nu_2 = N - c$. The distribution of F under H_1 (non-null distribution) is noncentral F with degrees of freedom ν_1 and ν_2 and noncentrality parameter

$$\lambda = \sum_{i=1}^{c} n_i(\mu_i - \bar{\mu})^2/\sigma^2, \quad \bar{\mu} = \sum_{i=1}^{c} n_i\mu_i/N \tag{1.4.3}$$

For details about a noncentral F distribution, one may refer to Biometrika Tables (Pearson and Hartley, 1972). See also Tiku (1965, 1985 b).

Asymptotic robustness: Using exactly the same argument as in Section 1.3 above, it follows that asymptotically the null distribution of F is central chi-square with ν_1 degrees of freedom. The non-null distribution of F is noncentral chi-square with ν_1 degrees of freedom and noncentrality parameter λ. For details about a noncentral chi-square distribution, one may refer to Tiku (1985a). Since λ in (1.4.3) does in no way depend on the underlying distribution, and $\bar{y}_i$ (and $\bar{y}$) are asymptotically normal and s_e^2 converges to σ^2, the F test above is robust (asymptotically) in the traditional sense. However, we will show that the F test does not have inference robustness, although it has criterion robustness (for large n_i, $1 \le i \le c$). In fact, we will develop tests in later chapters which have both the criterion robustness as well as the efficiency robustness (as defined in the Preface); see also Tiku et al. (1986, Preface).

1.5 ROBUSTNESS OF t AND F FOR SMALL SAMPLES

We have shown above that both t and F tests are asymptotically robust in the traditional sense, namely, they have the stability (from distribution to distribution) of both the type I error as well as the power. In practice, however, we seldom have very large samples. Therefore, the asymptotic robustness of the tests is hardly a consolation. To study the robustness for small sample sizes, several investigations have been carried out; four excellent review papers by Tiku (1975 a), Cressie (1980), Ito (1980), and Tan (1982a) discuss the major works on this topic. The method based on Laguerre expansion, due to Tiku (1964; 1971 a, b), is known to have wide applicability; see Tan and Wong (1977, 1980), Tan (1982a), and Tan and Tiku (1999). Therefore, we adopt this approach here. We first consider the two-sample t statistic

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}; \; s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}; \; s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2, \tag{1.5.1}$$

and the F statistic (1.4.2). Since t^2 is the same as F with $\nu_1 = 1$ and $\nu_2 = n_1 + n_2 - 2$, we will only consider the F statistic. In order not to interrupt the development of the subject matter, some of the details which are mathematically involved are deferred to Appendix 1B. Here, we only give the main steps in deriving approximations to the distribution of F for small sample sizes.

1.6 DISTRIBUTION OF F UNDER NON-NORMALITY

To derive the null distribution of F under non-normality, we adopt the approach of Tiku (1964). Consider the following fixed-effects model with c blocks and n observations in each block

$$y_{ij} = \mu + b_i + e_{ij} \quad (i = 1, 2,..., c; j = 1, 2,..., n), \tag{1.6.1}$$

where μ is a constant, b_i is the effect due to the i th block, and e_{ij} are independently distributed random errors all with mean zero and variance σ^2. Specifically, the distribution of e_{ij} (j = 1, 2,…, n) has the finite cumulants

$$\kappa_{1i}, \kappa_2, \kappa_{3i}, \kappa_{4i},........; \kappa_{1i} = 0. \tag{1.6.2}$$

Note that we are not assuming the error distributions to be identical from block to block except that they have the same variance $\kappa_2 = \sigma^2$.

Define the rth standard cumulant of the error distribution in the ith block as (i = 1, 2,…, c)

$$\lambda_{ri} = \kappa_{ri}/\kappa_2^{r/2} \quad (r = 3, 4,...) \tag{1.6.3}$$

Consider the identity

$$\sum_{i=1}^{c}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 \equiv n\sum_{i=1}^{n}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{c}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$$

$$= S_b + S_e \tag{1.6.4}$$

where $\bar{y}_{i.} = \sum_{j=1}^{n} y_{ij}/n$ is the ith block mean and $\bar{y}_{..} = \sum_{i=1}^{c}\sum_{j=1}^{n} y_{ij}/nk$ is the grand mean. As said earlier, the components S_b and S_e are called block SS (sum of squares) and error SS with $\nu_1 = c - 1$ and $\nu_2 = c(n - 1) = N - c$ degrees of freedom, respectively. Write

$$X = S_b/2\sigma^2 \text{ and } Y = S_e/2\sigma^2 \tag{1.6.5}$$

and let f(X, Y) denote the joint pdf of X and Y. Denote the Laguerre polynomials (Appendix 1B) by $L_r^{(m)}(X)$ (r = 1, 2,......). Tiku (1964) makes

$$f(x, y) = \left\{\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\beta_{rs}L_r^{(m)}(x)L_s^{(k)}(y)\right\}p_m(x)\,p_k(y), \quad 0 < x, y < \infty, \tag{1.6.6}$$

with $m = \nu_1/2$ and $k = \nu_2/2$;

$$p_m(X) = \frac{1}{\Gamma(m)}e^{-X}X^{m-1}, \; 0 < X < \infty,$$

and, similarly, $p_k(Y)$. The expressions for the Laguerre polynomials and some of their properties are given in Appendix 1B.

By virtue of the orthogonality property of Laguerre polynomials (Appendix 1B.3), we get

$$\beta_{rs} = E[L_r^{(m)}(X)L_s^{(k)}(Y)] \Big/ \binom{m+r-1}{r}\binom{k+s-1}{s}. \tag{1.6.7}$$

What we seek here is an approximation to f(X, Y). To that end, we consider the partial sum

$$f(x, y) \cong \left\{\sum\sum_{0 \le r+s \le 4}\beta_{rs}L_r^{(m)}(X)L_s^{(k)}(y)\right\}p_m(x)p_k(y); \quad 0 < x, y < \infty \tag{1.6.8}$$

The coefficients work out in terms of the standard cumulants λ_{ri}, $r \geq 3$. Write

$$(\lambda_r) = \frac{1}{c}\sum_{i=1}^{c} \lambda_{ri} \text{ and } (\Lambda_r\Lambda_s) = \frac{1}{c}\sum_{i=1}^{c} \lambda_{ri}\lambda_{si} - (\lambda_r)(\lambda_s). \tag{1.6.9}$$

Note that the latter is zero only if the error distributions are identical from block to block. For simplicity in algebraic presentation, the brackets in (1.6.9) will be dropped.

From (1.6.7), the coefficients β_{rs} work in terms of the product moments

$$E[X^rY^s] = \left(\frac{1}{2\sigma^2}\right)^{r+s} E[(S_b - E(S_b))^r (S_e - E(S_e))^s]. \tag{1.6.10}$$

The values of $E[S_b{}^r S_e{}^s]$ are given in David and Johnson (1951). After some laborious but straight forward algebra, we obtain the following (Tiku, 1964):

$$\beta_{00} = 1,\ \beta_{10} = \beta_{01} = 0,$$

$$\beta_{20} = \frac{\nu_1}{N(\nu_1+2)}\lambda_4,\ \beta_{11} = \frac{1}{N}\lambda_4,\ \beta_{02} = \frac{\nu_2}{N(\nu_2+2)}\lambda_4,$$

$$\beta_{30} = -\frac{1}{N(\nu_1+2)(\nu_1+4)}\left\{\frac{\nu_1{}^2}{N}\lambda_6 + 4(\nu_1-1)\lambda_3{}^2 + (10\nu_1{}^2 - 4\nu_1 + 4)\Lambda_3{}^2\right\},$$

$$\beta_{12} = -\frac{1}{N(\nu_2+2)}\left\{\frac{\nu_2}{N}\lambda_6 + 4\lambda_3{}^2 + \frac{2}{\nu_1}(2\nu_1+\nu_2)\Lambda_3{}^2\right\} \tag{1.6.11}$$

$$\beta_{21} = -\frac{1}{N(\nu_1+2)}\left\{\frac{\nu_1}{N}\lambda_6 + 4\Lambda_3{}^2\right\}$$

$$\beta_{03} = -\frac{1}{N(\nu_2+2)(\nu_2+4)}\left\{\frac{\nu_2{}^2}{N}\lambda_6 + 4(\nu_2-\nu_1-1)(\lambda_3{}^2+\Lambda_3{}^2)\right\}$$

and so on . It should be noted here that β_{rs}, $r + s \leq 4$, are all the coefficients to order N^{-3} (see David and Johnson, 1951); the second-order coefficients in (1.6.11) involve only the fourth-order standard cumulants, i.e. λ_4; third-order coefficients involve only the sixth-order standard cumulants, i.e. $\lambda_6, \lambda_3{}^2$ and $\Lambda_3{}^2$, and so on.

By virtue of the following property of Laguerre polynomials (Sansone, 1959)

$$\frac{(-1)^n}{n!}\int_0^\infty x^n L_r{}^{(m)}(x)p_m(x)dx = \begin{cases} 0 & \text{for } r > n \\ \binom{m+r-1}{r} & \text{for } r = n \end{cases}, \tag{1.6.12}$$

it is not difficult to work out the product moments $E(X^rY^s)$ from (1.6.8)-(1.6.11). For $r + s \leq 4$, they agree with the expressions given in David and Johnson (1951).

1.7 DISTRIBUTION OF THE ONE-WAY CLASSIFICATION VARIANCE RATIO

Since the error distributions are not necessarily normal, we write $w = s_b{}^2/s_e{}^2 = (N-c)S_b/(c-1)S_e$ $(N = nc)$; the ratio w is denoted by F if the error ditributions are all normal. Submitting (1.6.8) to the transformation

$$(\nu_1/\nu_2)w = X/Y \quad (\nu_1 = c-1,\ \nu_2 = N-c) \tag{1.7.1}$$

and integrating over Y from zero to infinity, we obtain a Laguerre series approximation to the distribution of w, namely,

$$p(w) \cong p_0(w) + \sum_{2 \le r+s \le 4}\sum \beta_{rs}\, p_{rs}(w), \quad 0 < w < \infty, \tag{1.7.2}$$

$p_0(w) = p(w; \nu_1/2, \nu_2/2)$ is the normal-theory F-distribution and the other terms may be called corrective-functions due to finite population cumulants and are given by (see also Tan and Tiku, 1999)

$$\frac{\Gamma(\nu_1/2 + r)\Gamma(\nu_2/2 + s)}{r!\,s!\,\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} \sum_{i=0}^{r} (-1)^i \binom{r}{i} \times \left[\sum_{j=0}^{s} (-1)^j \binom{s}{j} p(w; \nu_1/2 + i, \nu_2/2 + j\right]; \tag{1.7.3}$$

$$p(w; \nu_1/2 + i, \nu_2/2 + j) = \frac{(\nu_1/\nu_2)^{\nu_1/2+r}}{\beta(\nu_1/2 + r, \nu_2/2 + s)} \times \frac{w^{\nu_1/2+r-1}}{\{1 + (\nu_1/\nu_2)w\}^{(\nu_1+\nu_2)/2+r+s}}. \tag{1.7.4}$$

For $r = s = 0$, (1.7.4) reduces to a central F distribution with degrees of freedom ν_1 and ν_2, which is the distribution of w if the errors e_{ij} are all normally distributed. Note that (1.7.4) is not a central F distribution with degrees of freedom $\nu_1 + 2r$ and $\nu_2 + 2s$ as is generally perceived.

Moments of w: It is easy to show that for $a \ge 0$,

$$\int_0^\infty w^a p(w; \nu_1/2 + r, \nu_2/2 + s)\, dw = \frac{\beta(\nu_1/2 + r + a, \nu_2/2 + s - a)}{(\nu_1/\nu_2)^a \beta(\nu_1/2 + r, \nu_2/2 + s)} \tag{1.7.5}$$

so that the a^{th} order moment of w can be expressed in the form

$$E(w^a) \cong B_{00} + \sum_{2 \le r+s \le 4}\sum \beta_{rs} B_{rs}; \tag{1.7.6}$$

B_{rs} are given in Tiku (1964, p.86). For example for $a = 1$ and $a = 2$, B_{00} is equal to

$$\nu_2/(\nu_2 - 2) \quad \text{and} \quad \nu_2^2 (\nu_1 + 2)/\nu_1(\nu_2 - 2)(\nu_2 - 4), \tag{1.7.7}$$

respectively. To order N^{-2}, we obtain the following expressions for the mean and variance of w:

$$E(w) \cong 1 + \frac{2}{N} + \frac{2}{\nu_1 N}\Lambda_3^2 + O\left(\frac{1}{N^2}\right) \tag{1.7.8}$$

and

$$V(w) \cong \frac{2}{\nu_1}\left\{1 + \frac{(\nu_1 + 6)}{N} - \frac{1}{N}\lambda_4 + \frac{8}{(\nu_1 + 2)N}\Lambda_3^2 + O\left(\frac{1}{N^2}\right)\right\}. \tag{1.7.9}$$

Realize that $\nu_1 = c - 1$ is fixed (and is not too large an integer), and $N = nc$ increases with n. It is interesting to see that Λ_3^2 appears in both E(w) as well as V(w), but λ_4 appears only in the latter. One would, therefore, expect the kurtosis to be of consequence in the computation of the probability $P(w \ge w_0)$, but if the error distributions in the blocks are not identical, this probability would be seriously affected by skewness. See also Senoglu and Tiku (2002).

Probability integral: We are interested in the value of the probability $P(w_0) = P\{w \ge w_0\}$ which from (1.7.2) works out to be

$$P(w_0) = \int_{w_0}^\infty p(w)dw \cong P_0(w_0) + \sum_{2 \le r+s \le 4}\sum \beta_{rs} I_{rs}(x_0); \tag{1.7.10}$$

where $P_0(w_0)$ is the normal-theory tail area, and

$$I_{r,s}(x_0) = \frac{\Gamma(\nu_1/2 + r)\,\Gamma(\nu_2/2 + s)}{r\,!\,s\,!\,\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)} \times \sum_{i=0}^{r} (-1)^i \binom{r}{i}\left[\sum_{j=0}^{s} (-1)^j \binom{s}{j} I_{x_o}(\nu_2/2 + j, \nu_1/2 + i)\right] \quad (1.7.11)$$

$$x_0 = 1/\{1 + (\nu_1/\nu_2)w_0\} \quad \text{and} \quad I_x(a, b) = \frac{1}{\beta(a, b)}\int_0^x u^{a-1}(1-u)^{b-1}du$$

is Karl Pearson's incomplete Beta function. Using some recurrence relations (Tiku, 1964), the functions $I_{r,s}(x_0)$ simplify. For example,

$$I_{2,0} = \frac{1}{4}A\{(\nu_2 - 2) - (\nu_1 + \nu_2)x_0\}$$

$$I_{1,1} = -\frac{1}{2}A\{\nu_2 - (\nu_1 + \nu_2)x_0\} \quad (1.7.12)$$

$$I_{0,2} = \frac{1}{4}A\{(\nu_2 + 2) - (\nu_1 + \nu_1)x_0\},$$

$$A = x_0^{\nu_2/2}(1-x_0)^{\nu_{1/2}}/\beta(\nu_2/2, \nu_1/2),$$

and so on.

Retaining only the terms of order $O(N^{-2})$ in (1.6.11), the formula (1.7.10) can be written as

$$P(w_0) \cong P_0(w_0) - \lambda_4 A + \lambda_3{}^2 B + \Lambda_3{}^2\, C + \lambda_6 D - \lambda_4{}^2\, E - \lambda_4{}^2 F \quad (1.7.13)$$

where $P_0(w_0)$ is the normal-theory value and A, B, C, D, E, and F are non-normality corrective-functions. They are expressions in terms of $I_{r,s}(w_0)$, and are tabulated in Tan and Tiku (1999, pp. 74-79) for n_1=1(1)10, 12 and 24, and ν_2 = 2(1)10, 12, 15, 20, 24, 30, 40, 60, 120 and ∞, for values of w_0 for which the normal-theory value $P_0(w_0)$ is equal to 0.05 and 0.01 (i.e., the normal-theory 5 and 1 percent significance levels); see also Tiku (1964, Table 3). We give their values in Appendix 1C for a few representative values of ν_1 and ν_2. The values give an idea about the relative magnitudes of the terms in (1.7.13).

Numerical values: We now present some numerical values computed from (1.7.13). They give the type I error under non-normality, the corresponding values under normality being 0.050. The λ_6 term in (1.7.13) has been ignored although its contribution can be considerable for some populations. We consider two cases: (i) two blocks (ν_1 = 1), and (ii) four blocks (ν_1 = 3). We cover both the situations when the error distributions are not identical from block to block, and when they are identical. The values are given in Table 1.2 reproduced from Tiku (1964, p.89). It can be seen that the type I error is seriously affected if the error distributions are not identical; the type I error usually goes up. In such situations, therefore, the normal-theory F test does not have criterion robustness. If the error distributions are identical, there is no serious effect on the type I error unless the kurtosis λ_4 is large in which case the type I error usually goes down which, of course, is not an undesirable phenomenon provided it does not pull down the power with it. Unfortunately, it does exactly that as will be seen later.

Gayen's contribution should be noted here. Gayen (1950) assumed that the error distributions from block to block are identical and obtained the exact joint distribution of X and Y, and that of w. He assumed, however, that the distribution is exactly given by the first four terms of the Edgeworth series, that is, the probability density function of $z = e_{ij}/\sigma$ (for all i and j) is given by

$$f(z) = \left[1 + \frac{\lambda_3}{6} H_3(z) + \frac{\lambda_4}{24} H_4(z) + \frac{\lambda_4^{\,2}}{72} H_6(z)\right] \phi(z), \qquad -\infty < z < \infty, \quad (1.7.14)$$

where $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ and $H_r(z)$ is the Hermite polynomial of degree r (Szegö, 1959). We note, however, that by neglecting cumulants of order greater than 4, his expression for $P(w_0)$ is in perfect agreement with (1.7.13) with $\Lambda_3^{\,2}$, λ_6 and $\Lambda_4^{\,2}$ equated to zero. However, Barton and Dennis (1952) have shown that (1.7.14) has non-negative ordinates only if $0 \le \lambda_4 \le 2.4$ and $\lambda_3^{\,2} \le 0.2$ so that Gayen's result is of limited applicability.

On the other hand, Tan and Wong (1977) have studied distributions like (1.6.8) and concluded that they represent plausible distributions for a much wider range of λ_3 and λ_4 (and higher cumulants) than those anticipated by Gayen (1950).

EXAMPLE 1.1: Suppose that the underlying distribution is one of the three distributions described by E. S. Pearson (1963, pp. 101-102) under his "Case B", namely, the Pearson type IV curve (see also Tiku et. al., 1986)

$$f(y) \propto \left(1 + \frac{y^2}{a^2}\right)^{-q} \exp\left(-d \tan^{-1} \frac{y}{a}\right), \; -\infty < y < \infty, \qquad (1.7.15)$$

the noncentral t distribution

$$f(y) \propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \exp\left(-\frac{\nu\tau^2/2}{\nu + t^2}\right) Hh_\nu\left(-\frac{t\tau}{\sqrt{\nu + t^2}}\right) \qquad (1.7.16)$$

$$Hh_\nu = \frac{1}{\Gamma(\nu + 1)} \int_0^\infty u^\nu e^{-(u+y)^2/2} du, \quad -\infty < y < \infty,$$

and the Johnson S_u curve (Johnson, 1949, 1965)

$$f(y) \propto (\cosh^{-1} y) \exp\left[-\frac{1}{2}(\gamma_0 + \gamma_1 \sinh^{-1} y)^2\right], \quad -\infty < y < \infty, \qquad (1.7.17)$$

for which $\lambda_3^{\,2} = 0.58$, $\lambda_4 = 1.23$, have very similar distributions within the range of significance frequency and all have values of λ_5 and λ_6 not far from 3.0 and 11.0, respectively. Here, $\Lambda_3^{\,2}$ and $\Lambda_4^{\,2}$ are both zero and from the remaining terms in (1.7.13),

$$P(w_0) \cong 0.050 - 1.23(0.00296) + 0.58(0.00147) + 11.0(0.0026) - 1.5129(0.00076)$$
$$= 0.0489 \text{ for } \nu_1 = 1 \text{ and } \nu_2 = 8;$$

$$P(w_0) \cong 0.050 - 1.23(0.00240) + 0.58(0.00097) + 11.0(0.00015) - 1.5129(0.00040)$$
$$= 0.0487 \text{ for } \nu_1 = 4 \text{ and } \nu_2 = 20.$$

Table 1.2: Values of non-normal type I error $P(w_0)$, the normal-theory type I error is $P_0(w_0) = 0.050$.

Cases where λ_{3t} and λ_{4t} differ with the blocks

Two blocks					Four blocks				
	(1)		(2)			(3)		(4)	
t	λ_{3t}	λ_{4t}	λ_{3t}	λ_{4t}	t	λ_{3t}	λ_{4t}	λ_{3t}	λ_{4t}
1	1.5	1.0	1.5	3.0	1	– 1.5	1.0	– 1.5	3.0
2	– 1.5	1.0	– 1.5	3.0	2	1.5	1.0	1.5	3.0
					3	0.5	1.0	0.5	3.0
					4	– 0.5	1.0	– 0.5	3.0

ν_2	$\nu_1 = 1$		ν_2	$\nu_1 = 3$	
6	0.091	0.073	6	0.061	0.040
8	0.087	0.074	8	0.060	0.041
12	0.080	0.073	12	0.058	0.045
24	0.068	0.066	24	0.056	0.050
40	0.062	0.061	40	0.054	0.051

Cases where λ_{3t} and λ_{4t} are identical for all blocks

ν_2	$\lambda_3 = 1.5, \lambda_4 = 1.0$	$\lambda_3 = 1.5, \lambda_4 = 3.0$	ν_2	$\lambda_3 = \lambda_4 = 1.0$	$\lambda_3 = 1.0, \lambda_4 = 3.0$
6	0.051	0.033	6	0.051	0.029
8	0.050	0.037	8	0.049	0.030
12	0.049	0.042	12	0.048	0.035
24	0.049	0.046	24	0.048	0.042
40	0.050	0.048	40	0.048	0.045

It is clear that the effect of the foregoing non-normality on the normal-theory value 0.050 is negligible, even for values of ν_2 as small as 8.

EXAMPLE 1.2: Another important situation from a practical point of view is when the underlying distribution is a mixture of two normals, for example,

$$0.90N(\mu, \sigma^2) + 0.10N(\mu, 4\sigma^2) \quad \text{(scale mixture)} \tag{1.7.18}$$

with mean μ and variance $\kappa_2 = \sigma^2 = 1.3$, and $\lambda_3^2 = 0$, $\lambda_4 = 1.4379$ and $\lambda_6 = 13.2726$.

Here,
$$P(w_0) \cong 0.050 - 1.4379(0.00296) + 13.2726(0.00026) - 2.0676(0.00076)$$
$$= 0.0476 \text{ for } \nu_1 = 1 \text{ and } \nu_2 = 8;$$
$$P(w_0) \cong 0.050 - 1.4379(0.00240) + 13.2726(0.00015) - 2.0676(0.00040)$$
$$= 0.0477 \text{ for } \nu_1 = 4 \text{ and } \nu_2 = 20.$$

Again, the effect of this non-normality on the normal-theory type I error is negligible.

The contribution of Subrahmaniam et. al. (1975) may be noted here. They evaluated the exact values of the type I error when the underlying distribution is

$$\pi N(\mu, \sigma^2) + (1 - \pi) N(\mu + \delta\sigma, \sigma^2) \text{ (location mixture)} \tag{1.7.19}$$

and found them not very different from the normal-theory value.

EXAMPLE 1.3: In some situations, the underlying distribution is symmetric and short-tailed ($\lambda_4 < 0$). Consider, for example, the uniform distribution

$$f(y) = 1/(\theta_2 - \theta_1), \quad \theta_1 < y < \theta_2 \tag{1.7.20}$$

For this distribution

$$\lambda_3^2 = 0, \quad \lambda_4 = -1.2 \quad \text{and} \quad \lambda_6 = 6.8567$$

Here,
$$P(w_0) \cong 0.050 + 1.2(0.00296) + 6.8567(0.00026) - 1.44(0.00076)$$
$$= 0.0542 \text{ for } \nu_1 = 1 \text{ and } \nu_2 = 8;$$
$$P(w_0) \cong 0.050 + 1.2(0.00240) + 6.8567(0.00015) - 1.44(0.00040)$$
$$= 0.0533 \text{ for } \nu_1 = 4 \text{ and } \nu_2 = 20.$$

The effect of short-tailed symmetric distributions is in general to make the type I error larger than its normal-theory value, although not by appreciable amounts.

There are situations when one is interested in testing hypotheses about location parameters but the underlying distributions are extremely non-normal, either symmetric or skew (Cauchy and exponential, for example). Some of these distributions do not even have a finite mean. In such situations, the formula (1.7.13) cannot be expected to give close approximations since it ignores higher cumulants and, in spite of the fact that the correction terms due to these cumulants are of order $O(N^{-2})$, their contributions could be considerable. However, Donaldson (1968) carried out a Monte Carlo investigation to study the effect of such departures on the type I error of the F test above. His results are given in Table 1.3. It is seen that the effect of such extreme departures from normality is in general to make the type I error

Table 1.3: Simulated type I error of the F test, normal-theory value is 0.050.

		$\nu_1 = 3$ (c = 4)		
n	ν_2	Normal	Exponential	Log-normal
4	12	0.048	0.041	0.035
8	28	0.047	0.040	0.034
16	60	0.048	0.047	0.037

smaller than its normal-theory value. We reiterate that this phenomenon is not undesirable from a practical point of view unless, of course, it has no substantial downgrading effect on the power of the test. Unfortunately, it has and pulls down the power. We will develop precedures in later chapters which have considerably higher power in such situations.

The pioneering work on the robustness of classical procedures of the then GDR group headed by D. Rasch (see Rasch, 1980, and the references cited there) may be noted.

1.8 NON-NORMAL POWER FUNCTION

As said earlier, the null distribution of F under the assumption of normality is central F with degrees of freedom $\nu_1 = c - 1$ and $\nu_2 = N - c$ ($N = nc$); the non-null distribution is noncentral F with degrees of freedom ν_1 and ν_2 and noncentrality parameter $\lambda = n \sum_{i=1}^{c} (b_i/\sigma)^2$.

To find the non-null distribution of F under non-normality, Tiku (1971b) developed a Laguerre series expansion. He assumed that the error distributions are identical from block to block. Let $\lambda_r = \kappa_r/\sigma^r$ (r = 3, 4,......) be the rth standard cumulant of the error distribution, and write

$$\tau = n\sum_{i=1}^{c} (b_i/\sigma)^3 \text{ and } \delta = n\sum_{i=1}^{c} (b_i/\sigma)^4 ; \tag{1.8.1}$$

τ and δ are noncentrality parameters of order 3 and 4, respectively, where the noncentrality parameter λ (of the non-null distribution of F under normality) is of order 2.

Proceeding exactly along the same lines as in Sections 1.6-1.7, an approximation to order N^{-2}, for the non-normal power $1 - \beta^*$ of the F test can be obtained (Tiku, 1971b):

$$1 - \beta^* \cong (1 - \beta) - \lambda_2\tau A + \lambda_4(B + B_1\delta) - \lambda_3^2 C + \lambda_5\tau D - \lambda_6 E + \lambda_4^2 H \tag{1.8.2}$$

where
$$1 - \beta = \int_{F_\alpha}^{\infty} p(F, H_1)\, dF \tag{1.8.3}$$

is the normal-theory power and the remaining terms are non-normality corrective-functions. It may be noted that $p(F, H_1)$ in (1.8.3) is a noncentral F distribution with degrees of freedom $\nu_1 = c - 1$ and $\nu_2 = N - c$ (N = nc) and noncentrality parameter λ. The numerical values of A, B, B_1, C, D, E and H are given in Tiku (1971b, Table 3). These values give an idea about the relative magnitudes of the non-normality corrective-functions.

Tiku (1971b, Table 1) computed the values of the difference β-β^* for numerous non-normal populations. He concluded that the effect of moderate non-normality on the power of the F-test is unimportant. However, λ_4 has a greater effect on the power than λ_3^2. The following expressions of the mean and variance are of interest:

$$\mu_1' \cong \left(1 + \frac{\lambda}{\nu_1}\right)\left(1 + \frac{2}{N}\right) + \frac{1}{\nu_1 N}\lambda_4\lambda + O\left(\frac{1}{N^2}\right) \tag{1.8.4}$$

and

$$\mu_2 \cong \frac{2}{p}\left[\frac{p+2}{\nu_1+2}\left(1 + \frac{\nu_1+6}{N}\right)\left(1 + \frac{\lambda}{\nu_1}\right)^2 + \frac{1}{2}\frac{p+2}{N}\left\{\frac{p}{p+2} - \left(1 + \frac{\lambda}{\nu_1}\right)\right\}\lambda_4 - \frac{1}{N}\frac{p+2}{\nu_1+2}\left(1 + \frac{\lambda}{\nu_1}\right)\lambda_4\lambda - \frac{2}{N(\nu_1+2)(\nu_1+2\lambda)}\left(1 + \frac{\lambda}{\nu_1}\right)\lambda_4\lambda + O\left(\frac{1}{N^2}\right)\right]; \tag{1.8.5}$$

$p = (\nu_1 + \lambda)^2/(\nu_1 + 2\lambda)$. Note that the kurtosis λ_4 enters both, the mean as well as the variance. The effect of the kurtosis λ_4 generally is to diminish the power of the normal-theory F test.

The contribution of Srivastava (1959) may be noted here. He assumed the underlying distribution to be the Edgeworth series (1.7.14) and obtained the exact non-null distribution of F. If we ignore the terms in λ_5 and λ_6, (1.8.2) completely agrees with his expression.

1.9 EFFECT OF NON-NORMALITY ON THE t STATISTIC

Let $y_1, y_2, \ldots, y_n$ be a random sample of size n from a normal population $N(\mu, \sigma^2)$. To test the null hypothesis $H_0 : \mu = 0$ against the alternative $H_1 : \mu \neq 0$, the statistic $t = \sqrt{n}\,\bar{y}/s$ is employed. Large values of $|t|$ (or t^2) lead to the rejection of H_0 in favour of H_1. To study the effect of non-normality on the type I error and the power of the $|t|$ test, write

$$X = (\sqrt{n}\,\bar{y})^2/2\sigma^2 \quad \text{and} \quad Y = ns^2/2\sigma^2;$$

$$\nu = n - 1,\ \sigma^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1). \tag{1.9.1}$$

Tiku (1971a) developed a Laguerre series expansion for the joint distribution f(X, Y) of X and Y. Proceeding exactly along the same lines as in Section 1.8, he obtained to order $O(n^{-5/2})$, an approximation to the power

$$1 - \beta* = \text{Prob}\{\,|\,t\,| \geq t_\alpha \mid H_1\,|\} \tag{1.9.2}$$

of the | t | test. His formula is similar to (1.8.2). In fact,

$$1 - \beta* \cong (1-\beta) + \lambda_3 A + \lambda_4 B - \lambda_3^2 C - \lambda_5 D + \lambda_3\lambda_4 E - \lambda_6 F + \lambda_4^2 G + \lambda_3\lambda_5 H. \tag{1.9.3}$$

The expressions for $1 - \beta$, A, B, etc., on the right hand side involve the noncentrality parameter $\delta = n(\mu/\sigma)^2$; if we equate it to zero, $1 - \beta^*$ reduces to α^* (the non-normal type I error), $1 - \beta$ reduces to α (the normal-theory type I error) and the remaining terms reduce to the non-normality corrective-functions (to correct the normal-theory type I error α).

Gayen (1949) and Srivastava (1958) assumed the underlying distribution to be the Edgeworth series (1.7.14), and obtained the exact formula for α^* (the non-normal type I error) and $1 - \beta^*$ (the exact non-normal power) of the | t | test. Their expressions are exactly similar to (1.9.3), but they have worked out corrections only due to λ_3, λ_4 and λ_3^2. Our formula above includes corrections due to a number of population standard cumulants other than these three. Calculations show that the agreement between their formulae and the first four terms in (1.9.3) is very good. Note that Srivastava (1958, Table 1) considers only one-sided alternatives H_1 ($\mu > 0$) and, therefore, the non-normality corrections in his formula (3) should not be confused with similar terms in (1.9.3).

EXAMPLE 1.4: Consider the Pearson distribution given in (1.7.15) with $\lambda_3 = 0.76$, $\lambda_4 = 1.23$, $\lambda_5 = 3.0$ and $\lambda_6 = 11.0$, and $\nu = 10$, $\delta = 1$, and normal-theory type I error equal to 0.050. We calculate the value of the non-normal power $1 - \beta^*$; the values of A, B, C,..., in (1.9.3) are given in Tiku (1971a, Table 3):

$$\begin{aligned} 1 - \beta^* &\cong 0.146 - 0.76(0.1272) + 1.23(0.0091) + 0.58(0.0010) + 3.0(0.0021) \\ &\quad - 0.93(0.0014) - 11.0(0.0005) + 1.51(0.0016) + 2.28(0.0010) \\ &= 0.065. \end{aligned}$$

It is seen that the normal-theory power 0.146 reduces to 0.065, a substantial reduction. Similar calculations reveal that the type I error is almost the same as its normal-theory value 0.050. In other words, as far as the distributions (1.7.15)-(1.7.17) are concerned, the t test has criterion robustness but not efficiency robustness.

Tiku (1971a, Table 1) calculated the values of $1 - \beta^*$ for non-normal populations. For $\delta = 0$, $1 - \beta^*$ reduces to the non-normal type I error α^*. He compared these values with the corresponding normal-theory values $1 - \beta$ and α, respectively. He concluded that the effect of λ_3, say $|\lambda_3| > 0.5$, on the type I error and power of the two-sided t-test is considerable; the effect of other standard cumulants is unimportant.

Concluding remarks: For non-normal populations with finite cumulants, the exact distribution f(X, Y) of $X = S_b/2\sigma^2$ and $Y = S_e/2\sigma^2$ is expanded as an infinite series in terms of Laguerre polynomials and gamma density functions (Tiku 1964, 1971 a, b; Davis, 1976; Tan, 1982a; Tan and Tiku, 1999). An approximation to the distribution is obtained by retaining terms of order up to 4. The non-normal distribution of $w = s_b^2/s_e^2$ is obtained by a variable transformation. Alternatively, by assuming an Edgeworth series distribution, Gayen (1949,

1950) and Srivastava (1959) obtained the distributions of w and t. From the numerical computations presented above and those of Gayen (1949, 1950), Srivastava (1959), Donaldson (1968), Tiku (1964; 1971a, b) and Tan (1982a), the following conclusions are drawn:

1. For the F test, if the distributions are identical from block to block and the departure from normality is moderate and if $n_1 = n_2 = \ldots = n_k = n$ (balanced design), the type I error and the power are quite insensitive. However, for severe departures such as exponential, log-normal and Cauchy distributions, the effects are quite considerable even if n_i ($1 \le i \le c$) are all equal. In general, the effects of departure from normality are more pronounced for unbalanced designs ($n_i \neq n$, $i = 1, 2, \ldots, c$) than for the balanced designs. If the error distributions are not identical from block to block, the effects on the type I error and the power are devastating (high type I error but low power) irrespective of whether the design is balanced or not.
2. For the t test if the underlying distribution is symmetric but the departure is moderate, the type I error and the power are quite insensitive to non-normality. For skew distributions with skewness $|\lambda_3| > 0.5$, the effects are quite considerable often resulting in high type I error and low power as compared to the normal-theory values.

From the results above and various other studies (Benjamin, 1983; Bradley, 1980; Geary, 1947; Lee and D'Agostino, 1976; Lee and Gurland, 1977; Pearson, 1931; Pearson and Adyanthaya, 1929; Posten 1978, 1982; Posten et al. 1982; Rasch, 1980, 1983; Sansing and Owen, 1974; Tan, 1977), it is concluded that the t and F tests make good hypothesis testing procedures (i.e., they have type I error close to the presumed value and have high power) only if the underlying distribution is normal or close to it. One would, therefore, prefer those test procedures which have criterion robustness (type I error not substantially higher than a presumed level) and have also efficiency robustness (i.e., they have high power) over a reasonably wide range of distributions. Several such tests will be presented in later chapters.

1.10 TESTING EQUALITY OF TWO VARIANCES

Let $y_{i1}, \ldots, y_{in_i}$ ($i = 1, 2$) be two independent random samples from the normal populations $N(\mu_i, \sigma_i^2)$. To test the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$, the classical F test is based on the ratio of the two sample variances

$$s_1^2 = \sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2/(n_1 - 1) \quad \text{and} \quad s_2^2 = \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2/(n_2 - 1). \tag{1.10.1}$$

The convention is to place the larger sample variance in the numerator. If s_1^2 is the larger sample variance then the F statistic is defined as

$$F = s_1^2/s_2^2 \tag{1.10.2}$$

Large values of F lead to the rejection of H_0 in favour of $H_1 : \sigma_1^2 > \sigma_2^2$. The null distribution of F is central F with degrees of freedom $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$. The power of the test is given by

$$1 - \beta = P\{F \ge (\sigma_2^2/\sigma_1^2)\, F_\alpha(\nu_1, \nu_2)\} \tag{1.10.3}$$

where F_α is the $100(1 - \alpha)\%$ point of the central F distribution.

To derive the null distribution of F under non-normality, write $X = (n_1 - 1)s_1^2/2\sigma^2$ and $Y = (n_2 - 1)s_2^2/2\sigma^2$. The distribution of X and Y can be expressed in terms of Laguerre polynomials (Tiku, 1964; Tan and Tiku, 1999).

$$f(x, y) = \left\{\sum_{r=0}^{\infty} \alpha_{r(1)} L_r^{(m)}(x) p_m(y)\right\} \times \left\{\sum_{s=0}^{\infty} \alpha_{s(2)} L_s^{(k)}(y) p_s(y)\right\}, \; 0 < x, y < \infty, \quad (1.10.4)$$

since X and Y are independently distributed; $m = \nu_1/2 = (n_1 - 1)/2$ and $k = \nu_2/2 = (n_2 - 1)/2$. The coefficients $\alpha_{r(1)}$ and $\alpha_{s(2)}$ work out in terms of the standard cumulants

$$\lambda_{j(1)} = \kappa_{j(1)}/\kappa_2^{j/2} \quad \text{and} \quad \lambda_{j(2)} = \kappa_{j(2)}/\kappa_2^{j/2} \quad (j \geq 3) \quad (1.10.5)$$

of the two populations, respectively. For example (i = 1, 2)

$$\alpha_{0(i)} = 1, \quad \alpha_{1(i)} = 0, \quad \alpha_{2(i)} = \frac{\nu_i}{(\nu_i + 1)(\nu_i + 2)} \lambda_{4(i)},$$

$$\alpha_{3(i)} = -\frac{1}{(\nu_i + 1)(\nu_i + 2)(\nu_i + 4)} \left[\frac{\nu_i^2}{\nu_i + 1} \lambda_{6(i)} + 4(\nu_i - 1)\lambda_{3(i)}^2\right] \quad (1.10.6)$$

and so on. Again, $\alpha_{2(i)}$ involve the standard cumulants of order 4 only, $\alpha_{3(i)}$ involve standard cumulants of order 6 only, and so on.

Proceeding exactly along the same lines as before, an approximation to the non-normal probability Prob $\{F \geq F_\alpha(\nu_1, \nu_2) \mid H_0\}$ is obtained. To order $O(n^{-2})$, $n = \min(n_1, n_2)$, the formula includes terms in $\lambda_{4(i)}, \lambda_{3(i)}^2, \lambda_{3(1)}\lambda_{3(2)}, \lambda_{6(i)}, \lambda_{4(i)}^2$ and $\lambda_{4(1)}\lambda_{4(2)}$. From numerous computations it is seen that the non-normal type I error α^* is considerably larger than its normal-theory value, irrespective of how large ν_1 or ν_2 are. Consider, for example, the situation when both the samples have the Pearson distribution (1.7.15). The values of the type I error under this non-normality are given in Table 1.4. It is seen that the F test (1.10.2) is very sensitive to non-normality and the type I error is considerably higher than the normal-theory value irrespective of the sample sizes. This is a very undesirable phenomenon.

Mean and Variance: The following expressions of the mean and variance also reflect on the sensitivity of $F = s_1^2/s_2^2$ to non-normality; $n = \min(n_1, n_2)$:

Table 1.4: Values of type I error under non-normality, normal theory value is 0.050.

			ν_2		
ν_1	4	6	8	12	24
4	0.058	0.059	0.059	0.059	0.059
8	0.062	0.065	0.066	0.067	0.067
12	0.063	0.067	0.070	0.072	0.073
24	0.062	0.068	0.073	0.078	0.082

$$\mu_1' = 1 + \frac{2}{\nu_2}\left(1 + \frac{2}{\nu_2}\right) + \frac{\lambda_{4(2)}}{\nu_2}\left(1 - \frac{1}{\nu_2}\right) - \frac{4\lambda_{3(2)}^2}{\nu_2^2} - \frac{\lambda_{6(2)}}{\nu_2^2} + \frac{3\lambda_{4(2)}^2}{\nu_2^2} + O(n^{-3}) \quad (1.10.7)$$

and
$$\mu_2 = \frac{2}{\nu_1} + \frac{2}{\nu_2}\left(1 + \frac{6}{\nu_1}\right) + \frac{\lambda_{4(1)}}{\nu_1}\left(1 - \frac{1}{\nu_1}\right) + \frac{\lambda_{4(2)}}{\nu_2}\left(1 - \frac{5}{\nu_2}\right) - \frac{8\lambda_{3(2)}^2}{\nu_2^2}$$
$$- \frac{2\lambda_{6(2)}}{\nu_2^2} + \frac{2\lambda_{4(2)}^2}{\nu_2^2} + \frac{3\lambda_{4(1)}\lambda_{4(2)}}{\nu_1\nu_2} + O(n^{-3}) \quad (1.10.8)$$

If the variance s_2^2 in the denominator is calculated from a normal sample, then $\lambda_{s(i)} = 0$ for all $i \geq 3$. In that case, (1.10.7)-(1.10.8) reduce to

$$\mu_1' \cong 1 + \frac{2}{\nu_2}\left(1+\frac{2}{\nu_2}\right) \quad \text{and} \quad \mu_2 \cong \frac{2}{\nu_1} + \frac{2}{\nu_2}\left(1+\frac{6}{\nu_1}\right) + \frac{\lambda_{4(1)}}{\nu_1}\left(1-\frac{1}{\nu_1}\right). \qquad (1.10.9)$$

The effect of non-normality persists but is not as pronounced as in situations where both samples come from non-normal populations.

EXAMPLE 1.5: Consider the following Darwin's well-known data (Fisher, 1966) which represent the differences (in heights) between cross-and self-fertilized plants of the same pair grown together in one pot:

49 – 67 8 16 6 23 28 41 14 29 56 24 75 60 – 48

The problem is to estimate the difference d and to test the null hypothesis $H_0 : d = 0$ against the alternative $H_1 : d > 0$. Fisher assumed normality and obtained the folowing ML estimate:

$$\bar{d} = 20.933 \text{ with standard error } \pm \frac{s}{\sqrt{n}} = \frac{37.744}{\sqrt{15}} = \pm 9.745$$

He then calculated the t statistic

$$t = \sqrt{15}\ (20.933)/37.744$$
$$= 2.15 \text{ (99 percent point is 2.62)}$$

and concluded that there is no overwhelming reason to reject H_0.

We will show in Chapter 8 that this conclusion is erroneous resulting from the wrongful assumption of normality. By using a robust test we will show that H_0 should be rejected even at 0.1% significance level, let alone the 1% significance level considered by Fisher (1966).

SUMMARY

In this Chapter, we discuss the efficiencies of the sample mean $\bar{y}$ and the sample variance s^2. We show that they are inefficient unless the underlying distribution is normal or close to it. We show that for long-tailed symmetric distributions, the Tukey estimator $\hat{\mu}_T$ based on censored samples is enormously more efficient than $\bar{y}$. We develop a Laguerre series expansion for the joint distribution of $\bar{y}$ and s^2. By a variable transformation, we derive the distribution of the classical Student statistic $t = \sqrt{n}\,\bar{y}/s$ under non-normality. We evaluate the effect of non-normality on the type I error. We show that it is not substantial but the power of the test is greately diminished. In other words, the classical t test has criterion robustness but not efficiency robustness. We extend the results to testing the equality of c (≥ 2) location parameters in the context of experimental design. We show that if the observations in the blocks are nonidentical, the type I error (let alone the power) of the classical F test is affected drastically. We give numerical examples to illustrate these findings. We develop a Laguerre series expansion for the distribution of the ratio of two independent sample variances $w = s_1^2/s_2^2$. We show that the distribution of w is very sensitive to deviations from normality.

APPENDIX 1A

EXPECTED VALUES AND VARIANCES AND COVARIANCES

Let $y_{(1)} \le y_{(2)} \le \ldots \le y_{(n)}$ be the order statistics of a random sample of size n from a distribution of the type $(1/\sigma)f((y-\mu)/\sigma)$. Write $z = (y-\mu)/\sigma$ and

$$z_{(r)} = \{y_{(r)} - \mu\}/\sigma, \quad 1 \le r \le n. \tag{1A.1}$$

The probability density function of $u = z_{(r)}$ is given by (David, 1981, p.9)

$$p(u) = \frac{n!}{(r-1)!\,(n-r)!}[F(u)]^{r-1}[1-F(u)]^{n-r} f(u), \quad -\infty < u < \infty, \tag{1A.2}$$

where $F(z) = \int_{-\infty}^{z} f(z)dz$ is the cumulative distribution function. The joint probability density function of $u = z_{(r)}$ and $v = z_{(s)}$ $(r < s)$ is given by (David, 1981, p. 10)

$$p(u, v) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}[F(u)]^{r-1}[F(v)-F(u)]^{s-r-1} \times [1-F(v)]^{n-s} f(u) f(v), \quad -\infty < u < v < \infty. \tag{1A.3}$$

EXAMPLE 1A.1: Consider the uniform distribution U(0, 1) : $f(y) = 1$, $0 < y < 1$, with $F(z) = z$.

Here,
$$p(u) = \frac{n!}{(r-1)!\,(n-r)!} u^{r-1}(1-u)^{n-r}, \quad 0 < u < 1, \tag{1A.4}$$

which is a beta distribution $B(r, n-r+1)$, and

$$p(u, v) = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!} u^{r-1}(v-u)^{s-r-1}(1-v)^{n-s}, \quad 0 < u < v < 1 \tag{1A.5}$$

The following results follow immediately from (1A.4)–(1A.5):

$$\mu_{r:n} = E\{z_{(r)}\} = \frac{r}{n+1} \text{ and } \sigma_{r,s:n} = \mathrm{Cov}\{z_{(r)}, z_{(s)}\} = \frac{r(n-s+1)}{(n+1)^2(n+2)} \quad (r \ge s); \tag{1A.6}$$

see David (1981, pp. 35-36).

EXAMPLE 1A.2: Consider the exponential distribution

$$E(0, 1): f(z) = e^{-z}, \quad 0 < z < \infty, \tag{1A.7}$$

with $F(z) = \exp(-z)$. Here, we have the following results:

$$\mu_{r:n} = \sum_{i=1}^{r} \frac{1}{(n-i+1)} \text{ and } \sigma_{r,s:n} = \sum_{i=1}^{r} \frac{1}{(n-i+1)^2} \quad (r \ge s). \tag{1A.8}$$

Explicit algebraic expressions for $\mu_{r,\,r:n}$, $\sigma_{r,\,r:n}$ and $\sigma_{r,\,s:n}$ can be obtained for a few other distributions. For example, Malik (1967) and Huang (1975) have obtained these expressions for the power-function and Pareto distributions. In general, however, it is not possible to obtain explicit algebraic expressions for the means and the variances and covariances of order statistics. They have to be evaluated numerically, therefore; see, for example, Tietjen et al.

(1977) and Tiku and Kumra (1981). Of course, a number of authors have developed and discussed numerical and other procedures for computing the expected values and the variances and covariances of order statistics (see, for example, Singh, 1972; Balakrishnan, 1984; Balakrishnan and Chan, 1992a, b; Barnett, 1966b; David et al., 1977; Govindarajulu, 1966; Gupta et al., 1967; Harter, 1964; Mann et al., 1973; Tietjen et al., 1977; Shah 1966, 1970; Balakrishnan and Leung, 1988; Arnold et al., 1992).

The moments of order statistics satisfy certain recurrence relations and that makes their computation manageable; see Gavindarajulu (1963), Prescott (1974) and Arnold et al. (1992). A very useful recurrence relation that holds for any arbitrary distribution f(z) is

$$r\mu_{r+1:n} + (n-r)\mu_{r:n} = n\mu_{r:n-1} \quad (1 \geq r \geq n-1). \tag{1A.9}$$

For any arbitrary distribution f(z),

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\sigma_{r,s:n} = n \tag{1A.10}$$

If the distribution of z is symmetric then

$$\mu_{n-r+1} = -\mu_{r:n} \quad \text{and} \quad \sigma_{n-r+1,\, n-r+1:n} = \sigma_{r,\, r:n}, \quad \text{and} \quad \text{for } r \leq s \tag{1A.11}$$

$$\sigma_{n-s+1,\, n-r+1:n} = \sigma_{r,\, s:n} \quad \text{(double symmetry).} \tag{1A.12}$$

For a detailed treatment of order statistics and their properties, see David (1981). It may be noted that in (1A.1)

$$E\{y_{(r)}\} = \mu + \sigma\mu_{r:n}, \quad V\{y_{(r)}\} = \sigma_{r,\, r:n}\,\sigma^2 \text{ and } \mathrm{Cov}\{y_{(r)}, y_{(s)}\} = \sigma_{r,\, s:n}\,\sigma^2 \tag{1A.13}$$

A tabulation of the expected values and the variances and covariances of order statistics becomes voluminous very quickly. Therefore, most of the tables cover only the sample sizes $n \leq 20$. For large n (say $n > 20$), David and Johnson (1954) give equations which provide accurate approximations. These equations are reproduced in Tiku et al. (1986, pp. 71-73) and their terms worked out for normal, gamma, logistic, extreme-value and Student t distributions.

APPENDIX 1B

LAGUERRE POLYNOMIALS

The Laguerre polynomial $L_r^{(m)}(x)$ of degree r in x and parameter $m > 0$ is defined as (Szegö, 1959)

$$L_r^{(m)}(x) = \frac{1}{r!}\sum_{j=0}^{r}\binom{r}{j}(-x)^j\,\frac{\Gamma(m+r)}{\Gamma(m+j)}, \quad r = 0, 1, 2, \dots . \tag{1B.1}$$

In particular,

$$L_0^{(m)}(x) = 1$$

$$L_0^{(m)}(x) = m - x$$

$$L_0^{(m)}(x) = \frac{1}{2!}\,m(m+1) - (m+1)x + \frac{x^2}{2!} \tag{1B.2}$$

$$L_0^{(m)}(x) = \frac{1}{3!}\,m(m+1)(m+2) - \frac{1}{2!}\,(m+1)(m+2)x + (m+2)\,\frac{x^2}{2!} - \frac{x^3}{3!}$$

and so on. The Leguerre polynomials satisfy the following orthogonality property:

$$\int_0^\infty L_r^{(m)}(x)\,L_j^{(m)}(x)p_m(x)dx = \begin{cases} 0 & \text{if } r \neq j \\ C_r^{(m)} & \text{if } r = j \end{cases} \tag{1B.3}$$

where $$C_r^{(m)} = \frac{m(m+1)\ldots(m+r-1)}{r\,!} = \binom{m+r-1}{r},$$

and $$p_m(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}, \quad 0 < x < \infty, \tag{1B.4}$$

is the gamma density function.

Let $X > 0$ be a random variable with finite standard cumulants $\lambda_r = \kappa_r/\sigma^r$, $r \geq 3$; σ^2 is the variance of X. The probability density function of X can be represented exactly by a Laguerre series expansion (Tiku, 1964; Davis, 1976; Tan and Tiku, 1999), namely,

$$f(x) = \left\{\sum_{r=0}^{\infty} \alpha_r L_r^{(m)}(x)\right\} p_m(x), \quad 0 < x < \infty;\ \mu = E(X). \tag{1B.5}$$

From the orthogonality property (1B.3), it follows that

$$\alpha_r = E\{L_r^{(m)}(X)\} \Big/ \binom{m+r-1}{r}; \tag{1B.6}$$

$\alpha_0 = 1$ and $\alpha_1 = 0$. The coefficients α_r $(r \geq 2)$ work out in terms of the standard cumulants as in (1.10.4). First few terms of (1B.5) give acurate approximations for the distribution of X; see, for example, Tiku (1964, 1965).

Let $X > 0$ and $Y > 0$ be two random variables with finite (r, s) th mixed cumulants κ_{rs}, and $\lambda_{rs} = \kappa_{rs}/\sigma_1^r\sigma_2^s$ (σ_1^2 and σ_2^2 are the variances of X and Y, respectively); $r + s \geq 3$. The joint probability density function of X and Y can be represented exactly by

$$f(x, y) = \left\{\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \beta_{ij} L_i^{(m)}(x) L_j^{(k)}(y)\right\} p_m(x)\, p_k(y), \quad 0 < x, y < \infty; \tag{1B.7}$$

$m = E(X)$ and $k = E(Y)$. From the orhogonality property of Laguerre polynomials,

$$\beta_{ij} = E\{L_i^{(m)}(X) L_j^{(k)}(Y)\} \Big/ \binom{m+i-1}{i}\binom{k+j-1}{j}; \tag{1B.8}$$

$\beta_{00} = 1$ and $\beta_{10} = \beta_{01} = 0$. The coefficients β_{ij} work out in terms of the standard mixed cumulants as in (1.6.8). The first few terms give accurate approximations; see Tan and Tiku (1999) who discuss in detail the usefulness of Laguerre series expansions and give references to the enormous amount of work done in this area.

APPENDIX 1C

NON-NORMALITY CORRECTIVE-FUNCTIONS, NORMAL THEORY VALUE *$P_0(w_0)$ = 0.050.

ν_2	A	B	C	D	E	F	A	B	C	D	E	F
			$\nu_1 = 1$						$\nu_1 = 2$			
2	480	840	1882	76	194	72	428	910	1385	57	80	66
4	481	545	2263	68	212	45	522	818	1633	72	196	62
6	378	279	2052	42	129	– 6	471	525	1471	58	172	40
8	296	147	1802	26	76	– 22	405	326	1285	43	127	19
10	238	81	1590	16	47	– 23	348	205	1130	32	92	8
15	153	21	1213	6	18	– 16	250	67	863	16	46	– 0
20	110	5	977	3	9	– 10	192	19	698	10	26	– 0
30	69	– 2	702	1	3	– 4	129	– 9	506	4	11	0
60	31	– 2	379	– 0	1	– 1	64	– 14	278	1	3	1
120	20	– 1	272	– 0	0	– 0	30	– 9	141	0	1	0
			$\nu_1 = 4$						$\nu_1 = 8$			
2	322	904	1086	31	– 51	– 90	210	866	927	13	– 138	– 154
4	444	1024	1128	49	61	– 26	312	1134	914	23	– 89	– 128
6	446	807	876	49	111	3	339	1043	657	27	– 15	– 72
8	414	591	637	42	112	5	336	877	400	26	27	40
10	377	429	456	35	98	2	348	205	1130	32	92	8
15	296	199	192	22	62	– 4	277	441	– 140	18	50	– 11
20	240	97	69	15	40	– 5	238	281	698	10	26	– 0
30	172	19	– 25	7	19	– 3	182	150	– 422	8	24	– 6
60	91	– 18	– 62	2	5	– 1	104	18	– 401	3	8	– 3
120	50	– 17	– 50	1	1	– 0	57	– 7	– 282	1	2	– 1

*All the values are multiples of 10^5, e.g., in the first row A = 0.00480, B = 0.00840 and so on.

CHAPTER 2

Estimation of Location and Scale Parameters

2.1 INTRODUCTION

Consider a location-scale distribution (population) of the type $(1/\sigma)\, f((y - \mu)/\sigma)$, where μ is a location parameter and σ is a scale parameter; μ and σ might as well be the mean and standard deviation of the distribution. A location-scale distribution is one which under the transformation $z = (y - \mu)/\sigma$ reduces to $f(z)$, and $f(z)$ is free of μ and σ. Let $y_1, y_2,\ldots, y_n$ be a random sample of size n from $(1/\sigma)\, f((y - \mu)/\sigma)$. Our aim is to obtain efficient estimators of μ and σ (and σ^2). Ideally, one would like these estimators to be fully efficient at any rate for large n. As said earlier in Chapter 1, a fully efficient estimator is one which is unbiased and its variance is equal to the Cramér-Rao minimum variance bound (MVB). Of course, one realizes that such estimators do not exist for small sample sizes n other than for a few distributions in the exponential family, e.g, normal and exponential. For the normal

$$N(\mu, \sigma^2) : (1/\sqrt{2\pi}\sigma)\, \exp\{-(y - \mu)^2/2\sigma^2\}, \quad -\infty < y < \infty, \tag{2.1.1}$$

the sample mean $\bar{y}$ is fully efficient as an estimator of μ. For the exponential

$$E(0, \sigma) : (1/\sigma)\, \exp(-y/\sigma), \quad 0 < y < \infty, \tag{2.1.2}$$

the sample mean $\bar{y}$ is fully efficient as an estimator of σ (the scale parameter). The method that gives, under some very general regularity conditions, fully efficient estimators at any rate for large n, is the Fisher method of maximum likelihood.

2.2 MAXIMUM LIKELIHOOD

Consider first the estimation of a single parameter θ. Let $f(y, \theta)$ be the pdf (probability density function) of y, θ being an unknown parameter. Let $y_1, y_2,\ldots, y_n$ be a random sample of size n from $f(y, \theta)$. The likelihood function L is the joint pdf (probability density function) of $y_1, y_2,\ldots, y_n$;

$$L = \prod_{i=1}^{n} f(y_i, \theta) \tag{2.2.1}$$

The ML (maximum likelihood) estimator of θ is that value which maximizes L or ln L;

$$\ln L = \sum_{i=1}^{n} \ln f\,(y_i, \theta) \tag{2.2.2}$$

In fact, the ML estimator is the solution of the equation

$$\frac{d \ln L}{d\theta} = \sum_{i=1}^{n} g(y_i, \theta) = 0 \tag{2.2.3}$$

where $$g(y, \theta) = f'(y, \theta)/f(y, \theta) \tag{2.2.4}$$

is a function of y and θ. Note that $E(d \ln L/d\theta) = 0$ and $V(d \ln L/d\theta) = -E(d^2 \ln L/d\theta^2)$ (Kendall and Stuart, 1979, p. 9).

If g(y, θ) is linear in θ, (2.2.3) has an explicit and unique solution. If it is nonlinear, (2.2.3) does not in general have an explicit solution and has to be solved by iterative methods. That can be problematic for reasons of (i) multiple roots, (ii) nonconvergence of iterations, or (iii) convergence to wrong values (Barnett, 1966a; Lee et al., 1980; Tiku and Suresh, 1992; Vaughan, 1992a). In fact, if the data contains outliers, the iterations with likelihood equations might never converge (Puthenpura and Sinha, 1986). Moreover, the solution(s) of (2.2.3) might be inadmissible as an estimator(s). The following are illustrative examples.

EXAMPLE 2.1: Consider estimating the mean μ of a normal population $N(\mu, \sigma^2)$. Here,

$$L \propto \left(\frac{1}{\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}$$

and $$\frac{d \ln L}{d\mu} = \frac{1}{\sigma}\sum_{i=1}^{n} g(z_i) = 0, \quad g(z) = (y - \mu)/\sigma; \tag{2.2.5}$$

g(z) being linear in μ; the equation $d\ln L/d\mu = 0$ has an explicit and unique solution $\hat{\mu} = \sum_{i=1}^{n} y_i / n = \bar{y}$ (the sample mean). Incidentally, $d\ln L/d\mu$ can be reorganized to assume the form

$$\frac{d \ln L}{d\mu} = \frac{n}{\sigma^2}(\bar{y} - \mu). \tag{2.2.6}$$

A formulation like (2.2.6) implies that an estimator (in this case, $\bar{y}$) is fully efficient for all n (Kendall and Stuart, 1979, pp. 10-11). Realize that $E(d\ln L/d\mu) = 0$ and the Cramer-Rao minimum variance bound (MVB) is $1/\{-E(d^2 \ln L/d\mu^2)\} = \sigma^2/n$. The exact variance $V(\bar{y}) = \sigma^2/n$ is equal to the MVB. Also, $E(\bar{y}) = \mu$ since $E(d\ln L/d\mu) = 0$. Therefore, $\bar{y}$ is fully efficient as an estimator of μ. Realize that the reciprocal of the multiplier n/σ^2 of $(\bar{y} - \mu)$ in (2.2.6) gives the variance of $\bar{y}$ and this is a general result, namely, if

$$\frac{d\ln L}{d\theta} = A(\theta)\,(t - \theta), \tag{2.2.7}$$

A(θ) being free of $y_1, y_2, \ldots, y_n$, then t is an unbiased estimator of θ and $V(t) = 1/A(\theta)$, and t is the MVB estimator. In other words, t is fully efficient.

Remark. It may be noted that a MVB estimator is unique. If not, let t_1 and t_2 be two MVB estimators of θ with $V(t_1) = V(t_2) = V$, say. Consider the third estimator

$$t_3 = \frac{1}{2}(t_1 + t_2)$$

Now, $E(t_3) = \theta$ and $$V(t_3) = \frac{1}{4}\{2V + 2\,\mathrm{Cov}\,(t_1, t_2)\}.$$

Since V is the minimum variance bound,

$V(t_3) \geq V$ which implies that $Cov(t_1, t_2) \geq V$.

That is a contradiction unless $t_1 = ct_2$ where c is a constant. Clearly, c = 1 since $E(t_1) = E(t_2) = \theta$. That proves the result.

EXAMPLE 2.2: Consider estimating s in (2.1.1). Now,

$$\frac{d \ln L}{d\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (y_i - \mu)^2 = 0.$$

Writing $$S^2 = \sum_{i=1}^{n} (y_i - \mu)^2 / n$$

we have $$\frac{d \ln L}{d\sigma} = \frac{n}{\sigma^3} (S - \sigma)(S + \sigma) \tag{2.2.8}$$

$$= \frac{n}{\sigma^3} (S^2 - \sigma^2) = 0.$$

This equation has only one admissible root $\hat{\hat{\sigma}} = S$ which is the ML estimator provided μ is known. If μ is not known it is replaced by $\bar{y}$ (the solution of $d\ln L/d\mu = 0$), and

$$s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / (n-1)$$

is an unbiased estimator of σ^2. The formulation (2.2.8) may be noted; we will refer to it from time to time. Realize that the MVB $(\sigma) = 1/\{-E(d^2 \ln L/d\sigma^2)\} = \sigma^2/2n$, and $V(s) \cong \sigma^2/2n$ (equation 1.2.9) and $E(s) \cong \sigma$ for large n. The sample standard deviation s is, therefore, fully efficient for large n.

EXAMPLE 2.3: Consider estimating the mean (location parameter) μ of the distribution

$$f(y, p) \propto \frac{1}{\sigma} \left\{1 + \frac{(y-\mu)^2}{k\sigma^2}\right\}^{-p}, -\infty < y < \infty; \tag{2.2.9}$$

$k = 2p - 3$ and $p \geq 2$. Realize that $E(y) = \mu$ and $V(y) = \sigma^2$. For $1 \leq p < 2$, k is taken to be equal to 1 in which case σ is simply a scale parameter. For p = 1, E(Y) does not exist and μ is a location parameter. The distribution of $\sqrt{(\nu / k)}(y - \mu)/\sigma$ is the Student t with $\nu = 2p - 1$ degrees of freedom. Here,

$$\ln L = -n \ln \sigma - p \sum_{i=1}^{n} \ln\{1 + (1/k)z_i^2\},\ z_i = (y_i - \mu)/\sigma \quad \text{and} \tag{2.2.10}$$

$$\frac{d \ln L}{d\mu} = \frac{2p}{k\sigma} \sum_{i=1}^{n} g(z_i) = 0, \quad g(z) = z/\{1 + (1/k)z^2\}. \tag{2.2.11}$$

The function g(z) is nonlinear and, therefore, (2.2.11) has no explicit solution. Evaluation of the ML estimate is problematic since (2.2.11) has multiple roots for all $p < \infty$ (Vaughan, 1992a). For $p = \infty$, (2.2.9) reduces to $N(\mu, \sigma^2)$.

EXAMPLE 2.4: Consider the simple linear regression model (Islam et al., 2001)

$$y_i = \theta_0 + \theta_1 x_i + e_i, \quad 1 \leq i \leq n, \tag{2.2.12}$$

with usual interpretation of the parameters θ_0 and θ_1; e_i are iid (independently and identically distributed) random errors and have the Weibull distribution

$$W(p, \sigma) : (p/\sigma^p)e^{p-1} \exp\{-(e/\sigma)^p\}, \quad 0 < e < \infty \ (p > 0). \tag{2.2.13}$$

Here,
$$\frac{d \ln L}{d\mu} = -\frac{np}{\sigma} + \frac{p}{\sigma}\sum_{i=1}^{n} z_i^{\,p} = 0, \quad z_i = (y_i - \theta_0 - \theta_1 x_i)/\sigma. \tag{2.2.14}$$

The solution of (2.2.14) is the ML estimator of σ provided θ_0 and θ_1 are known,

$$\hat{\hat{\sigma}} = \left\{\sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^p / n\right\}^{1/p}. \tag{2.2.15}$$

However, θ_0 and θ_1 are not known in practice and will have to be replaced by their estimates $\tilde{\theta}_0$ and $\tilde{\theta}_1$. But $\tilde{e}_i = y_i - \tilde{\theta}_0 - \tilde{\theta}_1 x_i$ are not necessarily positive for all i, even if $\tilde{\theta}_0$ and $\tilde{\theta}_1$ are the ML estimates. Consequently, $\hat{\hat{\sigma}}$ is not necessarily real if p is not an integer. Thus, $\hat{\hat{\sigma}}$ is inadmissible. One way to rectify this situation is to equate $\tilde{e}_i$ to zero whenever it is negative. But then $\hat{\hat{\sigma}}$ develops substantial bias. That is not a good prospect.

2.3 MODIFIED LIKELIHOOD

What we need is a method of estimation which captures the beauty of maximum likelihood but alleviates its computational difficulties. One of these methods, and perhaps the most viable, is the method of modified likelihood estimation which originated with Tiku (1967a, b; 1968a, b, c; 1970; 1973), and Tiku and Suresh (1992). A remarkable property of this method is that it gives estimators which are highly efficient and have exactly the same forms irrespective of the underlying distribution. The method proceeds as follows:

Consider the likelihood equation to estimate an unknown location parameter θ,

$$\frac{d \ln L}{d\theta} = \frac{1}{\sigma}\sum_{i=1}^{n} g(z_i) = 0, \quad z_i = (y_i - \theta)/\sigma, \tag{2.3.1}$$

σ is a scale parameter. In the first place, let us assume that σ is known. Arrange y_i $(1 \le i \le n)$ in ascending order of magnitude and let

$$y_{(1)} \le y_{(2)} \le \ldots\ldots \le y_{(n)} \tag{2.3.2}$$

be the resulting order statistics. A few basic properties of order statistics are enunciated in Appendix 1A (Chapter 1). Express (2.3.1) in terms of the order statistics $y_{(i)}$, $1 \le i \le n$. Since complete sums are invariant to ordering,

$$\frac{d \ln L}{d\theta} = \frac{1}{\sigma}\sum_{i=1}^{n} g(z_{(i)}) = 0, \ z_{(i)} = (y_{(i)} - \theta)/\sigma. \tag{2.3.3}$$

Let $t_{(i)} = E(z_{(i)})$ be the expected value of the ith standardized ordered variate $z_{(i)}$, $(1 \le i \le n)$. Expand $g(z_{(i)})$ as a Taylor series around $t_{(i)}$. Realizing that a function g(z) is almost linear in a small interval $a < z < b$ (Tiku, 1967; 1968) and $z_{(i)}$ is located in the vicinity of $t_{(i)}$ at any rate for large n, we obtain a linear approximation from the first two terms of a Taylor series. That gives

$$g(z_{(i)}) \cong g(t_{(i)}) + (z_{(i)} - t_{(i)}) \left\{\frac{d}{dz} g(z)\right\}_{z = t_{(i)}} \tag{2.3.4}$$
$$= \alpha_i + \beta_i z_{(i)}, \quad 1 \le i \le n,$$

where $\beta_i = \left\{\frac{d}{dz} g(z)\right\}_{z = t_{(i)}}$ and $\alpha_i = g(t_{(i)}) - \beta_i t_{(i)}$. (2.3.5)

If g(z) is bounded and $z_{(i)}$ tends to its expected value $t_{(i)}$,

$$g(z_{(i)}) - (\alpha_i + \beta_i z_{(i)}), \quad 1 \le i \le n, \tag{2.3.6}$$

tend to zero as n tends to infinity (see Appendix 2A). Incorporating (2.3.4) in (2.3.3) gives the modified likelihood equation

$$\frac{d \ln L}{d\theta} \cong \frac{d \ln L^*}{d\theta} = \frac{1}{\sigma} \sum_{i=1}^{n} \{\alpha_i + \beta_i z_{(i)}\} = 0. \tag{2.3.7}$$

It may be noted that under some very general regularity conditions (Appendix 2A)

$$\lim_{n \to \infty} \frac{1}{n} \left\{\frac{d \ln L^*}{d\theta} - \frac{d \ln L}{d\theta}\right\} = 0. \tag{2.3.8}$$

It follows from (2.3.6) and (2.3.8) that a modified likelihood equation is asymptotically equivalent to the corresponding likelihood equation.

Since (2.3.7) is linear in θ, it has an explicit and unique solution called the MML (modified maximum likelihood) estimator

$$\hat{\theta} = \left\{\sigma \sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \beta_i y_{(i)}\right\} / m, \qquad m = \sum_{i=1}^{n} \beta_i. \tag{2.3.9}$$

If the underlying distribution is symmetric, then $t_{(i)} = -t_{(n-i+1)}$. Consequently,

$$\alpha_i = -\alpha_{n-i+1}, \quad \sum_{i=1}^{n} \alpha_i = 0; \quad \beta_i = \beta_{n-i+1} \quad (1 \le i \le n). \tag{2.3.10}$$

For a symmetric distribution, therefore, the MML estimator of θ is a linear function of the order statistics:

$$\hat{\theta} = \sum_{i=1}^{n} \beta_i y_{(i)} / m \tag{2.3.11}$$

and is free of σ. Realize that the MML estimator $\hat{\theta}$ is asymptotically equivalent to the ML estimator and is, therefore, fully efficient (asymptotically). This follows from equations (2.3.6) and (2.3.8), but see Appendix 2A.

EXAMPLE 2.5: Consider estimating the location parameter μ in (2.2.9). As mentioned earlier, the ML estimator is elusive since d ln L/dμ = 0 has multiple roots (Vaughan, 1992a). To derive the MML estimator, we express d ln L/dμ in terms of the order statistics $y_{(i)}$ $(1 \le i \le n)$:

$$\frac{d \ln L}{d\theta} = \frac{2p}{k\sigma} \sum_{i=1}^{n} g(z_{(i)}) = 0, \quad g(z) = z/\{1 + (1/k)z^2\} \tag{2.3.12}$$

From the first two terms of a Taylor series expansion around $t_{(i)} = E(z_{(i)})$, we have

$$g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)}, \quad 1 \le i \le n, \tag{2.3.13}$$

where $$\alpha_i = \frac{(2/k)t_{(i)}^{\ 3}}{[1+(1/k)t_{(i)}^{\ 2}]^2} \quad \text{and} \quad \beta_i = \frac{1-(1/k)t_{(i)}^{\ 2}}{[1+(1/k)t_{(i)}^{\ 2}]^2}. \tag{2.3.14}$$

The values of $t_{(i)}$ are available in the following publications:

Tiku and Kumra (1981) for $p = 2(0.5)10$, $n \le 20$;

Vaughan (1992 b) for $p = 1.5$, $n \le 20$.

For p=1 (Cauchy distribution), Vaughan (1994) gives algebraic expressions for computing the expected values and the variances and covariances of the order statistics $y_{(i)}$, $3 \le i \le n - 3$ $(n \ge 6)$; the expected values of the first two (and the last two) order statistics are infinite.

Incorporating (2.3.13) in (2.3.12) gives the modified likelihood equation

$$\frac{d \ln L}{d\mu} \cong \frac{d \ln L^*}{d\mu} = \frac{2p}{k\sigma} \sum_{i=1}^{n} \{\alpha_i + \beta_i z_{(i)}\} = 0. \tag{2.3.15}$$

Since $\sum_{i=1}^{n} \alpha_i = 0$, we get the MML estimator

$$\hat{\mu} = \sum_{i=1}^{n} \beta_i y_{(i)} / m, \qquad m = \sum_{i=1}^{n} \beta_i . \tag{2.3.16}$$

Since the underlying distribution is symmetric, $E(\hat{\mu}) = \mu$. The exact variance of $\hat{\mu}$ is given by

$$V(\hat{\mu}) = (\beta' \Omega \beta)\sigma^2/m^2 \tag{2.3.17}$$

where $\beta = (\beta_1, \beta_2, \ldots \beta_n)$ and Ω is the variance-covariance matrix of the ordered variates $z_{(i)}$, $1 \le i \le n$.

Remark: The coefficients β_i $(1 \le i \le n)$ increase until the middle value and then decrease in a symmetric fashion, i.e., β_i have umbrella ordering. Thus, the extreme observations automatically receive small weights which depletes their effect. This is instrumental in achieving robustness to long-tailed symmetric distributions and to outliers in a sample (Chapter 8).

Asymptotic properties: The equation (2.3.15) can be reorganized to assume the form

$$\frac{d \ln L}{d\mu} \cong \frac{d \ln L^*}{d\mu} = \frac{2pm}{k\sigma^2}(\hat{\mu} - \mu). \tag{2.3.18}$$

Since $d \ln L^*/d\mu$ is asymptotically equivalent to $d \ln L/d\mu$, it follows that $\hat{\mu}$ is asymptotically the MVB estimator with variance σ^2/M $(M = 2pm/k)$ and is normally distributed. In other words, $\hat{\mu}$ is the BAN (best asymptotically normal) estimator.

The MVB for estimating μ is $(p \ge 2)$

$$\text{MVB}(\mu) = 1 \Big/ \left\{ -E\left(\frac{d^2 \ln L}{d\mu^2} \right) \right\} \tag{2.3.19}$$

$$= \frac{(p - 3/2)(p+1)}{np(p - 1/2)} \sigma^2;$$

this follows from the fact that

$$\int_{-\infty}^{\infty} \{1 + (1/k)z^2\}^{-j} \, dz = \sqrt{k}\,\Gamma(1/2)\Gamma(j - 1/2)/\Gamma(j) \quad (j \ge 1) \tag{2.3.20}$$

Now (Appendix 2A),

$$\lim_{n \to \infty} \frac{m}{n} = E\,[\{1 - (1/k)z^2\}/\{1 + (1/k)z^2\}] \qquad (2.3.21)$$
$$= (p - 1/2)/(p + 1);$$

follows from (2.3.20). Asymptotically, therefore, σ^2/M is exactly equal to the MVB (2.3.19). Thus, $\hat{\mu}$ is asymptotically fully efficient; see also Appendix 2A.

To have an idea about the efficiency of $\hat{\mu}$ for small sample sizes, we give in Table 2.1 the values of (a) 1/M, (b) $(1/\sigma^2)$ V $(\hat{\mu})$ calculated from (2.3.17), and (c) $(1/\sigma^2)$ MVB (μ) calculated from (2.3.19). It can be seen that the MML estimator $\hat{\mu}$ is highly efficient even for small sample sizes. Moreover, σ^2/M provides accurate approximations to the true values of the variance V($\hat{\mu}$). Realize that the variances and covariances of order statistics are not required in calculating M. The MML estimator being so highly efficient is one of its properties, besides being unique and explicit.

Table 2.1: Values of (a) 1/M, (b) $(1/\mu^2)$ V $(\hat{\mu})$ and (c)$(1/\sigma^2)$ MVB (μ).

n =	5	10	15	20	n =	5	10	15	20
		p = 2					p = 4		
(a)	0.106	0.052	0.034	0.025		0.169	0.088	0.059	0.044
(b)	0.120	0.055	0.036	0.026		0.190	0.092	0.061	0.045
(c)	0.100	0.050	0.033	0.025		0.179	0.089	0.060	0.045
		p = 7					p = 10		
(a)	0.184	0.094	0.064	0.048		0.189	0.096	0.065	0.049
(b)	0.197	0.098	0.065	0.049		0.199	0.099	0.066	0.049
(c)	0.193	0.097	0.064	0.048		0.197	0.098	0.066	0.049

Remark: The MVB estimator of μ does not exist for the family (2.2.9). Therefore, every estimator of μ will have its variance bigger than the MVB (2.3.19). It is indeed very pleasing to see that the variance of the MML estimator is, even for sample sizes as small as n=10, only marginally bigger than the MVB. We conclude that the MML estimator is highly efficient.

2.4 ESTIMATING LOCATION AND SCALE

Consider the problem of estimating both the location and scale parameters μ and σ. Given a random sample $y_1, y_2, \ldots, y_n$ from the distribution $(1/\sigma)\, f((y - \mu)/\sigma)$, the likelihood equations for estimating μ and σ, when expressed in terms of the order statistics $y_{(i)}$ $(1 \le i \le n)$, have typically the following forms:

$$\partial \ln L/\partial\mu = \sum_{i=1}^{n} g(z_{(i)}) = 0$$

and

$$\partial \ln L/\partial\sigma = -(n/\sigma) + \sum_{i=1}^{n} z_{(i)} g(z_{(i)}) = 0, \qquad (2.4.1)$$

$z_{(i)} = (y_{(i)} - \mu)/\sigma$. Replacing $g(z_{(i)})$ by the linear function $g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)}$, the modified likelihood equations have typically the following forms:

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n}\{\alpha_i + \beta_i z_{(i)}\} \qquad (2.4.2)$$

$$= \frac{m}{\sigma^2}(K + D\sigma - \mu) = 0 \text{ and}$$

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)}\{\alpha_i + \beta_i z_{(i)}\}$$

$$= -\frac{1}{\sigma^3}[(n\sigma^2 - B\sigma - C) - m(K - \mu)(K + D\sigma - \mu)] = 0 \qquad (2.4.3)$$

where $K = \left(\sum_{i=1}^{n}\beta_i y_{(i)}\right)/m$, $m = \sum_{i=1}^{n}\beta_i$, $D = \sum_{i=1}^{n}\alpha_i/m$; $B = \sum_{i=1}^{n}\alpha_i(y_{(i)} - K)$

and $C = \sum_{i=1}^{n}\beta_i(y_{(i)} - K)^2 = \sum_{i=1}^{n}\beta_i {y_{(i)}}^2 - mK^2$.

The solutions of these equations are the MML estimators:

$$\hat{\mu} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \{B + \sqrt{(B^2 + 4nC)}\}/2n; \qquad (2.4.4)$$

for skew distributions $D \neq 0$. Of course, the divisor 2n may be replaced by $2\sqrt{\{n(n-1)\}}$ (or some other appropriate constant) to reduce the bias if any. If $\beta_i \geq 0$ for all i = 1, 2,..., n, then $\hat{\sigma}$ is real and positive. Note that $\hat{\mu}$ is nonlinear if the underlying distribution is not symmetric.

Remark: It is interesting to note that all the modified likelihood equations $\partial \ln L^*/\partial\mu = 0$ and $\partial \ln L^*/\partial\sigma = 0$ in this chapter assume forms exactly similar to (2.4.2) and (2.4.3), respectively.

EXAMPLE 2.6: Consider estimating both μ and σ in (2.2.9). The modified likelihood equations are

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma}\sum_{i=1}^{n}\{\alpha_i + \beta_i z_{(i)}\} \qquad (2.4.5)$$

$$= \frac{M}{\sigma^2}(\hat{\mu} - \mu) = 0$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{n} z_{(i)}\{\alpha_i + \beta_i z_{(i)}\} \qquad (2.4.6)$$

$$= -\frac{1}{\sigma^3}[(n\sigma^2 - B\sigma - C) - M(\hat{\mu} - \mu)^2] = 0$$

where $M = \frac{2pm}{k}$, $B = \frac{2p}{k}\sum_{i=1}^{n}\alpha_i y_{(i)}$

and

$$C = \frac{2p}{k}\sum_{i=1}^{n}\beta_i(y_{(i)} - \hat{\mu})^2 = \frac{2p}{k}\left\{\sum_{i=1}^{n}\beta_i {y_{(i)}}^2 - m\hat{\mu}^2\right\} \qquad (2.4.7)$$

The constant coefficients α_i and β_i are given in (2.3.14). The solutions of the equations (2.4.5)–(2.4.6) are the MML estimators:

$$\hat{\mu} = \sum_{i=1}^{n} \beta_i y_{(i)}/m \quad \left(m = \sum_{i=1}^{n} \beta_i\right) \tag{2.4.8}$$

and

$$\hat{\sigma} = \{B + \sqrt{(B^2 + 4nC)}\}/2\sqrt{n(n-1)}.$$

We have already shown that $\hat{\mu}$ is asymptotically fully efficient. We have also shown that $\hat{\mu}$ is unbiased for all n, and is highly efficient.

The MVB for estimating σ is (follows from the very interesting result in Appendix 2A.10)

$$\text{MVB}(\sigma) = 1/\{-E(\partial^2 \ln L/\partial\sigma^2)\} = (p + 1)\,\sigma^2/\{2n(p - 1/2)\}. \tag{2.4.9}$$

The estimator $\hat{\sigma}$ is real and positive provided $\beta_i \geq 0$ for all i = 1, 2,..., n, and is highly efficient for all n. In fact, $\hat{\sigma}$ is asymptotically fully efficient since $\partial \ln L^*/\partial\sigma$ is equivalent to $\partial \ln L/\partial\sigma$ (Tiku and Suresh, 1992; Vaughan, 1992a). If $\beta_i < 0$ for some i, $\hat{\sigma}$ can cease to be real. Since $t_{(i)}$ is a decreasing sequence of values until the middle value and $|t_{(i)}| = |t_{(n-i+1)}|$, it follows that if β_1 is positive, then all the remaining β_i coefficients are positive. For small values of p ($\leq$ 3) and large n, however, a few β_i coefficients can be negative as a result of which $\hat{\sigma}$ can cease to be real and positive. In that case, the linear approximation (2.3.13) is recast as follows (Tiku et al., 2000):

Asymptotically $z_{(i)} - t_{(i)} \cong 0$ ($p \geq 2$). Hence,

$$\frac{1}{n}\sum_{i=1}^{n} h(t_{(i)})\,(z_{(i)} - t_{(i)}) \cong 0, \quad h(t) = t^2/\{1 + (1/k)t^2\}^2, \tag{2.4.10}$$

h(t) being a bounded function. For large n, therefore,

$$\frac{1}{n}\sum_{i=1}^{n} \{\alpha_i + \beta_i z_{(i)}\} \cong \frac{1}{n}\sum_{i=1}^{n} \{\alpha_i^* + \beta_i^* z_{(i)}\} \tag{2.4.11}$$

where (Tiku et al., 2000)

$$\alpha_i^* = 0 \quad \text{and} \quad \beta_i^* = 1/\{1 + (1/k)t_{(i)}^2\}. \tag{2.4.12}$$

If for a sample C in (2.4.8) assumes a negative value, the MML estimators are calculated from the sample with α_i and β_i replaced by α_i^* and β_i^* ($1 \leq i \leq n$), respectively. For p > 3, C assumes a negative value very rarely. The coefficients (2.4.12) are not used that often, therefore. Alternatively, the observations which correspond to negative values of β_i are censored and the MML estimators calculated from the resulting censored sample. The estimators so obtained are highly efficient (Vaughan, 1992a). This is discussed in Chapter 7.

It might be argued that the expected values $t_{(i)}$ ($1 \leq i \leq n$) are not available for n > 20. For $n \geq 10$, their approximate values obtained from the following equations may be used ($p \geq 2$),

$$\frac{1}{\sqrt{k}\,\beta(1/2,\, p - 1/2)} \int_{-\infty}^{t_{(i)}} \left\{1 + \frac{z^2}{k}\right\}^{-p} dz = \frac{i}{n+1}, \quad 1 \leq i \leq n. \tag{2.4.13}$$

The use of the approximate values for $n \geq 10$ does not affect the efficiency of the MML estimators (2.4.8) in any substantial way. To evaluate (2.4.13), an IMSL subroutine is available. Realize that $t = \sqrt{(\nu/k)}z$ has the Student t distribution with $\nu = 2p - 1$ degrees of freedom.

Remark: We show in Appendix 2B that the MML estimates $\hat{\mu}$ and $\hat{\sigma}$ when substituted in the likelihood equations (2.4.1) make them virtually equal to zero even for small sample sizes. The MML and the ML estimates are , therefore, numerically the same (almost). Tan (1985) has similar results.

2.5 GENERALIZED LOGISTIC

Consider the family of generalized logistic distributions (b > 0)

$$GL(b, \sigma) : f(y) = \frac{b}{\sigma} \frac{\exp\{-(y-\mu)/\sigma\}}{[1+\exp\{-(y-\mu)/\sigma\}]^{b+1}}, \quad -\infty < y < \infty. \tag{2.5.1}$$

For b < 1, b = 1 and b > 1, (2.5.1) represents negatively skew, symmetric and positively skew distributions, respectively. For b = 1, in fact, (2.5.1) is the well-known logistic distribution which has been used extensively in many areas of application (Agresti, 1996; Berkson, 1951; Pearl, 1940; Perks, 1932; Reed and Berkson, 1929). Numerous biomedical and industrial applications of the logistic distribution are given in Chapter 4.

Given a random sample $y_1, y_2, ..., y_n$, the likelihood equations for estimating μ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma} \sum_{i=1}^{n} g(z_i) = 0 \tag{2.5.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} z_i - \frac{(b+1)}{\sigma} \sum_{i=1}^{n} z_i g(z_i) = 0; \tag{2.5.3}$$

$z = (y-\mu)/\sigma$ and $g(z) = e^{-z}/(1+e^{-z}) = 1/(1+e^z)$. The equations have no explicit solutions.

To obtain the modified likelihood equations, we first express (2.5.2) – (2.5.3) in terms of the ordered variates $z_{(i)} = (y_{(i)} - \mu)/\sigma$ $(1 \le i \le n)$, simply by replacing z_i by $z_{(i)}$. We then linearize $g(z_{(i)})$:

$$g(z_{(i)}) \cong \alpha_i - \beta_i z_{(i)}, \quad 1 \le i \le n; \tag{2.5.4}$$

$$\alpha_i = (1+e^t+te^t)/(1+e^t)^2 \quad \text{and} \quad \beta_i = e^t/(1+e^t)^2, \quad t = t_{(i)}; \tag{2.5.5}$$

$t_{(i)} = E(z_{(i)})$. Values of $t_{(i)}$ for $n \le 15$ are available in Balakrishnan and Leung (1988). For $n \ge 10$, however, the approximate values of $t_{(i)}$ may be used, namely,

$$t_{(i)} = -\ln(q_i^{-1/b} - 1), \; q_i = i/(n+1), \tag{2.5.6}$$

which are the solutions of

$$\int_{-\infty}^{t_{(i)}} f(z)\,dz = q_i \quad (1 \le i \le n) \tag{2.5.7}$$

The modified likelihood equations are

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma} \sum_{i=1}^{n} (\alpha_i - \beta_i z_{(i)}) \tag{2.5.8}$$

$$= \frac{M}{\sigma^2}(K + D\sigma - \mu) = 0$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} z_{(i)} - \frac{(b+1)}{\sigma} \sum_{i=1}^{n} z_{(i)}(\alpha_i - \beta_i z_{(i)}) \tag{2.5.9}$$

$$= -\frac{1}{\sigma^3}\,[(n\sigma^2 - B\sigma - C) - M(K-\mu)(K + D\sigma - \mu)] = 0$$

where $M = (b+1)m,\ K = \left(\sum_{i=1}^{n} \beta_i y_{(i)}\right)/m \qquad \left(m = \sum_{i=1}^{n} \beta_i\right),$

$$D = \sum_{i=1}^{n} \Delta_i/m,\ \ \Delta_i = (b+1)^{-1} - \alpha_i,\quad B = (b+1)\sum_{i=1}^{n} \Delta_i\,(y_{(i)} - K) \qquad (2.5.10)$$

and $C = (b+1)\left(\sum_{i=1}^{n} \beta_i y_{(i)}^{\,2} - mK^2\right).$

The solutions of these equations are unique and explicit, the MML estimators:

$$\hat{\mu} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \{B + \sqrt{(B^2 + 4nC)}\}/2\sqrt{\{n(n-1)\}}\,. \qquad (2.5.11)$$

Realize that $\hat{\sigma}$ is always real and positive since $\beta_i > 0$ $(1 \le i \le n)$. For b = 1 (logistic distribution), D = 0 (follows from symmetry) in which case $\hat{\mu} = K$ is a linear function of order statistics and is free of σ (and $\hat{\sigma}$).

Comment: We show in Appendix 2B that the MML estimators (2.5.11) make the likelihood equations (2.5.2) – (2.5.3) virtually equal to zero even for small sample sizes. This establishes the fact that the ML and the MML estimates are numerically the same (almost). From numerous calculations, Tiku (1967a, p.164), Tiku (1968a, p.137), Tan (1985), Tiku and Vaughan (1997, pp. 890-892), Tiku et al., (1986, pp.101, 106-107) and Vaughan (2002) arrive at the same conclusion. Also, the MML estimators are asymptotically fully efficient and are highly efficient for small sample sizes. Smith et al. (1973), Lee et al. (1980), Tan (1985), Vaughan (1992a, 1994) and Senoglu and Tiku (2001) have similar results. There is, therefore, no reason whatsoever for not using the MML estimators in place of the elusive ML estimators.

Covariance matrix: The asymptotic variances and the covariance of $\hat{\mu}$ and $\hat{\sigma}$ are given by $I^{-1}(\mu, \sigma)$ where I is the Fisher information matrix:

$$I(\mu,\sigma) = \frac{n}{\sigma^2} \times \begin{bmatrix} \dfrac{b}{(b+2} & \dfrac{b\{\psi(b+1) - \psi(2)\}}{(b+2)} \\ \dfrac{b\{\psi(b+1) - \psi(2)\}}{(b+2)} & 1 + \dfrac{b\{[\psi'(b+1) + \psi'(2)] + [\psi(b+1) - \psi(2)]^2\}}{(b+2)} \end{bmatrix}; \qquad (2.5.12)$$

$\psi(x) = \Gamma'(x)/\Gamma(x)$ is the psi-function and $\psi'(x)$ is the derivative of $\psi(x)$ with respect to x. The values of $\psi(x)$ and its derivative $\psi'(x)$, and other details for working out the elements of the matrix in (2.5.12), are given in Appendix 2D.

It is not difficult to invert I and we write

$$V = I^{-1}(\mu,\sigma) = \frac{\sigma^2}{n}\begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix}, \quad (v_{12} = v_{21}). \qquad (2.5.13)$$

For b=1 (logistic distribution) $v_{12} = v_{21} = 0$ in which case v_{11} and v_{22} are the minimum variance bounds, that is,

$$MVB(\mu) = 3\sigma^2/n \qquad (2.5.14)$$

and $MVB(\sigma) = 1.56\sigma^2/2n \quad (\text{since } \psi'(2) = 0.4228).$

The MML estimators are remarkably efficient. For b = 1, the values of the minimum variance bounds and the exact variance of $\hat{\mu}$ and the simulated variance of $\hat{\sigma}$ are given below; $\hat{\mu}$ is unbiased because of symmetry and the simulated values of $E(\hat{\sigma})$ are given below ($\sigma = 1$ without loss of generality). Also given are the values of the relative efficiency $E = 100\{V(\hat{\mu})/V(\bar{y})\}$ of the sample mean. It may be noted that E decreases with increasing n.

	n = 10	n = 20		n = 10	n = 20
MVB(μ)	0.300	0.150	$E(\hat{\sigma})$	1.027	1.019
$V(\hat{\mu})$	0.310	0.152	MVB($\hat{\sigma}$)	0.078	0.039
E	94.2	92.4	$V(\hat{\sigma})$	0.081	0.039

The asymptotic variances of $\hat{\mu}$ and $\hat{\sigma}$ are given in Table 2.2 for b = 0.5, 2, 4 and 6 (the estimators are asymptotically unbiased). Also given are the simulated means and variances of $\hat{\mu}$ and $\hat{\sigma}$. It is clear that the MML estimators have negligible bias and are highly efficient. We reiterate, there is hardly any loss of precision involved in using the MML estimators in place of the elusive ML estimators.

Table 2.2: Means and variances of the MML estimators for GL(b, σ): $\mu = 0$ and $\sigma = 1$.

		n = 10				n = 20			
		b = 0.5	2	4	6	b = 0.5	2	4	6
Asymp:	$nV(\hat{\mu})$	5.11	2.16	2.24	2.65	5.11	2.16	2.24	2.65
	$nV(\hat{\sigma})$	0.76	0.66	0.63	0.63	0.76	0.66	0.63	0.63
Simul:	$E(\hat{\mu})$	– 0.025	0.026	0.068	0.086	– 0.017	0.009	0.003	0.048
	$E(\hat{\sigma})$	1.032	1.020	1.001	0.997	1.014	1.014	1.006	0.994
	$nV(\hat{\mu})$	5.46	2.25	2.34	2.83	5.30	2.18	2.28	2.66
	$nV(\hat{\sigma})$	0.89	0.79	0.73	0.71	0.82	0.72	0.68	0.66

All the simulated values given in this monograph are based on [100,000/n] (integer value) Monte Carlo runs and have standard errors well within ±0.005.

2.6 EXTREME VALUE DISTRIBUTION

Consider the extreme-value distribution

$$EV(\delta, \eta) : f(y) = \frac{1}{\eta}\exp\left[-\left(\frac{y-\delta}{\eta}\right) - \exp\left(-\frac{y-\delta}{\eta}\right)\right], \quad -\infty < y < \infty, \qquad (2.6.1)$$

which is of major importance for many of its applications. This distribution is also important due to the fact that if U has the Weibull distribution

$$f(u) = \frac{p}{\sigma}\left(\frac{u}{\sigma}\right)^{p-1} \exp\left\{-\left(\frac{u}{\sigma}\right)^{p}\right\}, \quad 0 < u < \infty, \tag{2.6.2}$$

then Y = ln U has the extreme-value distribution (2.6.1) with $\delta = \ln \sigma$ and $\eta = 1/p$.

Given a random sample $y_1, y_2, \ldots, y_n$ from (2.6.1), we want to estimate δ and η. The likelihood equations expressed in terms of the ordered variates $z_{(i)} = (y_{(i)} - \delta)/\eta$ $(1 \le i \le n)$ are

$$\frac{\partial \ln L}{\partial \delta} = \frac{n}{\eta} - \frac{1}{\eta}\sum_{i=1}^{n} \exp(-z_{(i)}) = 0 \tag{2.6.3}$$

and

$$\frac{\partial \ln L}{\partial \delta} = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} z_{(i)} - \frac{1}{\eta}\sum_{i=1}^{n} z_{(i)} \exp(-z_{(i)}) = 0. \tag{2.6.4}$$

The ML estimators of δ and η are elusive since (2.6.3) – (2.6.4) do not have explicit solutions. Here, $g(z) = e^{-z}$ and we have

$$e^{-z_{(i)}} \cong e^{-t_{(i)}} + (z_{(i)} - t_{(i)})\left\{\frac{d}{dz} e^{-z}\right\} \tag{2.6.5}$$

$$z = t_{(i)} = \alpha_i - \beta_i z_{(i)}, \quad 1 \le i \le n,$$

where $\beta_i = -\left\{\frac{d}{dz} e^{-z}\right\}_{z = t_{(i)}} = e^{-t_{(i)}}$ and $\alpha_i = e^{-t_{(i)}}(1 + t_{(i)})$. (2.6.6)

Note that β_i $(1 \le i \le n)$ is a decreasing sequence of positive values, i.e., β_i have half-umbrella ordering.

The expected values $t_{(i)}$ are given by (Lieblien, 1953; White, 1969)

$$t_{(i)} = 0.57722 + i\binom{n}{i}\sum_{j=0}^{n-i}\binom{n-i}{j}(-1)^j \frac{\ln(i+j)}{i+j}, \tag{2.6.7}$$

$0.57722 \cong \int_0^{\infty} (\ln u) \exp(-u)\, du$ being the Euler constant. For $n \ge 10$, the approximate values of $t_{(i)}$ obtained from the equation

$$\int_{-\infty}^{t_{(i)}} \exp\{-z - e^{-z}\}\, dz = \frac{i}{n+1} \tag{2.6.8}$$

may be used; (2.6.8) gives

$$t_{(i)} = -\ln[-\ln\{i/(n+1)\}], \quad 1 \le i \le n. \tag{2.6.9}$$

The modified likelihood equations are obtained by incorporating (2.6.5) in (2.6.3) – (2.6.4) and have forms exactly similar to (2.4.2) – (2.4.3). The solutions of these equations are the MML estimators:

$$\hat{\delta} = K + D\hat{\eta} \quad \text{and} \quad \hat{\eta} = \{B + \sqrt{(B^2 + 4nC)}\}/2n; \tag{2.6.10}$$

$$K = \sum_{i=1}^{n} \beta_i y_{(i)}/m, \quad m = \sum_{i=1}^{n} \beta_i, \quad D = \sum_{i=1}^{n} (1 - \alpha_i)/m;$$

$$B = \sum_{i=1}^{n} (1 - \alpha_i)(y_{(i)} - K) \quad \text{and} \quad C = \sum_{i=1}^{n} \beta_i (y_{(i)} - K)^2 = \sum_{i=1}^{n} \beta_i y_{(i)}^2 - mK^2.$$

There is a little bias in $\hat{\delta}$ and $\hat{\eta}$ for $n \le 15$ but that is corrected by replacing D by $-(1/m)\sum_{i=1}^{n} \beta_i t_{(i)}$ and the divisor 2n in (2.6.10) by 2m. For large n, $m/n \cong 1$, and the modified estimator $\hat{\delta}$ is equivalent to that in (2.6.10) (Vaughan and Tiku, 2000, p. 59).

Asymptotically, the estimators $\hat{\delta}$ and $\hat{\eta}$ are unbiased and their variance-covariance matrix is

$$\frac{6\eta^2}{\pi^2 n}\begin{bmatrix} (c-1)^2 + \dfrac{\pi^2}{6} & 1-c \\ 1-c & 1 \end{bmatrix}, \quad c \cong 0.57722, \qquad (2.6.11)$$

obtained by inverting $I(\delta, \eta)$, I being the Fisher information matrix (Vaughan and Tiku, 2000); see Appendix 2A.

Given in Table 2.3 are the simulated values of the variances, and the corresponding MVB; the bias in the estimators is negligible (Vaughan and Tiku, 2000, p. 59) and is not, therefore, reported:

Table 2.3: Values of (1) $(1/\eta^2)\, V(\hat{\delta})$, (2) $(1/\eta^2)\, MVB(\hat{\delta})$; (3) $(1/\eta^2)\, V(\hat{\eta})$ and (4) $(1/\eta^2)\, MVB(\hat{\eta})$.

n	(1)	(2)	(3)	(4)	n	(1)	(2)	(3)	(4)
5	0.229	0.222	0.170	0.122	20	0.057	0.055	0.035	0.030
10	0.115	0.111	0.077	0.061	25	0.046	0.044	0.027	0.024
15	0.076	0.074	0.049	0.041	30	0.037	0.037	0.022	0.022

It can be seen that the MML estimators are remarkably efficient. As said earlier, this is a typical feature of the MML estimators. Moreover, $\hat{\mu}$ and $\hat{\sigma}$ are unique and explicit. See also Smith et al. (1973) who have similar results for a log-normal distribution. In fact, they add a quadratic term to the linear approximation (2.3.4) but that does not improve the efficiencies in any substantial way.

Remark: We show later in Chapter 8 that the Huber (1964, 1981) M-estimators (Section 2.11) of μ are remarkably eficient for long-tailed symmetric distributions such as (2.2.9) but not for skew and short-tailed symmetric distributions. For skew distributions, they develop substantial bias. For short-tailed symmetric distributions, they are inefficient. The M-estimators of $\hat{\sigma}$, however, can have substantial downward bias even for long-tailed symmetric distributions (Chapter 8).

We now consider another very useful method of estimating a location and a scale parameter as follows:

2.7 BEST LINEAR UNBIASED ESTIMATORS

Let $y_{(1)} \le y_{(2)} \le \ldots\ldots \le y_{(n)}$ be the order statistics of a random sample of size n from a location-scale distribution $(1/\sigma)\, f((y-\mu)/\sigma)$. Let $t_{(i)} = E(z_{(i)})$, $\sigma_{ii:n} = V(z_{(i)})$ and $Cov\,(z_{(i)}, z_{(j)}) = \sigma_{ij:n}$ be the expected values and variances and covariances of the standardized variates $z_{(i)} = (y_{(i)} - \mu)/\sigma$, $1 \le i \le n$. They are available for numerous distributions, mostly for $n \le 20$. Since $E(y_{(i)}) = \mu + \sigma t_{(i)}$ $(1 \le i \le n)$, we formulate it as a linear model:

$$Y = W\theta + e \tag{2.7.1}$$

where $Y = \begin{pmatrix} y_{(1)} \\ y_{(2)} \\ \cdot \\ \cdot \\ \cdot \\ y_{(n)} \end{pmatrix}$, $W = \begin{pmatrix} 1 & t_{(1)} \\ 1 & t_{(2)} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & t_{(n)} \end{pmatrix}$, $\theta = \begin{pmatrix} \mu \\ \sigma \end{pmatrix}$ and $e = \begin{pmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{pmatrix}$; (2.7.2)

the errors e_i are not independently distributed. In fact, their variance-covariance matrix is

$$\Omega = (\sigma_{ij:n})_{i,\, j=1,\, 2, \ldots, n} \tag{2.7.3}$$

To obtain the BLUE (best linear unbiased estimators) of μ and σ, we minimize the generalized error variance (Lloyd, 1952)

$$e' \Omega^{-1} e = (Y - W\theta)' \Omega^{-1} (Y - W\theta) \tag{2.7.4}$$

This gives the BLUE θ^*:

$$\theta^* = \begin{pmatrix} \mu^* \\ \sigma^* \end{pmatrix} = (W' \Omega^{-1} W)^{-1} W' \Omega^{-1} Y. \tag{2.7.5}$$

The right hand side yields two linear functions

$$\mu^* = \sum_{i=1}^{n} a_i y_{(i)} \quad \text{and} \quad \sigma^* = \sum_{i=1}^{n} b_i y_{(i)}; \tag{2.7.6}$$

a_i and b_i are constant coefficients which work in terms of the expected values and the variances and covariances of the standardized variates $z_{(i)}$, $1 \le i \le n$.

It is easy to show that $E(\theta^*) = \theta$ so that μ^* and σ^* are both unbiased, and the variance-covariance matrix V of μ^* and σ^* is

$$V = (W' \Omega^{-1} W)^{-1} \sigma^2. \tag{2.7.7}$$

If the underlying distribution is symmetric, then $(i \le j)$

$$t_{(i)} = -t_{(n-i+1)} \quad \text{and} \quad \sigma_{ij:n} = \sigma_{n-j+1,\, n-i+1:n} \quad \text{(double symmetry)}, \tag{2.7.8}$$

as said earlier. Consequently, V is a diagonal matrix and $\text{Cov}(\mu^*, \sigma^*) = 0$.

Among linear functions, BLUE are the ideal estimators. The difficulty is the availability of the expected values and the variances and covariances of order statistics, and the laborious computations involved in calculating the coefficients a_i and b_i, $1 \le i \le n$. Small sample distribution theory of BLUE is also unavailable other than for a few distributions, for example, exponential; see the monumental book by David (1981).

For a general class of linear estimators (called L-estimators) of μ and σ, see Chernoff and Gastworth (1967) and Schneider (1986, p.89); L-estimators are asymptotically normally distributed.

To compare the efficiencies of the MML estimator $\hat{\mu}$ (whenever it is a linear function of order statistics) and the BLUE μ^*, consider the family of distributions (2.2.9). Like μ^*, the MML estimator $\hat{\mu}$ is unbiased for all n. The exact variances of $\hat{\mu}$ are given in Table 2.1. Given below are the corresponding values for the BLUE μ^*;

Exact values of $(1/\sigma^2)\, V(\mu^*)$

n = 5	10	15	20	n = 5	10	15	20
	p = 2				p = 4		
0.120	0.055	0.035	0.026	0.189	0.092	0.061	0.045
	p = 7				p = 10		
0.197	0.098	0.065	0.049	0.199	0.099	0.066	0.049

It is seen that $\hat{\mu}$ is essentially as efficient as μ^* and that is typical of the MML estimator of a location parameter (whenever it is a linear function of order statistics). In general, $\hat{\mu}$ assumes the form $\hat{\mu} = K + D\,\hat{\sigma}$ and is somewhat more efficient than μ^*. It may be remembered that in computing $\hat{\mu}$, only the expected values of order statistics are required and not their variances and covariances.

The MML estimator $\hat{\sigma}$ is generally of the form (2.4.4) and is nonlinear. In fact, it is somewhat more efficient than the BLUE σ^*. For large n, $\hat{\sigma}$ is unbiased. For small n, $\hat{\sigma}$ may have a little bias which in most situations can be reduced by replacing the divisor n by $\sqrt{\{n(n-1)\}}$ or some other appropriate constant.

2.8 NUMERICAL EXAMPLES

We now consider a few numerical examples to illustrate the computations of the MML estimators.

EXAMPLE 2.8: The following data represent 100 times the white blood counts of patients who had acute myelogenous leukemia (Gross and Clark, 1975):

23 7.5 43 26 60 105 100 170 54 70 94 320 350 1000 520 1000

The underlying distribution is presumed to be the Weibull (Chapter 9)

$$W(\theta, \sigma) : \frac{p}{\sigma}\left(\frac{u-\theta}{\sigma}\right)^{p-1} \exp\left\{-\left(\frac{u-\theta}{\sigma}\right)^{p}\right\}, \quad \theta < y < \infty, \tag{2.8.1}$$

with p not far from 1.2. To formulate the MML estimators of θ and σ, the likelihood equations expressed in terms of the ordered variates $z_{(i)} = (u_{(i)} - \theta)/\sigma$ $(1 \le i \le n)$ are

$$\frac{\partial \ln L}{\partial \theta} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{(i)}^{p-1} = 0 \tag{2.8.2}$$

and

$$\frac{\partial \ln L}{\partial \theta} = -\frac{n}{p} - \frac{p-1}{\sigma}\sum_{i=1}^{n} z_{(i)} z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{(i)}^{p} = 0 \tag{2.8.3}$$

The equation (2.8.3) simplifies:

$$\frac{\partial \ln L}{\partial \theta} = -\frac{np}{p} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{(i)}^{p} = 0 \tag{2.8.4}$$

The equation (2.8.2) is not defined if $z_{(1)}$ tends to zero which is more likely to happen when n becomes large. The simplified equation (2.8.4) gives the ML estimator (for given θ)

$$\hat{\sigma} = \left[\sum_{i=1}^{n} (u_{(i)} - \theta)^p / n\right]^{1/p} \tag{2.8.5}$$

In practice θ is not known and has to be replaced by an estimate $\tilde{\theta}$, and $\tilde{\theta}$ should not be greater than the smallest order statistic $u_{(1)}$ for $\hat{\sigma}$ to be real and positive for all p. In fact, θ is often replaced by $u_{(1)}$. But $u_{(1)}$ is not necessarily the solution of (2.8.2). There is no such difficulty with modified likelihood.

To obtain the MML estimators, we note that (Harter, 1964)

$$E\{z_{(i)}\} = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{i-1} \frac{(-1)^j \binom{i-1}{j} \Gamma\left(1+\frac{1}{p}\right)}{(n-i+j+1)^{(p+1)/p}}. \tag{2.8.6}$$

The approximate values of $t_{(i)}$ are given by

$$\int_0^{t_{(i)}} p z^{p-1} \exp(-z^p)\, dz = \frac{i}{n+1},$$

i.e.,

$$t_{(i)} = \left[-\ln\left\{1 - \frac{i}{n+1}\right\}\right]^{1/p}, \quad 1 \le i \le n; \tag{2.8.7}$$

these values may be used for all $n \ge 10$ without any serious detrimental effect on the efficiencies of the MML estimators given below.

To obtain the modified likelihood equations, we linearize $z_{(i)}{}^{p-1}$ and $z_{(i)}{}^{-1}$ by using the first two terms of a Taylor series expansion around $t_{(i)}$. This gives (Islam et al., 2001)

$$z_{(i)}{}^{p-1} \cong \alpha_{(i)} + \beta_i z_{(i)}; \quad \alpha_i = (2-p)t_{(i)}{}^{p-1} \text{ and} \tag{2.8.8}$$

$$\beta_i = (p-1)t_{(i)}{}^{p-2} \qquad (1 \le i \le n);$$

and, similarly, $z_{(i)}{}^{-1} \cong \alpha_{i0} - \beta_{i0} z_{(i)}$; $\alpha_{i0} = 2t_{(i)}{}^{-1}$ and $\beta_{i0} = t_{(i)}{}^{-2}$ $(1 \le i \le n)$. (2.8.9)

Realize that $\beta_i > 0$ for $p > 1$, and $\beta_{i0} > 0$ $(1 \le i \le n)$.

Incorporating (2.8.8) – (2.8.9) in (2.8.2) – (2.8.3) gives the modified likelihood equations:

$$\frac{\partial \ln L}{\partial \theta} \cong \frac{\partial \ln L^*}{\partial \theta} = -\frac{p-1}{\sigma} \sum_{i=1}^{n} \{\alpha_{i0} - \beta_{i0} z_{(i)}\} + \frac{p}{\sigma} \sum_{i=1}^{n} \{\alpha_i + \beta_i z_{(i)}\} \tag{2.8.10}$$

$$= \frac{m}{\sigma^2}(K + D\sigma - \mu) = 0$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} - \frac{p-1}{\sigma} \sum_{i=1}^{n} z_{(i)}\{\alpha_i - \beta_i z_{(i)}\} + \frac{p}{\sigma} \sum_{i=1}^{n} z_{(i)}\{\alpha_i + \beta_i z_{(i)}\} \tag{2.8.11}$$

$$= -\frac{1}{\sigma^3}[(n\sigma^2 - B\sigma - C) - m(K-\mu)(K + D\sigma - \mu)] = 0$$

where $\delta_i = (p-1)\beta_{i0} + p\beta_i$, $m = \sum_{i=1}^{n} \delta_i$; $\Delta_i = p\alpha_i - (p-1)\alpha_{i0}$,

$$D = \sum_{i=1}^{n} \Delta_i / m, \quad K = \left(\sum_{i=1}^{n} \delta_i u_{(i)}\right)/m, \quad B = \sum_{i=1}^{n} \Delta_i (u_{(i)} - K) \tag{2.8.12}$$

and $$C = \sum_{i=1}^{n} \delta_i (u_{(i)} - K)^2 = \sum_{i=1}^{n} \delta_i u_{(i)}{}^2 - mK^2.$$

The solutions of these equations are the MML estimators:

$$\hat{\theta} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \{B + \sqrt{(B^2 + 4nC)}\} / 2\sqrt{\{n(n-1)\}}. \tag{2.8.13}$$

Note that $\hat{\sigma}$ is always real and positive since $\delta_i > 0$ $(1 \le i \le n)$ for $p > 1$. In practice, of course, p is often greater than 1 (Cohen and Whitten, 1982).

The Fisher information matrix exists for $p > 2$ and is given by

$$I(\theta, \sigma) = \frac{np^2}{\sigma^2} \begin{bmatrix} \left(1 - \frac{1}{p}\right)^2 \Gamma\left(1 - \frac{2}{p}\right) & \Gamma\left(2 - \frac{1}{p}\right) \\ \Gamma\left(2 - \frac{1}{p}\right) & 1 \end{bmatrix}. \tag{2.8.14}$$

For $p > 2$, the MML estimators $\hat{\theta}$ and $\hat{\sigma}$ above are asymptotically fully efficient, i.e., they are unbiased and their variance-covariance matrix is given by I^{-1}.

The standard errors of $\hat{\theta}$ and $\hat{\sigma}$ are approximately $(p > 2)$

$$\pm \frac{\hat{\sigma}}{p\sqrt{n}} \left[\frac{1}{\left(1 - \frac{1}{p}\right)^2 \Gamma\left(1 - \frac{2}{p}\right) - \left\{\Gamma\left(2 - \frac{1}{p}\right)\right\}^2} \right]^{1/2}$$

and $$\pm \frac{\hat{\sigma}}{p\sqrt{n}} \left[\frac{\left(1 - \frac{1}{p}\right)^2 \Gamma\left(1 - \frac{2}{p}\right)}{\left(1 - \frac{1}{p}\right)^2 \Gamma\left(1 - \frac{2}{p}\right) - \left\{\Gamma\left(2 - \frac{1}{p}\right)\right\}^2} \right]^{1/2}, \tag{2.8.15}$$

respectively.

For the data above, we have the following MML estimates of θ and σ:

$$\hat{\theta} = 5.98 \quad \text{and} \quad \hat{\sigma} = 202.$$

Since $p = 1.2$ and is less than 2, (2.8.14) – (2.8.15) cannot be used to calculate the approximate standard errors in these estimates; see, however, Example 2.10. If $p \le 2$, we recommend the transformation $Y = \ln U$. The distribution of Y is the extreme-value distribution. The estimation can then proceed. We illustrate this in Example 2.10.

EXAMPLE 2.9: The following 15 observations represent the measurements of a certain characteristic in blood cells (Elveback et al., 1970):

8.921	8.982	9.048	9.262	9.689	9.715	9.774	9.830
10.128	10.485	10.591	10.766	10.840	10.881	11.263.	

Elveback et al. assume a logistic distribution. To calculate the BLUE, the coefficients a_i and b_i in (2.7.6) are available (Gupta et al., 1967). To calculate the MML estimates, we only need the expected values of order statistics which we calculate from (2.5.6) with b = 1. From equations (2.5.10) and (2.7.6), we obtain the following:

	Estimate		$(1/\sigma^2)$ Variance	
	μ	σ	μ	σ
BLUE	10.010	0.843	0.063	0.048
MML	10.012	0.831	0.062	0.039

Incidentally, the ML estimates obtained by a laborious iterative process are 10.011 and $0.801\sqrt{15/14} = 0.829$ for μ and σ, respectively. There is close agreement between the three methods. The modified likelihood method is analytically and computationally the easiest, however.

EXAMPLE 2.10: Consider the following data obtained by taking the logarithm of the observations in Example 2.8:

3.135 2.015 3.761 3.258 4.094 4.654 4.605 5.136 3.989 4.248 4.543
5.768 5.858 6.908 6.254 6.908

The distribution of Y = ln U is assumed to be the extreme-value distribution (2.6.1).

From equations (2.6.10), we obtain the following estimates:

$$\hat{\delta} = 4.073 \quad \text{and} \quad \hat{\eta} = 1.248 \qquad (2.8.16)$$

The approximate standard errors in these estimates obtained from (2.6.11) are

$$\pm\left[\frac{6(1.248)^2}{(3.143)^2(16)}(0.1787 + 1.6464)\right]^{1/2} = \pm\,0.328 \quad \text{and} \quad \pm\,\frac{6(1.248)^2}{(3.143)^2(16)} = \pm\,0.243,$$

respectively. The standard errors are pleasantly small which indicates a high degree of precision in the estimates.

We now discuss the asymptotic distributions of the MML estimators. We will show in Section 2.10 (and Chapter 8) that the results are applicable even for small sample sizes.

2.9 ASYMPTOTIC DISTRIBUTIONS OF THE MML ESTIMATORS

We assume that the range of the random variable y does not depend on the parameter(s) we are estimating. To obtain the distributions of $\sqrt{m}(\hat{\mu} - \mu)/\sigma$ and $(n-1)\hat{\sigma}^2/\sigma^2$, we first note the following important results:

Let $y_1, y_2, \ldots\ldots y_n$ be a random sample of size n from a normal population $N(\mu, \sigma^2)$. Here,

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma^2}(\bar{y} - \mu) \quad \text{and} \quad \frac{\partial \ln L}{\partial \sigma} = \frac{n}{\sigma^3}(S^2 - \sigma^2), \quad S^2 = \sum_{i=1}^{n}(y_i - \mu)^2/n. \qquad (2.9.1)$$

Bartlett (1953) showed that (i) the cumulants of $\partial \ln L/\partial\mu$ are determined by $E(\partial^r \ln L/\partial\mu^r)$ which are zero for all $r \geq 3$ and that gives the distribution of $\sqrt{n}(\bar{y} - \mu)/\sigma$ as normal N(0, 1), (ii) the cumulants of $\partial \ln L/\partial\sigma$ are determined by $E(\partial^r \ln L/\partial\sigma^r)$ $(r \geq 1)$ which gives the distribution of nS^2/σ^2 as chi-square with n degrees of freedom, and (iii) the mixed cumulants of the random variables $\partial \ln L/\partial\mu$ and $\partial \ln L/\partial\sigma$ are determined by $E(\partial^{r+s} \ln L/\partial\mu^r \partial\sigma^s)$ which are zero for all $r \geq 1$ and $s \geq 1$, together with the Cochran identity

$$nS^2 \equiv n(\bar{y} - \mu)^2 + (n-1)s^2, \quad s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1), \tag{2.9.2}$$

gives the result that $\bar{y}$ and s^2 (or s) are independently distributed, and the distribution of $(n-1)s^2/\sigma^2$ is chi-square with n – 1 degress of freedom; for a beautiful geometrical proof of these results, see Patnaik (1949).

Consider the situation when the underlying distribution is symmetric. Here, the modified likelihood equations are

$$\frac{\partial \ln L}{\partial\mu} \cong \frac{\partial \ln L^*}{\partial\mu} = \frac{m}{\sigma^2}(\hat{\mu} - \mu) \tag{2.9.3}$$

and

$$\frac{\partial \ln L}{\partial\sigma} \cong \frac{\partial \ln L^*}{\partial\sigma}$$

$$= -\frac{n}{\sigma^3}\left(\sigma - \frac{B_0 + \sqrt{B_0^2 + 4nC_0}}{n}\right)\left(\sigma - \frac{B_0 - \sqrt{B_0^2 + 4nC_0}}{n}\right) \tag{2.9.4}$$

$$\cong \frac{n}{\sigma^3}\left(\frac{C_0}{n} - \sigma^2\right)$$

since $B_0/\sqrt{(nC_0)} \cong 0$, α_i and β_i being bounded;

$$B_0 = \sum_{i=1}^{n} \alpha_i (y_{(i)} - \mu) \quad \text{and} \quad C_0 = \sum_{i=1}^{n} \beta_i (y_{(i)} - \mu)^2 \quad (\beta_i > 0).$$

Moreover, $$\sum_{i=1}^{n} \beta_i (y_{(i)} - \mu)^2 \equiv m(\hat{\mu} - \mu)^2 + \sum_{i=1}^{n} \beta_i (y_{(i)} - \hat{\mu})^2, \quad m = \sum_{i=1}^{n} \beta_i, \tag{2.9.5}$$

and $\hat{\mu}$ is linear and is normally distributed (asymptotically). Realizing that the structural relationships here are exactly the same as in (2.9.1)–(2.9.2), and the modified likelihood equations are asymptotically equivalent to the corresponding likelihood equations, and satisfy all the Bartlett conditions above, we have the following results for symmetric distributions.

Lemma 2.1: The asymptotic distribution of $\sqrt{m}(\hat{\mu} - \mu)/\sigma$ is normal N(0, 1);

$$\sigma^2/m \cong 1/\{-E(\partial^2 \ln L^*/\partial\mu^2)\}.$$

Lemma 2.2: For large n, the distribution of $(n-1)\hat{\sigma}^2/\sigma^2$ is chi-square with n – 1 degrees of freedom;

$$E(\hat{\sigma}) \cong \sigma \quad \text{and} \quad \sigma^2/2n \cong 1/\{-E(\partial^2 \ln L^*/\partial\sigma^2)\}.$$

Lemma 2.3: Asymptotically, the MML estimators $\hat{\mu}$ and $\hat{\sigma}^2$ are independently distributed since

$$E(\partial^{r+s} \ln L^*/\partial\mu^r \partial\sigma^s) = 0 \quad \text{for} \quad r \geq 1 \quad \text{and} \quad s \geq 1.$$

For an alternative proof developed by solving the differential equations $\partial \ln L^*/\partial\mu = 0$ and $\partial \ln L^*/\partial\sigma = 0$, see Appendix 2C.

For skew distributions, we have the following results; $\hat{\mu}(\sigma) = K + D\sigma$ and D is a constant.

Lemma 2.4: The conditional (σ known) distribution of $\sqrt{m}(\hat{\mu}(\sigma) - \mu)/\sigma$ is asymptotically normal N(0, 1);

$$\sigma^2/m \cong 1/\{-E(\partial^2 \ln L^*/\partial\mu^2)\}.$$

This follows from the factorization

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{m}{\sigma^2}(\hat{\mu}(\sigma) - \mu). \tag{2.9.6}$$

Remark: Alternatively, the asymptotic normality of $\hat{\mu}$ and $\hat{\mu}$ (σ) follows from the fact that they are linear functions of order statistics with uniformly bounded coefficients (David, 1981).

Lemma 2.5: The conditional (μ known) distribution of $n\hat{\sigma}(\mu)^2/\sigma^2$ is, for large n, chi-square with n degrees of freedom;

$$E\{\hat{\sigma}(\mu)\} \cong \sigma \quad \text{and} \quad \sigma^2/2n \cong 1/\{-E(\partial^2 \ln L^*/\partial\sigma^2)\}.$$

This follows from (2.9.4) and the values of $E(\partial^r \ln L^*/\partial\sigma^r)$ which are exactly the same as in the situation when the underlying distribution is symmetric. Note also the factorization.

$$\sum_{i=1}^{n} \beta_i (y_{(i)} - \mu)^2 \equiv m(K - \mu)^2 + \sum_{i=1}^{n} \beta_i (y_{(i)} - K)^2, \tag{2.9.7}$$

K being a linear function. However, $\hat{\mu}$ and $\hat{\sigma}$ are not independent of one another, even asymptotically, since the mixed derivatives $E(\partial^{r+s} \ln L^*/\partial\mu^r\partial\sigma^s)$ are not zero for $r \ge 1$ and $s \ge 1$.

Remark: We show in Section 2.11 that the distribution of $(n-1)\hat{\sigma}^2/\sigma^2$ is more accurately a multiple of chi-square (equations 2.11.16 and 2.11.19).

Lemma 2.6: The asymptotic distribution of $\sqrt{(n/v_{11})}(\hat{\mu}(\hat{\sigma}) - \mu)/\hat{\sigma}$ is normal N(0, 1), since $\hat{\sigma}$ converges to σ as n tends to infinity; $\hat{\sigma}$ is given in (2.4.4). This follows from the well-known Slutsky's theorem, along the same lines as in Chapter 1 (Section 1.3); v_{11} is the first element in the asymptotic covariance matrix such as (2.5.13).

Comment: If we have a linear contrast

$$\sum_{i=1}^{c} l_i\hat{\mu}_i, \quad \sum_{i=1}^{c} l_i = 0, \tag{2.9.8}$$

of c MML estimators calculated from independent samples of size n, and $\hat{\sigma}^2 = \sum_{i=1}^{c} \hat{\sigma}_i^2/c$ is the "pooled" MML estimator of σ^2, the Bartlett conditions are satisfied and $\sum_{i=1}^{k} l_i\hat{\mu}_i$ and $\hat{\sigma}^2$ are independently distributed (asymptotically). As a result, the asymptotic distribution of

$$\sqrt{m}\sum_{i=1}^{c} l_i\hat{\mu}_i \Bigg/ \left(\hat{\sigma}\sqrt{\sum_{i=1}^{c} l_i^2}\right) \tag{2.9.9}$$

is normal N(0, 1). For small n, the distribution is closely approximated by the Student t with c(n – 1) degrees of freedom. This will be illustrated in Chapters 6-8. The results also extend to situations where the sample sizes are not necessarily equal. Applications of the statistic (2.9.9) are particularly in the area of Experimental Design and are discussed in Chapter 6.

Comment: In addition to the location and scale parameters, numerous distributions have a third parameter (called shape parameter) which usually determines the shape of a distribution. For example, the parameter p in (2.2.9) or in (2.2.13) is a shape parameter which might, in practice, be unknown. Determination of a shape parameter is discussed in Chapter 9. It is, however, shown in Chapter 8 that the MML estimators $\hat{\mu}$ and $\hat{\sigma}$ and the hypothesis testing procedures based on them are robust to plausible deviations from the true value of a shape parameter. A very precise estimate of p is not, therefore, too crucial for applications of the method of modified likelihood. Unlike the Huber M-estimators, the MML estimators (and the hypothesis testing procedures based on them) are not confined to long-tailed symmetric distributions. In fact, the MML estimators are defined and have good efficiency and robustness properties for all the three types of distributions: skew, short-tailed symmetric and long-tailed symmetric distributions. This is illustrated in Chapter 8. See also Tables 2.1–2.4. It may be noted that skew distributions occur more frequently in practice than do symmetric distributions (Pearson, 1931).

2.10 HYPOTHESIS TESTING

Testing H_0: $\mu = 0$ against H_1: $\mu > 0$ is an important problem both in theory and in practice. If the underlying distribution is normal $N(\mu, \sigma^2)$, the statistic

$$t = \sqrt{n}\,(\bar{y}/s) \tag{2.10.1}$$

is used. The null distribution of t is the Student t with $\nu = n - 1$ d.f. The non-null distribution of t is noncentral t with d.f. ν and noncentrality parameter λ_1, $\lambda_1^2 = n(\mu/\sigma)^2$. The test is UMP (uniformly most powerful). For large n, the power of the test is given by

$$\text{Power} = P(Z \geq z_\alpha - |\lambda_1|), \tag{2.10.2}$$

where Z is a normal N(0, 1) variate and z_α is its $100(1 - \alpha)$% point. Consider now non-normal distributions.

Symmetric: Suppose the distribution is one of (2.2.9). The test is now based on the MML estimators (2.4.8). We note from the equation (2.3.18) that

$$V(\hat{\mu}) \cong \sigma^2/M \quad (M = 2pm/k) \quad (p \geq 2).$$

Since $\sigma^2/M \cong (p - 3/2)(p + 1)\,\sigma^2/np(p - 1/2)$ for large n and the expression on the right hand side gives closer approximations to the true variances of $\hat{\mu}$ (Table 2.1), we define the statistic

$$T = \sqrt{\left\{\frac{np(p - 1/2)}{(p - 3/2)(p + 1)}\right\}}\left(\frac{\hat{\mu}}{\hat{\sigma}}\right). \tag{2.10.3}$$

Large values of T lead to the rejection of H_0 in favour of H_1. In view of Lemmas 2.1–2.3, the null distribution of T is referred to the Student t with $\nu = n - 1$ d.f. For large n, the power of the test is given by

$$P(Z \geq z_\alpha - |\lambda_2|), \tag{2.10.4}$$

where $\lambda_2^2 = \{np(p - 1/2)/(p - 3/2)(p + 1)\}(\mu/\sigma)^2$. Since $\lambda_2^2/\lambda_1^2 > 1$, the T test is more powerful than the t test; this is also illustrated in the next chapter. We will show in Chapter 8 that the MML estimators $\hat{\mu}$ and $\hat{\sigma}$ and the T test are remarkably robust to long-tailed distributions and to outliers in a sample. This is due to the umbrella ordering of the coefficients β_i $(1 \leq i \leq n)$, namely, the coefficients increase until the middle value and then decrease in a symmetric fashion. For p = 3.5 and n = 20, for example, the first ten β_i coefficients are, $\beta_{n-i+1} = \beta_i$:

0.006 0.259 0.452 0.604 0.725 0.820 0.893 0.946 0.981 0.998 (2.10.5)

Thus, the extreme observations on both sides of an ordered sample automatically receive small weights. This depletes the influence of long tails and outliers.

Skew distributions: Suppose that the distribution is one of (2.5.1). The MML estimators $\hat{\mu}$ and $\hat{\sigma}$ are given in (2.5.11). To test H_0: $\mu = 0$ against H_1: $\mu \neq 0$, for example, we define the statistic

$$T = \sqrt{\left(\frac{n}{\nu_{11}}\right)}\left(\frac{\hat{\mu}}{\hat{\sigma}}\right) \tag{2.10.6}$$

where ν_{11} is the first element in (2.5.13). For b = 1 (logistic distribution), for example, $\nu_{11} = (b + 2)/b = 3$. Large values of | T | lead to the rejection of H_0 in favour of H_1. The null distribution of T is referred to the Student t with $\nu = n - 1$ d.f. and gives remarkably accurate approximations to the percentage points. For example, we have the following simulated values of the probability:

$$\text{Prob}(\,|\,T\,| \geq t_{0.025}(\nu)\mid H_0): \tag{2.10.7}$$

n	b = 0.5	1.0	2.0	4.0
10	0.045	0.040	0.046	0.069
20	0.045	0.046	0.048	0.061
30	0.050	0.051	0.051	0.053
50	0.046	0.049	0.050	0.052

It may be noted that the Generalized Logistic has considerable amount of skewness for $b \geq 4$ (Appendix 2D). That seems to be the reason that a larger sample size n is needed for (2.10.7) to give accurate approximations, for $b \geq 4$.

For large n, the power of the T test is given by

$$\text{Prob}(\,|\,Z\,| \geq z_{\alpha/2} - |\,\lambda\,|\,); \tag{2.10.8}$$

$\lambda^2 = (n/\nu_{11})(\mu/\sigma)^2$. Realizing that the variance of the Generalized Logistic distribution is $\{\psi'(b) + \psi'(1)\}\,\sigma^2$ (Appendix 2D), the power of the t test is given by (2.10.8) with λ^2 replaced by

$$\lambda^2 = \frac{n}{\psi'(b) + \psi'(1)}\left(\frac{\mu}{\sigma}\right)^2. \tag{2.10.9}$$

Since $\nu_{11} < \psi'(b) + \psi'(1)$, e.g. $\nu_{11} = 3$ and $\psi'(b) + \psi'(1) = 3.29$ for $b = 1$, the T test is more powerful than the t test. We will show in Chapter 8 that the T test is remarkably robust to skew distributions and outliers. This is due to the half-umbrella ordering of the coefficients β_i $(1 \leq i \leq n)$: they decrease in the direction of the long tail. For n = 10, for example, we have the following values of these coefficients calculated from (2.5.5); b = 8:

0.206 0.161 0.130 0.105 0.084 0.065 0.049 0.034 0.020 0.007

Thus, the extreme order statistics in the direction of the long tail automatically receive small weights (for b = 8, the long tail is on the right hand side). This depletes the effect of the long tail.

Before we consider testing for an assumed value of a scale parameter, it is pertinent to introduce Huber (1964, 1981) estimators and the tests based on them.

2.11 HUBER M-ESTIMATORS

Let $y_1, y_2, ..., y_n$ be a random sample from a distribution of the type $(1/\sigma)\, f((y - \mu)/\sigma)$. Huber (1964,1981) assumed in particular that f is long-tailed symmetric (kurtosis $\mu_4/\mu_2^2 > 3$) and proposed a new method to estimate μ as follows:

The log-likelihood function is

$$\ln L = -n \ln \sigma + \sum_{i=1}^{n} \ln f(z_i), \quad z_i = (y_i - \mu)/\sigma. \tag{2.11.1}$$

If the functional form f is known, the ML estimator of μ is the solution of the equation

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma} \sum_{i=1}^{n} \xi(z_i) = 0, \quad \xi(z) = -f'(z)/f(z). \tag{2.11.2}$$

Writing $w_i = w_i(z) = \xi(z_i)/z_i$, (2.11.2) can be written as $\sum_{i=1}^{n} w_i (y_i - \mu) = 0$ so that

$$\mu = \sum_{i=1}^{n} w_i y_i \Big/ \sum_{i=1}^{n} w_i . \tag{2.11.3}$$

Given σ and $\xi(z)$, one may solve (2.11.3) by iteration (Low, 1959; Gross, 1976).

In practice, however, neither $\xi(z)$ nor σ are known. Huber (1964) proposed a function $\xi(z)$ as

$$\xi(z) = \begin{cases} z & \text{if } |z| \le c \\ c\,\mathrm{sgn}(z) & \text{if } |z| > c \end{cases}. \tag{2.11.4}$$

This is a combination of the normal distribution in the middle and the double-exponential in the tails. The popular choices of c have been 1.345, 1.5 and 2.0 (Birch and Myers,1982) because these choices correspond roughly to 10, 5 and 2.5 percent censoring on either side of a normal sample.

For the unknown σ, Huber (1964,1977), Gross (1976,1977) and many others have used $\tilde{\sigma}_0$ = mad = median | y_i-median (y_i) | as an estimate. Later, Huber(1981) and Birch and Myers (1982) suggested changing mad to mad/0.6745 (an asymptotically unbiased estimator of σ in case of a normal distribution).

The solutions of (2.11.2)–(2.11.4) are referred to as Huber M-estimators (Huber, 1964, 1977, 1981), to be denoted by $\hat{\mu}_H$. Huber showed that the asymptotic variance of the M- estimator is

$$(\sigma^2/n)\, E(\xi^2(z))/\{E(\xi'(z))\}^2. \tag{2.11.5}$$

Hence, the estimator of σ which matches with $\hat{\mu}_H$ is

$$\hat{\sigma}_H = \left\{ n\tilde{\sigma}_0^{\,2} \left[\sum_{i=1}^{n} \xi^2\left(\frac{y_i - \hat{\mu}_H}{\tilde{\sigma}_0} \right) \right] \Big/ \left[\sum_{i=1}^{n} \xi'\left(\frac{y_i - \hat{\mu}_H}{\tilde{\sigma}_0} \right) \right]^2 \right\}^{1/2} \tag{2.11.6}$$

Descending functions: A function like $\xi(z) = z/(1 + z^2)$ which decreases as $|z|$ increases is called a descending ξ-function. When the functional form f(z) is not known and one wishes to deplete the effect of tails, one may then approximate f(z) by descending functions. Three most popular descending functions are:

1. The wave function (Andrews et al.,1972; Andrews, 1974)

$$\xi(z) = \begin{cases} \sin(z) & \text{if } |z| \le \pi \\ 0 & \text{if } |z| > \pi \end{cases} \tag{2.11.7}$$

2. The bisquare function (Beaton and Tukey, 1974)

$$\xi(z) = \begin{cases} z(1 - z^2)^2 & \text{if } |z| \le 1 \\ 0 & \text{if } |z| > 1 \end{cases} \tag{2.11.8}$$

3. The Hampel piecewise linear function (Hampel, 1974)

$$\xi(z) = \text{sgn}(z) \begin{cases} |z| & 0 \le |z| < a \\ a & a \le |z| < b \\ \dfrac{c - |z|}{c - b} & b \le |z| < c \\ 0 & c \le |z| \end{cases} \tag{2.11.9}$$

This results in different estimators for different values of a, b and c.

In an extensive numerical study, Gross (1976) examined 25 representative estimators (out of 65 discussed in Andrews et al.,1972) of μ (and σ) and particularly recommended three of them, namely, the wave estimators w24, the bisquare estimators BS82 and the Hampel estimators H22. Since the three have similar efficiency and robustness properties (Gross, 1976; Tiku, 1980; Dunnett, 1982; Tiku et al., 1986), we reproduce the equations only for the pair w24:

$$T_0 = \text{median}\,\{y_i\} \text{ and } S_0 = \text{median}\,\{|\, y_i - T_0 \,|\} \quad (1 \le i \le n),$$

$$\hat{\mu}_w = T_0 + (hS_0)\tan^{-1}\left[\frac{\Sigma_i \sin(z_i)}{\Sigma_i \cos(z_i)}\right]$$

and

$$\hat{\sigma}_w = (hS_0)\left[n\,\frac{\Sigma_i \sin^2(z_i)}{(\Sigma_i \cos(z_i))^2}\right]^{1/2}; \tag{2.11.10}$$

$h = 2.4$, and the summations include only those i for which $|z_i| < \pi$, $z_i = (y_i - T_0)/(hS_0)$. The estimates are calculated in a single iteration.

The estimators w24, BS82 and H22 censor observations (smallest and largest) implicitly; the number of censored observations is not the same for every sample. For symmetric distributions, the Huber estimators of μ are unbiased and uncorrelated with the matching estimators of σ. For long-tailed symmetric distributions, $\hat{\mu}_w$ (and other such Huber estimators) are remarkably efficient, but not for skew and short-tailed symmetric distributions. Other estimators which are equally efficient for long-tailed symmetric distributions are the Tukey estimator and the Tiku (1980a) MML estimator based on Type II censored samples

$$y_{(r+1)} \le y_{(r+2)} \le \ldots \le y_{(n-r)}. \tag{2.11.11}$$

The formulae for calculating these estimators are given in Appendix 2E. For long-tailed symmetric distributions with finite mean and variance (e.g., the family 2.2.9 with $p \ge 2$), r in (2.11.11) is chosen to be the integer value $[0.5 + 0.1n]$. For more extreme non-normal symmetric distributions (e.g., Cauchy and normal/uniform) which can easily be identified by using

Q-Q plots or goodness-of-fit tests, r in (2.11.11) is chosen to be [0.5 + 0.3n]. A detailed comparison of Huber, Tukey and Tiku estimators is given in Tiku (1980a) and Tiku et al. (1986). See also Dunnett (1982).

To test the null hypothesis H_0: $\mu = 0$, the statistic based on the w24 estimators is

$$t_w = \sqrt{n}\,(\hat{\mu}_w/\hat{\sigma}_w). \tag{2.11.12}$$

The null distribution of t_w is referred to the Student t with $\nu = n - 1$ d.f. The statistics based on BS82 and H22 are exactly similar to (2.11.12). For sample sizes $n \le 20$, however, the t_w test (and those based on BS82 and H22) have type I error considerably larger than the presumed level (Tiku et al., 1986). For the logistic distribution, for example, we have the following values;

$t_B = \sqrt{n}\,(\hat{\mu}_B/\hat{\sigma}_B)$ is based on the BS82 estimators mentioned above:

Values of the Prob($t_w \le t_{0.05}(\nu)$ | H_0)

	n = 10	20	30	50
t_w	0.072	0.061	0.054	0.046
t_B	0.072	0.064	0.055	0.049

When the percentage points of t_w and t_B and other such statistics are corrected (by simulation since it is difficult to do this analytically), they have power and robustness properties similar to the Tukey and the Tiku statistics

$$t_{Trim} = \sqrt{(n-2r)}\,(\hat{\mu}_{Trim}/\hat{\sigma}_{Trim}) \quad \text{and} \quad t_c = \sqrt{m}\,(\hat{\mu}_c/\hat{\sigma}_c); \tag{2.11.13}$$

see, for example, Tiku (1980a) and Dunnett (1982). It may be noted that the statistics (2.11.12) and (2.11.13) are applicable only to long-tailed symmetric distributions. They do not give good results if used for skew or short-tailed symmetric distributions. We illustrate this now with respect to the Huber M-estimators.

Consider the Generalized Logistic distribution (2.5.1). The MML estimators $\hat{\mu}$ and $\hat{\sigma}$ are given in (2.5.11). Without bias corrections, the M-estimators give misleading results. For $\mu = 0$, $b = 0.5$ and $n = 20$, for example, the simulated mean of $\hat{\mu}$ is -0.017σ but that of $\hat{\mu}_w$ is -1.236σ. The bias corrected M-estimators are

$$\text{w24}: \mu_w^* = \hat{\mu}_w - [\psi(b) - \psi(1)]\,\sigma_w^*$$

and
$$\sigma_w^* = \hat{\sigma}_w/\sqrt{\{\psi'(b) + \psi'(1)\}} \tag{2.11.14}$$

and, similarly, BS82 and H22. The bias corrected LS estimators are

$$\text{LS}: \tilde{\mu} = \bar{y} - [\psi(b) - \psi(1)]\tilde{\sigma}$$

and
$$\tilde{\sigma} = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1)\,\{\psi'(b) + \psi'(1)\}}\,. \tag{2.11.15}$$

It may be noted that no such bias correction is needed for the MML estimators $\hat{\mu}$ and $\hat{\sigma}$; they are self bias-correcting. Given in Table 2.4 are the simulated means and MSE (mean square errors) of the estimators above. Since the values for BS82 and H22 are almost the same as w24, they are not reproduced. It can be seen that the M-estimators are considerably less efficient than the MML estimators. They are even somewhat less efficient than the LS estimators (overall). Moreover, σ_w* has substantial bias.

Table 2.4: Values of the mean and mean square error; $\mu = 0$ without loss of generality.

n	b	$\hat{\mu}$	μ_w*	$\tilde{\mu}$	$\hat{\sigma}$	σ_w*	$\tilde{\sigma}$
				$(1/\sigma)$ Mean			
10	0.5	– 0.025	– 0.052	– 0.063	1.032	0.843	0.947
	1.0	– 0.003	0.002	0.004	1.027	0.884	0.966
	4.0	0.070	0.162	0.080	1.000	0.863	0.954
20	0.5	– 0.017	0.006	– 0.031	1.014	0.880	0.975
	1.0	– 0.003	– 0.008	0.002	1.019	0.924	0.977
	4.0	0.023	0.080	0.043	1.008	0.912	0.975
				$(1/\sigma^2)$ MSE			
10	0.5	0.546	0.686	0.563	0.090	0.102	0.095
	1.0	0.311	0.322	0.326	0.083	0.089	0.086
	4.0	0.239	0.264	0.268	0.071	0.088	0.084
20	0.5	0.265	0.379	0.288	0.041	0.051	0.052
	1.0	0.153	0.158	0.164	0.038	0.042	0.048
	4.0	0.120	0.127	0.133	0.034	0.042	0.044

We show in Chapter 7 that for symmetric short-tailed distributions also the M-estimators are considerably less efficient than the MML estimators. The use of the M-estimators should, therefore, be limited to symmetric long-tailed distributions. However, there is a problem: the M-estimators of σ can have substantial downward bias (Chapter 8). This is also illustrated in Chapter 11.

Scale parameter: To test H_0: $\sigma = \sigma_0$ against H_0: $\sigma > \sigma_0$, the MML estimator $\hat{\sigma}$ is used. The statistic $d\hat{\sigma}^2/\sigma_0^2$ is employed and its null distribution is referred to a chi-square with n – 1 d.f. (Lemmas 2.2 and 2.5). A sharper result, however, is obtained by using the technique of Lawless (1982):

Denote the variance of $\hat{\sigma}/\sigma$ by $V = \nu_{22}/n$ where ν_{22} is the last element of the covariance matrix such as (2.5.13). Let

$$\sqrt{\chi^2} = \sqrt{d}\,(\hat{\sigma}/\sigma) \tag{2.11.16}$$

where χ^2 is a chi-square variate with ν d.f. The constant coefficients d and ν are determined by equating the expected values of the variables and their squares on both sides of (2.11.16). Realizing that $E(\hat{\sigma}/\sigma) \cong 1$ and writing Var $(\hat{\sigma}/\sigma) = V$, $d = \nu/(V + 1)$ since $E(X^2) = Var(X) + E^2(X)$ for any random variable X, and ν is determined by the equation

$$\sqrt{\{2(V+1)\}} = \sqrt{\nu}\,\Gamma\left(\frac{\nu}{2}\right)\Big/\Gamma\left(\frac{\nu+1}{2}\right). \tag{2.11.17}$$

Consider, for example, the extreme-value distribution (2.6.1). The MML estimator $\hat{\sigma} = \hat{\eta}$ is given in (2.6.10). The bias-corrected estimator has the divisor 2n replaced by 2m. Therefore, $V = (n/m^2)(6/\pi^2)$ from (2.6.11); ν and d can now be determined. For example, we have the following values:

n =	5	10	15	20	25	30
ν	3.2	6.6	10.4	14	18	22
d	2.73	6.13	9.92	13.53	17.52	21.53

To illustrate the accuracy of this two-moment $\sqrt{\chi^2}$ approximation, Vaughan and Tiku (2000, p.60) give the simulated values of the probability

$$\text{Prob}\left\{\frac{d\hat{\sigma}^2}{\sigma_0{}^2} \geq \chi_{0.05}^2\ (\nu)\right\} \tag{2.11.18}$$

as follows:

n =	5	10	15	20	25	30
Prob	0.052	0.051	0.050	0.050	0.050	0.050

It can be seen that the approximation is remarkably accurate. Large values of $d(\hat{\sigma}/\sigma_0)^2$ lead to the rejection of H_0 in favour of H_1.

A simpler chi-square approximation, although somewhat less accurate than (2.11.16), is obtained as follows.

Write
$$\chi^2 = \frac{(n-1)}{h}\left(\frac{\hat{\sigma}}{\sigma}\right)^2 \tag{2.11.19}$$

where χ^2 is a chi-square variate with n – 1 degrees of freedom. Equating the expected values on both sides, we get

$$h = 1 + V, \quad V = \text{Var}(\hat{\sigma}/\sigma); \quad E(\hat{\sigma}/\sigma) \cong 1. \tag{2.11.20}$$

Thus, the distribution of $(n-1)\hat{\sigma}^2/\sigma^2$ is a multiple of χ^2; $h \cong 1$ if $V \cong 0$. The chi-square approximation (2.11.20) is computationally much easier than (2.11.16).

SUMMARY

In this Chapter, we consider the problem of estimating the location parameter (mean) μ and the scale parameter (standard deviation) σ of a given distribution. We consider three families of distributions: (i) skew, (ii) short-tailed symmetric, and (iii) long-tailed symmetric. The maximum likelihood estimators (MLE) being generally intractable, we derive the modified maximum likelihood estimators (MMLE). The latter are obtained by (i) expressing the likelihood equations in terms of the order statistics of a random sample and (ii) linearizing the intractable terms. The resulting equations have explicit solutions called MMLE. We show that the MMLE are asympototically fully efficient and have high efficiency for small sample sizes n. We discuss the distributions of the MMLE $\hat{\mu}$ and $\hat{\sigma}$. For large n, $\sqrt{m}(\hat{\mu}-\mu)/\sigma$ is normal N(0, 1) and $(n-1)\hat{\sigma}^2/\sigma^2$ is a multiple of chi-square with n – 1 degrees of freedom. We compare the MMLE with the BLUE (best linear unbiased estimators) and show that the former are jointly more efficient besides being much easier to compute. We give numerical examples to illustrate these findings. We also compare the MMLE with the Huber M-estimators and show that the former are generally more efficient. We study the distributions of the Student type statistics $\sqrt{m}\,\hat{\mu}/\hat{\sigma}$ for the three families above. They are asymptotically

normal N(0, 1) and, for small sample sizes n, they are closely approximated by the Student t distribution with n – 1 degrees of freedom. We show that the MLE and the MMLE are numerically very close to one another.

Note: As pointed out in Tiku (1996, p. 266) and Bian and Tiku (1997b, p.86), Balakrishnan in several of his publications and in his book (Balakrishnan and Cohen, 1991) use the method of modified likelihood and claim originality of the method. That is not true, however; see Balakrishnan (1989a , b) and his retractions. See also the book review by Saleh (1994).

APPENDIX 2A

ASYMPTOTIC EQUIVALENCE

Asymptotic equivalence of a modified likelihood equation $d \ln L^*/d\theta = 0$ (and its derivative) and the corrsponding likelihood equation $d \ln L^*/d\theta = 0$ (and its derivative) follows from the following result due to Hoeffding (1953, Theorem 1).

Let z_i $(1 \le i \le n)$ be independent standard random variates (with location parameter $\mu = 0$ and scale parameter $\sigma = 1$) with a common cumulative density function F(z). Let

$$z_{(1)} \le z_{(2)} \le \ldots \le z_{(n)} \tag{2A.1}$$

be the ordered variates with $t_{i:n} = E\{z_{(i)}\}$, $1 \le i \le n$. It is assumed that

$$\int_{-\infty}^{\infty} |z| \, dF(z) < \infty \tag{2A.2}$$

which implies that $t_{i:n}$ are finite for all i = 1, 2, ..., n. Let g(z) be a real-valued continuous function such that $|\, g(z) \,| \le h(z)$ where the function h(z) is convex and

$$\int_{-\infty}^{\infty} h(z) \, dF(z) < \infty . \tag{2A.3}$$

Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} g\{t_{i:n}\} = \int_{-\infty}^{\infty} g(z) \, dF(z) . \tag{2A.4}$$

Consider, for example, the likelihood equations (2.6.3)-(2.6.4) and the corresponding modified likelihood equations obtained by replacing exp $\{-z_{(i)}\}$ by $\alpha_i - \beta_i z_{(i)}$; the coefficients α_i and β_i $(1 \le i \le n)$ are given in (2.6.6). The differences

$$(1/n)\{d \ln L^*/d\delta) - d \ln L/d\delta)\}, \quad (1/n)\{d^2 \ln L^*/d\delta^2) - (d^2 \ln L/d\delta^2)\},$$

etc., are in the limit (n tends to infinity) expressions in terms of $(t_{(i)} = t_{i:n})$

$$\lim_{n \to \infty} E\left[\frac{1}{n} \sum_{i=1}^{n} ({z_{(i)}}^2 - {t_{(i)}}^2)\right], \quad \lim_{n \to \infty} E\left[\frac{1}{n} \sum_{i=1}^{n} (e^{-z_{(i)}} - e^{-t_{(i)}})\right],$$

$$\lim_{n \to \infty} E\left[\frac{1}{n} \sum_{i=1}^{n} z_{(i)} (e^{-z_{(i)}} - e^{-t_{(i)}})\right]$$

and

$$\lim_{n \to \infty} E\left[\frac{1}{n} \sum_{i=1}^{n} {z_{(i)}}^2 (e^{-z_{(i)}} - e^{-t_{(i)}})\right] \tag{2A.5}$$

Realizing that the absolute values of the functions $g_1(z) = z^2$, $g_2(z) = \exp(-z)$, $g_3(z) = z \exp(-z)$ and $g_4(z) = z^2 \exp(-z)$ are dominated by non-negative convex functions with finite expectations (with respect to the underlying extreme-value distribution), it follows from (2A.4) that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g_j\{t_{i:n}\} = E\{g_j(z)\} \quad (j = 1, 2, 3, 4). \tag{2A.6}$$

Further,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_{ii:n} \exp(-t_{i:n}) = 0, \quad \sigma_{ii:n} = Var\{z_{(i)}\}, \tag{2A.7}$$

since the variances $\sigma_{ii:n}$ $(1 \le i \le n)$ are bounded above (David and Groeneveld, 1982), and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sigma_{ii:n} = 0; \tag{2A.8}$$

(2A.8) follows from the fact that

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} t_{(i)}^2 &= E(Z^2) \\ &= \lim_{n\to\infty} E\left[\frac{1}{n} \sum_{i=1}^{n} z_{(i)}^2\right] \\ &= \lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=1}^{n} (t_{i:n}^2 + \sigma_{ii:n})\right]. \end{aligned} \tag{2A.9}$$

From the results (2A.5 – 2A.9), the asymptotic equivalence follows immediately (Vaughan and Tiku, 2000, Appendix A).

Remark: The equation (2A.4) can be used to obtain the minimum variance bounds. Consider, for example, the equation (2.4.6). Since $\sum_{i=1}^{n} \alpha_i = 0$,

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right) = \frac{2n}{\sigma^2}\left[1 - \frac{p}{nk} \sum_{i=1}^{n} \alpha_i E(z_{(i)})\right], \quad E(z_{(i)}) = t_{(i)}; \tag{2A.10}$$

α_i is given in (2.3.14). From (2A.4),

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \alpha_i t_{(i)} &= E\left(\frac{(2/k)z^4}{\{(1+(1/k)z^2\}^2}\right) \\ &= 2k[1 - 2E\{1 + (1/k)z^2\}^{-1} + E\{1 + (1/k)z^2\}^{-2}]. \end{aligned} \tag{2A.11}$$

The expected values in (2A.11) are easy to evaluate from (2.3.20) and when substituted in (2A.10) give the result that

$$-E(\partial^2 \ln L^*/\partial\sigma^2) = 2n(p - 1/2)/(p + 1)\sigma^2.$$

It is very interesting to note that to compute the MMLE and the second derivatives of ln L*, only the values of $E(z_{(i)})$ are required and not $E(z_{(i)}^2)$.

APPENDIX 2B

NUMERICAL COMPARISON

The MML estimators of μ and σ are asymptotically equivalent to the ML estimators. To verify how close the two are numerically for small n, we followed Tan (1985, p.332). We calculated the MMLE from the modified likelihood equations and substituted them in the likelihood equations.

Symmetric distributions: For the symmetric family (2.2.9) we calculated the MML estimates $\hat{\mu}$ and $\hat{\sigma}$ from (2.4.7) with the divisor $2\sqrt{n(n-1)}$ replaced by 2n in the expression for $\hat{\sigma}$ since the true solution has divisor 2n. We used the exact values of $t_{(i)} = E(z_{(i)})$ given in Tiku and Kumra (1981). We substituded them in the likelihood equations (2.4.1) to obtain the values of

$$\hat{\tau} = \frac{1}{n}\left(\frac{\partial \ln L}{\partial \mu}\right)_{\mu=\hat{\mu},\,\sigma=\hat{\sigma}} \quad \text{and} \quad \hat{\delta} = \frac{1}{n}\left(\frac{\partial \ln L}{\partial \sigma}\right)_{\mu=\hat{\mu},\,\sigma=\hat{\sigma}}. \tag{2B.1}$$

We repeated these calculations for N = [100,000/n] random samples of size n generated from (2.2.9). The means and variances of the N values of $\hat{\tau}$ and $\hat{\delta}$ are given below; $\mu = 0$ and $\sigma = 1$ without loss of generality:

		p = 3.5			p = 5.0		
	n =	6	10	20	n = 6	10	20
$10^4 \times$ Mean	$\hat{\tau}$	2.11	0.18	0.17	– 1.36	0.65	0.10
$10^4 \times$ Var	$\hat{\tau}$	6.35	2.33	0.69	4.49	1.84	0.55
$10^2 \times$ Mean	$\hat{\delta}$	– 5.70	– 4.57	– 2.98	– 4.24	– 3.50	– 2.42
$10^2 \times$ Var	$\hat{\delta}$	0.33	0.18	0.07	0.16	0.11	0.01

The closeness of the ML and the MML estimates is apparent. Tan (1985) has similar results.

Skew distributions: A similar process with the family (2.5.1) and the MML estimates (2.5.10) substituted in (2.5.2)-(2.5.3) gave the following values. For n = 6 and 10, we used the exact values of $t_{(i)}$ given in Balakrishnan and Leung (1988):

		b = 1.0			b = 4.0		
	n =	6	10	20	n = 6	10	20
$10^4 \times$ Mean	$\hat{\tau}$	– 0.43	0.36	– 0.44	– 2.86	– 1.64	– 0.87
$10^4 \times$ Var	$\hat{\tau}$	1.71	0.53	0.38	5.65	1.36	0.39
$10^2 \times$ Mean	$\hat{\delta}$	– 5.13	– 3.56	– 3.11	– 5.93	– 3.89	– 2.71
$10^2 \times$ Var	$\hat{\delta}$	0.23	0.09	0.05	0.32	0.10	0.05

The ML and MML estimates are very close to one another for both symmetric as well as skew distributions. There is all the reason, therefore, to use the MML estimators in situations where the ML estimators are not readily available.

APPENDIX 2C

ASYMPTOTIC DISTRIBUTIONS

It follows from (2.3.18) that $\hat{\mu}$ is asymptotically distributed as $N(\mu, \sigma^2/M)$ From (2.4.8), $\hat{\sigma}^2 \cong C/(n-1)$ since $B/\sqrt{nC} \cong 0$ for large n. Solving the differential equations (2.4.5)-(2.4.6), we get

$$L^* \cong (\sigma^2)^{-(n-1)/2} \exp\left[-\frac{(n-1)\hat{\sigma}^2}{2\sigma^2} - \frac{M}{2\sigma^2}(\hat{\mu}-\mu)^2\right] \quad H(y), \tag{2C.1}$$

where H(y) is an analytical function free of μ and σ^2. Making the transformation $\hat{\mu} = \hat{\mu}(y)$, $\hat{\sigma}^2 = \hat{\sigma}^2(y)$, $w_3 = y_3, \ldots, w_n = y_n$ and integrating out $(w_3, \ldots, w_n)$, we obtain the joint density of $(\hat{\mu}, X = (n-1)\hat{\sigma}^2/\sigma^2)$ as

$$f(\hat{\mu}, X) = f_1(\hat{\mu}; \mu, \sigma^2/M)\, f_2(X; n-1)\, R(\hat{\mu}, z) \tag{2C.2}$$

where f_1 is the normal density $N(\mu, \sigma^2/M)$, f_2 is the chi-square density with $n-1$ degrees of freedom, and $R(\hat{\mu}, Z)$ is an analytical function free of μ and σ^2. This follows readily from the fact that $-\infty < \hat{\mu} < \infty$ and $\hat{\sigma}^2 > 0$, and H(y) is free of μ and σ^2. Now, $(\hat{\mu}, \hat{\sigma}^2)$ are complete sufficient statistics from (2C.2) (see Hogg and Craig, 1978). Write $U(\hat{\mu}, z) = R(\hat{\mu}, z) - 1$; then

$$\int_{-\infty}^{\infty}\int_0^{\infty} f_1(\hat{\mu}; \mu, \sigma^2/M) f_2\,(X; n-1) U(\hat{u}, Z)\, d\hat{\mu} dZ = 0$$

for all $-\infty < \hat{\mu} < \infty$ and $\sigma^2 > 0$. Hence $U(\hat{\mu}, z) = 0$, i.e. $R(\hat{\mu}, z) = 1$ almost surely. Thus

$$f(\hat{\mu}, X) = f_1(\hat{\mu}; \mu, \sigma^2/M)\, f_2(X; n-1) \tag{2C.3}$$

and that proves the result.

APPENDIX 2D

THE PSI-FUNCTION

Consider the Generalized Logistic distribution (b > 0)

$$f(z) = be^{-z}/\{1 + e^{-z}\}^{b+1}, \quad -\infty < z < \infty \tag{2D.1}$$

The moment generating function of Z is

$$M_\theta(z) = E(e^{\theta z}) = b\Gamma(1-\theta)\,\Gamma(b+\theta)/\Gamma(b+1) \quad (|\theta| < 1). \tag{2D.2}$$

The rth moment of Z is

$$\mu_r' = \left\{\frac{d^r}{d\theta^r} M_\theta(z)\right\}_{\theta=0} \tag{2D.3}$$

and is an expression in terms of the psi-function $\psi(x) = \Gamma'(x)/\Gamma(x)$ and its derivative. In particular,

$$E(Z) = \psi(b) - \psi(1) \quad \text{and} \quad V(Z) = \psi'(b) + \psi'(1). \tag{2D.4}$$

Given below are the values of $\psi(b)$ and $\psi'(b)$. For several other properties of $\psi(x)$ and its derivatives, one may refer to Abramowitz and Stegun (1985):

b	$\psi(b)$	$\psi(b+1)$	$\psi'(b)$	$\psi'(b+1)$
0.5	– 1.9635	0.0365	4.9348	0.9348
1.0	– 0.5772	0.4228	1.6449	0.6449
2.0	0.4228	0.9228	0.6449	0.3949
4.0	1.2561	1.5061	0.2838	0.2213
6.0	1.7061	1.8728	0.1813	0.1536
8.0	2.0156	2.1406	0.1331	0.1175
10.0	2.2518	2.3518	0.1051	0.0951

The following results are needed in deriving the elements of the Fisher Information matrix. Writing

$$g(z) = e^{-z}/(1 + e^{-z}) \quad \text{and} \quad h(z) = e^{-z}/(1 + e^{-z})^2,$$

it is not difficult to show that

$$\begin{aligned} &E\{g(Z)\} = (b+1)^{-1},\ E\{Zg(Z)\} = (b+1)^{-1}(\psi(b) - \psi(2)) \\ &E\{h(Z)\} = b(b+1)^{-1}(b+2)^{-1} \\ &E\{Zh(Z)\} = b(b+1)^{-1}(b+2)^{-1}[\psi(b+1) - \psi(2)] \\ &E\{Z^2 h(Z)\} = b(b+1)^{-1}(b+2)^{-1}[\psi'(b+1) + \psi'(2) + \{\psi(b+1) - \psi(2)\}^2]. \end{aligned} \tag{2D.5}$$

The Pearson coefficients of skewness and kurtosis of the Generalized Logistic are given below:

	b = 0.5	1	2	4	6
Skewness $\mu_3/\mu_2^{3/2}$	– 0.855	0	0.577	0.868	0.961
Kurtosis μ_4/μ_2^2	5.400	4.2	4.333	4.758	4.951

APPENDIX 2E

ESTIMATORS BASED ON CENSORED SAMPLES

The Tukey estimators of μ and σ are usually denoted by $\hat{\mu}_{Trim}$ and $\hat{\sigma}_{Trim}$; the expression for $\hat{\mu}_{Trim}$ is given in (1.2.5) and ($\hat{\mu}_T = \hat{\mu}_{Trim}$)

$$\hat{\sigma}_{Trim} = \left[\frac{\sum_{i=r+1}^{n-r} (y_{(i)} - \hat{\mu}_T)^2 + r(y_{(r+1)} - \hat{\mu}_T)^2 + r(y_{(n-r)} - \hat{\mu}_T)^2}{n - 2r - 1} \right]^{1/2}. \tag{2E.1}$$

Tiku (1982) states that “non-normality essentially comes from the tails and once the extreme order statistics are censored, there is hardly any difference between a normal and a non-normal sample”. See also Tiku et al. (1986, pp. 22-23). He assumes normality for (2.11.11) and recommends the use of the MML estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ (Tiku, 1967, 1968, 1973) in the framework of robustness:

$$\hat{\mu}_c = \left\{ \sum_{i=r+1}^{n-r} y_{(i)} + r\beta(y_{(r+1)} + y_{(n-r)}) \right\} \Big/ m \quad (m = n - 2r + 2r\beta), \quad \text{and} \quad (2E.2)$$

$$\hat{\sigma}_c = \{B + \sqrt{(B^2 + 4AC)}\} / 2\sqrt{\{A(A-1)\}}, \quad A = n - 2r; \quad (2E.3)$$

$$B = r\alpha(y_{(n-r)} - y_{(r+1)}) \quad \text{and} \quad C = \sum_{i=r+1}^{n-r} y_{(i)}{}^2 + r\beta(y_{(n-r)}{}^2 + y_{(r+1)}{}^2) - m\hat{\mu}^2.$$

The coefficients α and β are calculated from the equations

$$\alpha = \frac{f(t)}{q} - \beta t \quad \text{and} \quad \beta = -\frac{f(t)}{q}\left[t - \frac{f(t)}{q}\right]; \quad (2E.4)$$

$q = r/n$, $f(t) = (2\pi)^{-1/2} \exp(-t^2/2)$ and t is determined by $\int_{-\infty}^{t} f(t)\,dt = 1 - q$; α and β are both positive fractions.

CHAPTER 3

Linear Regression with Normal and Non-normal Error Distributions

3.1 INTRODUCTION

In Chapter 2, we discussed the modified likelihood estimation of location and scale parameters of distributions of the type $(1/\sigma)f((y - \mu)/\sigma)$. We showed that the MML (modified maximum likelihood) estimators are explicit functions of sample observations and are highly efficient. In fact, they are asymptotically the MVB (minimum variance bound) estimators under some very general regularity conditions (Appendix 2A). We also showed that they are numerically the same (almost) as the ML estimators for all sample sizes (Appendix 2B); see also Chapter 4. In this chapter and the next three, we consider certain very important (from theoretical as well as practical point of view) topics, namely, linear regression, binary regression, autoregression and exprimental design. We derive the MML estimators of the parameters and study their efficiency properties. We show, in particular, that the MML estimators are considerably more efficient than the LS (least squares) estimators. It may be remembered that the LS estimators are widely used in these areas. In fact, we show that the LS estimators have a disconcerting feature, namely, their efficiencies (relative to the MML estimators) decrease as the sample size n increases and stabilize at values considerably less than 100 percent. We show that it is advantageous to use the modified likelihood methodology presented here, since it gives enormously more efficient estimators than the least squares estimators. We illustrate this by considering three families of distributions representing skew, short-tailed and long-tailed symmetric distributions. As in Chapter 2, it is shown that the weights given to ordered residuals in the formulation of the MML estimators have distinct features: (i) for skew distributions they decrease in the direction of the long tail, (ii) for short-tailed symmetric distributions they decrease untill the middle value and then increase in a symmetric fashion, and (iii) for long-tailed symmetric distributions they increase untill the middle value and then decrease in a symmetric fashion. This is instrumental in achieving robustness, considered in Chapter 8.

3.2 LINEAR REGRESSION

Consider the simple linear regression model

$$y_i = \theta_0 + \theta_1 x_i + e_i, \quad 1 \le i \le n; \tag{3.2.1}$$

θ_0 is called the intercept and θ_1 is called the slope (more popularly the regression coefficient). The x_i – values are presumed to be nonstochastic in nature and measurable without error. The errors e_i $(1 \le i \le n)$ are presumed to be random and iid (identically and independently distributed). There are numerous situations, however, where x_i $(1 \le i \le n)$ are stochastic in nature. One such important situation is considered in Section 3.11.

Maximum Likelihood: Traditionally, e_i $(1 \le i \le n)$ have been assumed to be iid normal $N(0, \sigma^2)$. Given a random sample $y_1, y_2, \ldots, y_n$ and the corresponding design values $x_1, x_2, \ldots, x_n$, the likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^n e^{-\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)^2/2\sigma^2}.$$

The ML estimators are the solutions of the likelihood equations $\partial \ln L/\partial\theta_0 = 0$, $\partial \ln L/\partial\theta_1 = 0$ and $\partial \ln L/\partial\sigma = 0$. These equations have explicit and unique solutions, the ML estimators:

$$\hat{\hat{\theta}}_0 = \bar{y} - \hat{\hat{\theta}}_1\bar{x}, \hat{\hat{\theta}}_0 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3.2.2}$$

and

$$\hat{\hat{\sigma}} = \left[\sum_{i=1}^{n}\left\{y_i - \bar{y} - \hat{\hat{\theta}}_1(x_i - \bar{x})\right\}^2 \Big/ (n-2)\right]^{1/2}; \tag{3.2.3}$$

$\bar{y} = \sum_{i=1}^{n} y_i/n$ and $\bar{x} = \sum_{i=1}^{n} x_i/n$. The divisor $n - 2$ in (3.2.3) replacing n renders $\hat{\hat{\sigma}}^2$ unbiased, i.e., $E(\hat{\hat{\sigma}}^2) = \sigma^2$. The exact variances and the covariance of $\hat{\hat{\theta}}_0$ and $\hat{\hat{\theta}}_1$ are given by $I^{-1}(\theta_o, \theta_1)$, where

$$I(\theta_0, \theta_1) = \frac{1}{\sigma^2}\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}. \tag{3.2.4}$$

The Fisher information matrix is

$$I(\theta_0, \theta_1, \sigma) = \frac{1}{\sigma^2}\begin{pmatrix} n & \sum_{i=1}^{n} x_i & 0 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & 0 \\ 0 & 0 & 2n \end{pmatrix}. \tag{3.2.5}$$

Thus, $Var(\hat{\hat{\sigma}}) \cong \sigma^2/2n$ asymptotically which agrees with the equation (1.2.9) since $\lambda_4 = 0$ for a normal distribution. Morever, $\hat{\hat{\theta}}_1$ (being a linear function of y_i, $1 \le i \le n$) is normally distributed with mean θ_1 and variance $\sigma^2/\sum_{i=1}^{n}(x_i - \bar{x})^2$. If $\theta_1 = 0$, then $\hat{\hat{\theta}}_1 Q/\sigma^2$ has a chi-square distribution with one degree of freedom; $Q = \sum_{i=1}^{n}(x_i - \bar{x})y_i$. It may be noted that in view of $\partial \ln L/\partial\theta_0 = 0$, $\partial \ln L/\partial\theta_1$ can be reorganized to assume the form

$$\frac{\partial \ln L}{\partial\theta_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sigma^2}(\hat{\hat{\theta}}_1 - \theta_1). \tag{3.2.6}$$

Therefore, $\hat{\hat{\theta}}_1$ is the MVB estimator. Note also the identity

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 \equiv \hat{\hat{\theta}}_1 Q + \sum_{i=1}^{n}\{(y_i - \bar{y} - \hat{\hat{\theta}}_1(x_i - \bar{x})\}^2, \; Q = \sum_{i=1}^{n}(x_i - \bar{x})y_i. \tag{3.2.7}$$

Consequently, $(n-2)\hat{\hat{\sigma}}^2/\sigma^2$ has a chi-square distribution with $n-2$ degrees of freedom, and $\hat{\hat{\theta}}_1 Q$ and $\hat{\hat{\sigma}}^2$ are independently distributed. Therefore, the maximum likelihood methodology has in this case all the beauty. Unfortunately, however, this beauty is lost as soon as we move on to non-normal distributions which are more prevalent in practice. We illustrate this in Sections 3.3–3.9.

Least Squares: No distributional assumptions as such are made in applying the least squares methodology. Under the assumption that e_i $(1 \le i \le n)$ are iid, the LS estimators are obtained by minimizing the error SS (sum of squares)

$$\sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)^2.$$

The resulting estimators $\tilde{\theta}_0$ and $\tilde{\theta}_1$, under the assumption of normality $N(0, \sigma^2)$, are exactly the same as the ML estimators (3.2.2). The LS estimator of σ^2 is defined as

$$\tilde{\sigma}^2 = \min \sum_{i=1}^{n} e_i^{\,2}/(n-r) \text{ (r is the number of parameters estimated, besides } \sigma) \tag{3.2.8}$$

$$= \sum_{i=1}^{n}\{y_i - \bar{y} - \tilde{\theta}_1 (x_i - \bar{x})\}^2/(n-2);$$

$\tilde{\sigma}$ is exactly the same as the ML estimator (3.2.3). Under the assumption of normality, the LS estimators have all the desirable properties. We will see, however, that the LS estimators have low efficiencies for non-normal distributions. In passing, the following generalization of the least squares methodology may be noted.

Weighted LS: Suppose that the random errors e_i $(1 \le i \le n)$ in (3.2.1) are independently distributed with a common mean $E(e_i) = a\sigma$ and variances $\mathrm{Var}(e_i) = V_i\sigma^2$. Write $w_i = 1/V_i$ $(1 \le i \le n)$. The weighted LS estimators of θ_0 and θ_1 are obtained by minimizing

$$\sum_{i=1}^{n} w_i e_i^{\,2} = \sum_{i=1}^{n} w_i (y_i - \theta_0 - \theta_1 x_i)^2. \tag{3.2.9}$$

This gives $$\tilde{\theta}_1 = \sum_{i=1}^{n} w_i (x_i - \bar{x}) y_i \Big/ \sum_{i=1}^{n} w_i (x_i - \bar{x})^2, \bar{\theta}_0 = \bar{y} - \tilde{\theta}_1 \bar{x} \tag{3.2.10}$$

and $$\tilde{\sigma}^2 = \sum_{i=1}^{n} w_i (y_i - \bar{y} - \tilde{\theta}_1 (x_i - \bar{x}))^2/(n-2);$$

$$\bar{y} = \sum_{i=1}^{n} w_i y_i \Big/ \Big(\sum_{i=1}^{n} w_i\Big) \quad \text{and} \quad \bar{x} = \sum_{i=1}^{n} w_i x_i \Big/ \Big(\sum_{i=1}^{n} w_i\Big).$$ Realize that

$$E(\tilde{\theta}_1) = \theta_1 \quad \text{and} \quad \mathrm{Var}(\tilde{\theta}_1) = \sigma^2 \Big/ \sum_{i=1}^{n} w_i (x_i - \bar{x})^2. \tag{3.2.11}$$

The LS estimator $\tilde{\theta}_0$, however, needs a bias-correction. Specifically,

$$\tilde{\theta}_0 = \bar{y} - \tilde{\theta}_1 \bar{x} - a\tilde{\sigma} \tag{3.2.12}$$

is the bias-corrected estimator.

A further extension is the generalized least squares estimation but that has essentially been discussed in Chapter 2 (Section 2.7).

3.3 THE WEIBULL DISTRIBUTION

In the linear model (3.2.1) suppose that e_i are iid and have a Weibull distribution with shape parameter p(>0),

$$W(p,\sigma): \frac{p}{\sigma^p} e^{p-1} \exp\left\{-\left(\frac{e}{\sigma}\right)^p\right\}, \; 0 < e < \infty. \tag{3.3.1}$$

The mean and variance are

$$E(e) = \Gamma(1 + 1/p)\sigma \quad \text{and} \quad Var(e) = \{\Gamma(1 + 2/p) - \Gamma^2(1 + 1/p)\}\sigma^2. \tag{3.3.2}$$

To have an idea about the nature of the Weibull W(p, σ), the values of its skewness $\mu_3/\mu_2^{3/2}$ and kurtosis μ_4/μ_2^2 are given below:

p =	1.5	2	2.5	3	4	6
Skewness	1.064	0.631	0.358	0.168	– 0.087	– 0.158
Kurtosis	4.365	3.246	2.858	2.705	2.752	2.538

Clearly, W(p, σ) represents a wide variety of skew distributions, both with kurtosis greater than as well as less than 3.

In the context of linear regression, we are primarily interested in values of $p \geq 1.3$. The reason is that $\theta_0 + \theta_1 x + E(e)$ will often be used as a predictor of the expected response E(Y). Therefore, a value of p that results in a value of the probability

$$\text{Prob}\,\{y \geq \theta_0 + \theta_1 x + E(e)\} = \exp\{-[\Gamma(1 + 1/p)]^p\} \tag{3.3.3}$$

substantially smaller or larger than 0.5 (say, ≤ 0.4 or ≥ 0.6) is hardly of any interest. If e has the Weibull distribution W(p, σ) above, then the values of this probability are as follows:

p =	0.5	1.0	1.1	1.2	1.3	1.5	2.0	3.0	6.0
Prob =	0.24	0.37	0.38	0.39	0.41	0.42	0.46	0.49	0.53

We are, therefore, primarily interested in values of $p \geq 1.3$; see also Cohen and Whitten (1988) who state that in most applications p > 1 (increasing failure rate). Writing

$$z_i = e_i/\sigma = (y_i - \theta_0 - \theta_i x_i)/\sigma \quad (1 \leq i \leq n),$$

the likelihood equations are

$$\frac{\partial \ln L}{\partial \theta_0} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} z_i^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_i^{p-1} = 0, \tag{3.3.4}$$

$$\frac{\partial \ln L}{\partial \theta_1} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} x_i z_i^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} x_i\, z_i^{p-1} = 0 \tag{3.3.5}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} - \frac{p-1}{\sigma}\sum_{i=1}^{n} z_i z_i^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_i\, z_i^{p-1} = 0 \tag{3.3.6}$$

The last equation can be simplified to give

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{np}{\sigma} + \frac{p}{\sigma}\sum_{i=1}^{n} z_i^{\,p} = 0 \tag{3.3.7}$$

As mentioned earlier in Chapter 2 (Example 2.8), the estimator of σ obtained from (3.3.7) might not be even real, let alone having the optimality properties one would desire; see also Akkaya and Tiku (2001b) who give such an example in the context of time series. We will, therefore, use (3.3.6) in estimating σ as in the next section (Islam et al., 2001).

3.4 MODIFIED LIKELIHOOD

To obtain the MML estimators, we first order $w_i = y_i - \theta_1 x_i$ (for a given θ_1) so that

$$w_{(1)} \le w_{(2)} \le \ldots\ldots \le w_{(n)}\ ;\ w_{(i)} = y_{[i]} - \theta_1 x_{[i]} \quad (1 \le i \le n). \tag{3.4.1}$$

We define the standardized ordered variates $z_{(i)} = \{w_{(i)} - \theta_0\}/\sigma$ $(1 \le i \le n)$; $(y_{[i]}, x_{[i]})$ may be called concomitants of $z_{(i)}$ and is that pair of (y_j, x_j) values which determines $w_{(i)}$. Since complete sums are invariant to ordering, the likelihood equations can be written in terms of $z_{(i)}$, namely,

$$\frac{\partial \ln L}{\partial \theta_0} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{(i)}^{\,p-1} = 0, \tag{3.4.2}$$

$$\frac{\partial \ln L}{\partial \theta_1} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} x_{[i]} z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} x_{[i]} z_{(i)}^{\,p-1} = 0 \tag{3.4.3}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} - \frac{p-1}{\sigma}\sum_{i=1}^{n} z_{(i)}\, z_{(i)}^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{(i)} z_{(i)}^{\,p-1} = 0. \tag{3.4.4}$$

Realize the difficulty that can arise if $z_{(1)}$ tends to zero in which case (3.4.2) – (3.4.3) are not defined. This is more likely if n is large. Write $t_{(i)} = E\{z_{(i)}\}$ $(1 \le i \le n)$. Harter (1964) gives exact values of $t_{(i)}$. For $n \ge 10$, however, we use the approximate values of $t_{(i)}$ calculated from (2.8.7). To obtain the modified likelihood equations, we incorporate the linear approximations (2.8.8)-(2.8.9) in (3.4.2) – (3.4.4). That gives

$$\frac{\partial \ln L}{\partial \theta_0} \cong \frac{\partial \ln L^*}{\partial \theta_0} = -\frac{p-1}{\sigma}\sum_{i=1}^{n}\{\alpha_{i0} - \beta_{i0} z_{(i)}\} + \frac{p}{\sigma}\sum_{i=1}^{n}\{\alpha_i + \beta_i z_{(i)}\} = 0, \tag{3.4.5}$$

$$\frac{\partial \ln L}{\partial \theta_1} \cong \frac{\partial \ln L^*}{\partial \theta_1} = -\frac{p-1}{\sigma}\sum_{i=1}^{n} x_{[i]}\{\alpha_{i0} - \beta_{i0} z_{(i)}\} + \frac{p}{\sigma}\sum_{i=1}^{n} x_{[i]}\{\alpha_i + \beta_i z_{(i)}\} = 0 \tag{3.4.6}$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} - \frac{p-1}{\sigma}\sum_{i=1}^{n} z_{[i]}\{\alpha_{i0} - \beta_{i0} z_{(i)}\} + \frac{p}{\sigma}\sum_{i=1}^{n} z_{[i]}\{\alpha_i + \beta_i z_{(i)}\} = 0. \tag{3.4.7}$$

The solutions of these equations are the following MML estimators ($p > 1$):

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{[.]} - (\Delta/m)\hat{\sigma}, \hat{\theta}_1 = K - D\hat{\sigma} \tag{3.4.8}$$

and

$$\hat{\sigma} = \{-B + \sqrt{(B^2 + 4nC)}\} / 2\sqrt{\{n(n-2)\}}; \tag{3.4.9}$$

$$\delta_i = (p-1)\beta_{i0} + p\beta_i, \Delta_i = (p-1)\alpha_{10} - p\alpha_i, m = \sum_{i=1}^{n} \delta_i, \Delta = \sum_{i=1}^{n} \Delta_i, \tag{3.4.10}$$

$$\bar{y}_{[.]} = (1/m)\sum_{i=1}^{n} \delta_i y_{[i]}, \quad \bar{x}_{[.]} = (1/m)\sum_{i=1}^{n} \delta_i x_{[i]},$$

$$K = \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} / \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2,$$

$$D = \sum_{i=1}^{n} \Delta_i (x_{[i]} - \bar{x}_{[.]}) / \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2$$

$$B = \sum_{i=1}^{n} \Delta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\}$$

and

$$C = \sum_{i=1}^{n} \delta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\}^2 = \sum_{i=1}^{n} \delta_i (y_{[i]} - \bar{y}_{[.]})^2 - K\sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]}. \tag{3.4.11}$$

Since $\delta_i > 0$ (for all $p > 1$), the MML estimator $\hat{\sigma}$ is always real and positive.

Remark: Because of the asymptotic equivalence of the likelihood and modified likelihood equations, the MML estimators above are fully efficient (asymptotically).

Computations: The computations are carried out in two iterations. In the first iteration, $w_{(i)}$ are obtained by ordering $w_i = y_i - \tilde{\theta}_1 x_i$ $(1 \le i \le n)$ in increasing order of magnitude, $\tilde{\theta}_1 = \sum_{i=1}^{n} (x_i - \bar{x}) y_i / \sum_{i=1}^{n} (x_i - \bar{x})^2$ being the LS (least squares) estimator of θ_1. Then, $\tilde{\theta}_1$ is calculated from (3.4.9) – (3.4.11). In the second iteration, $w_{(i)}$ are obtained by ordering $w_i = y_i - \hat{\theta}_1 x_i$ $(1 \le i \le n)$. The resulting concomitants $(y_{[i]}, x_{[i]})$, $1 \le i \le n$, are used to compute the MML estimators from (3.4.8) – (3.4.11). Only two iterations are needed for the estimates to stabilise sufficiently enough (Islam et al., 2001; Tiku et al., 2001). The reason is that the MML estimators above only depend on the concomitants $(y_{[i]}, x_{[i]})$ and the concomitant indices [i] are determined by the relative magnitudes, not necessarily the true values, of w_i $(1 \le i \le n)$.

The following result is true because of the asymptotic equivalence of the ML and MML estimators (Chapter 2, Appendix A2).

THEOREM 3.1: For $p > 2$, the MML estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ are asymptotically fully efficient, i.e., they are asymptotically unbiased and their variance-covariance matrix (asymptotic) is $I^{-1}(\theta_0, \theta_1, \sigma)$, where ($p > 2$)

$$I = \frac{np^2}{\sigma^2} \times \begin{bmatrix} \left(1-\frac{1}{p}\right)^2 \Gamma\left(1-\frac{2}{p}\right) & \left(1-\frac{1}{p}\right)^2 \Gamma\left(1-\frac{2}{p}\right)\sum_{i=1}^{n} x_i/n & \Gamma\left(2-\frac{1}{p}\right) \\ \left(1-\frac{1}{p}\right)^2 \Gamma\left(1-\frac{2}{p}\right)\sum_{i=1}^{n} x_i/n & \left(1-\frac{1}{p}\right)^2 \Gamma\left(1-\frac{2}{p}\right)\sum_{i=1}^{n} x_i^2/n & \Gamma\left(2-\frac{1}{p}\right)\sum_{i=1}^{n} x_i/n \\ \Gamma\left(2-\frac{1}{p}\right) & \Gamma\left(2-\frac{1}{p}\right)\sum_{i=1}^{n} x_i/n & 1 \end{bmatrix} \quad (3.4.12)$$

is the Fisher information matrix consisting of the elements $-E(\partial^2 \ln L/\partial\theta_0^2)$, $-E(\partial^2 \ln L/\partial\theta_0\partial\theta_1)$, etc. These elements are worked out in Appendix 3A.

If x_i are such that their mean $\bar{x} = 0$ (e.g., the design is symmetric in the interval –1 to 1), then it follows from (3.4.12) that $\hat{\theta}_1$ is uncorrelated with $\hat{\theta}_0$ and $\hat{\sigma}$ for large n. Or, if x_i are measured from their mean $\bar{x}$ and the design points are taken to be $(x_i - \bar{x})$ $(1 \le i \le n)$, then $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$. In that case the elements I_{12} and I_{23} are both zero and $\hat{\theta}_1$ is uncorrelated with $\hat{\theta}_0$ and $\hat{\sigma}$ (for large n) and $V(\hat{\theta}_1) \cong I_{22}^{-1}$,

$$I_{22} = (p-1)^2\,\Gamma(1-2/p)\sum_{i=1}^{n}(x_i - \bar{x})^2/\sigma^2;$$

the result about the variance of $\hat{\theta}_1$ is, however, true irrespective of whether x_i $(1 \le i \le n)$ are measured from their mean $\bar{x}$ or not.

Variances: The elements of I^{-1} give the asymptotic variances and covariances of the MML estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$. In particular,

$$\text{Var}(\hat{\theta}_0) \cong \frac{\sigma^2}{n(p-1)}\left\{\frac{1}{\Gamma(1-2/p) - \Gamma^2(1-1/p)} + \frac{n\bar{x}^2}{\Gamma(1-2/p)\sum_{i=1}^{n}(x_i - \bar{x})^2}\right\} \quad (3.4.13)$$

$$\text{Var}(\hat{\theta}_1) \cong \frac{\sigma^2}{(p-1)^2\,\Gamma(1-2/p)\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (3.4.14)$$

and

$$\text{Var}(\hat{\sigma}) \cong \frac{\Gamma(1-2/p)\sigma^2}{np^2\{\Gamma(1-2/p) - \Gamma^2(1-1/p)\}} \quad (3.4.15)$$

Realize that the variance of $\hat{\sigma}$ does not involve the design points x_i $(1 \le i \le n)$.

Since the variance of $\hat{\theta}_1$ is inversely proportional to $\sum_{i=1}^{n}(x_i - \bar{x})^2$, a design which is optimal for normal is also optimal for the Weibull error distribution, at any rate for large n. This avoids the need of re-inventing optimal designs when the error distribution is Weibull rather than normal.

Small Samples: The MVB estimators of θ_0, θ_1 and σ do not exist for small n. All estimators, therefore, will have their variances greater than the asymptotic variances given in (3.4.13) – (3.4.15). Small differences will imply high efficiency. Islam et al. (2001) simulated the means and variances of the MML estimators $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$. The bias is negligible in all the three estimators. For $n \geq 50$ and $p > 2.5$, the variances of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ are only marginally bigger than their asymptotic values (computed from equations 3.4.13-3–4.15). The values are given in Table 3.1. It can be seen that the MML estimators are highly efficient, at any rate for $p \geq 3$.

Table 3.1: Values of (n / σ^2)Variance of the MML Estimators

		$\hat{\theta}_0$		$\hat{\theta}_1$		$\hat{\sigma}$	
n	p	Asymp	Simul	Asymp	Simul	Asymp	Simul
50	2.5	0.57	0.83	1.20	1.61	0.31	0.41
	3.0	0.66	0.79	1.16	1.27	0.35	0.43
	4.0	0.66	0.70	0.78	0.78	0.41	0.42
100	2.5	0.49	0.65	1.18	1.42	0.31	0.39
	3.0	0.59	0.67	1.14	1.23	0.35	0.40
	4.0	0.61	0.66	0.76	0.79	0.41	0.43

It may be noted that in applications of the Weibull distribution, large samples are needed to achieve high efficiency of estimation (Menon, 1963; Harter, 1964; Lawless, 1982; Smith, 1985).

3.5 LEAST SQUARES FOR THE WEIBULL

Assuming the Weibull distribution $W(p, \sigma)$ for the errors, the LS estimators (corrected for bias) are from (3.2.10)–(3.2.12):

$$\tilde{\theta}_0 = \bar{y} - \tilde{\theta}_1\bar{x} - \Gamma(1 + /p)\tilde{\sigma}, \tilde{\theta}_1 = \sum_{i=1}^{n} (x_i - \bar{x})y_i / \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3.5.1}$$

and

$$\tilde{\sigma} = \{\sum_{i=1}^{n} (y_i - \bar{y} - \tilde{\theta}_1 (x_i - \bar{x}))^2/(n-2)(\Gamma(1 + 2/p) - \Gamma^2(1 + 1/p))\}^{1/2}; \tag{3.5.2}$$

$\bar{y} = \sum_{i=1}^{n} y_i/n$ and $\bar{x} = \sum_{i=1}^{n} x_i/n$. No such bias correction is needed in the MML estimators; they are, in fact, self bias-correcting. Consider, in particular, the LS estimator $\tilde{\theta}_1$, which is of primary interest from a practical point of view. Now,

Table 3.2: Values of the relative efficiencies of the LS estimators

n	E_1	E_2	E_3	E_1	E_2	E_3
		p = 1.3			p = 1.5	
10	97	109	61	61	69	70
20	79	94	58	52	56	64
30	77	94	58	45	47	65
50	70	81	57	40	40	59
100	70	75	56	30	34	61
		p = 2.0			p = 6.0	
10	76	86	79	95	94	92
20	75	79	78	96	94	94
30	68	72	77	96	93	94
50	64	67	73	96	93	93
100	55	60	67	96	91	92

$$E(\tilde{\theta}_1) = \theta_1 \quad \text{and} \quad \text{Var}(\tilde{\theta}_1) = \{\Gamma(1 + 2/p) - \Gamma^2(1 + 1/p)\}\sigma^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 . \tag{3.5.3}$$

The asymptotic relative efficiency of the LS estimator $\tilde{\theta}_1$ is, therefore,

$$\text{RE}(\tilde{\theta}_1) = 100\{\text{Var}(\hat{\theta}_1)/\text{Var}(\tilde{\theta}_1)\}$$
$$= 100\left\{\frac{1}{(p-1)^2\,\Gamma(1-2/p)\{\Gamma(1+2/p) - \Gamma^2(1+1/p)\}}\right\} \tag{3.5.4}$$

which assumes the values 67, 88, 91 and 81 percent for p = 2.5, 3.0, 6.0 and 10.0, respectively. Clearly, the LS estimator $\tilde{\theta}_1$ is considerably less efficient than the MML estimator.

The LS estimators $\tilde{\theta}_0$ and $\tilde{\sigma}$ are, of course, asymptotically unbiased. Islam et al.(2001) simulated the means of the LS and the MML estimators and the relative efficiencies of the LS estimators, namely,

$$E_1 = 100\{\text{Var}(\hat{\theta}_0)/\text{Var}(\tilde{\theta}_0)\}, \quad E_2 = 100\{\text{Var}(\hat{\theta}_1)/\text{Var}(\tilde{\theta}_1)\}$$

and $$E_3 = 100\{\text{Var}(\hat{\sigma})/\text{Var}(\tilde{\sigma})\}. \tag{3.5.5}$$

The bias in the estimators is negligible and, therefore, not reported. The values of E_1, E_2 and E_3 are given in Table 3.2. The design points x_i $(1 \le i \le n)$ were generated from a uniform(0, 1) distribution only once and were common to all the N = [100,000/n] samples $(y_1, y_2, \ldots, y_n)$ generated to satisfy the model (3.2.1) with Weibull W(p, σ) iid errors e_i $(1 \le i \le n)$. It can be seen that the LS estimators are considerably less efficient than the MML estimators. For design points generated from the normal N(0, 1), the relative efficiencies are essentially the same. In fact, the relative efficiencies are essentially the same as those in Table 3.2, irrespective of the nature of the design points x_i $(1 \le i \le n)$.

The following results are true which will be used later for hypothesis testing, e.g., H_0: $\theta_1 = 0$.

Lemma 3.1: Conditionally (σ known), $\hat{\theta}_1(\sigma) = K - D\sigma$ is asymptotically the MVB estimator and is normally distributed with mean θ_1 and variance

$$\sigma^2 \Big/ \sum_{i=1}^{n} \delta_i \, (x_{[i]} - \bar{x}_{[.]})^2 \quad (p > 2). \tag{3.5.6}$$

Proof: In view of (3.4.5), $\partial \ln L^*/\partial\theta_1$ can be re-organized to assume the form

$$\frac{\partial \ln L}{\partial \theta_1} \cong \frac{\partial \ln L^*}{\partial \theta_1} = \frac{\sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]})^2}{\sigma^2} \{(K - D\sigma) - \theta_1\}. \tag{3.5.7}$$

The result then follows from the fact that $\partial \ln L^*/\partial\theta_1$ is asymptotically equivalent to $\partial \ln L/\partial\theta_1$ and the third and higher derivatives of ln L* are zero (Section 2.9). Also ($p > 2$)

$$\frac{1}{\sigma^2} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta_i \, (x_{[i]} - \bar{x}_{[.]})^2 = \frac{(p-1)^2}{n\sigma^2} \Gamma\left(1 - \frac{2}{p}\right) \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{3.5.8}$$

Since the expression on the left hand side gives more accurate approximations to the true values, we use (3.5.6) for the variance.

Lemma 3.2: Conditionally (θ_1 known), $\hat{\sigma}(\theta_1)$ is asymptotically the MVB estimator of σ and $(n-1)\,\hat{\sigma}^2(\theta_1)$ is a multiple of chi-square.

Proof: In view of (3.4.5), $\partial \ln L^*/\partial\sigma$ can be re-organized to assume the form

$$\frac{\partial \ln L^*}{\partial \sigma} = -\frac{n}{\sigma^3} \left(\frac{-B_0 + \sqrt{(B_0{}^2 + 4nC_0)}}{2n} - \sigma \right) \left(\frac{-B_0 - \sqrt{(B_0{}^2 + 4nC_0)}}{2n} - \sigma \right) \tag{3.5.9}$$

where B_0 and C_0 are exactly the same as B and C in (3.4.11), respectively, with K replaced by θ_1. Since the only admissible root of (3.5.9) is

$$\sigma^2(\theta_1) = \left\{ -B_0 + \sqrt{(B_0{}^2 + 4nC_0)} \right\} \Big/ 2n,$$

the result follows (Section 2.9). Also, $B_0/\sqrt{(nC_0)} \cong 0$ for large n.

Consequently, $$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{n}{\sigma^3}\left(\frac{C_0}{n} - \sigma^2 \right). \tag{3.5.10}$$

This equation is structurally of the same form as (2.9.4). Therefore, $\hat{\sigma}^2(\theta_1) \cong C_0/(n-1)$ is the MVB of σ^2 (asymptotically). It also follows from the results given in Section 2.9 that $(n-1)\ \hat{\sigma}^2\ (\theta_1)/\sigma^2$ is a chi-square with $n-1$ degrees of freedom, for large n. The distribution of $(n-2)\ \hat{\sigma}^2/\sigma^2$ is referred to a chi-square with $n-2$ degrees of freedom. The result can be improved as in (2.11.16) and (2.11.19).

The results of the two lemmas above will be used later in Section 3.13 for testing an assumed value of θ_1.

Example 3.1: Consider the following data (Johnson and Johnson, 1979) which represents the ordered survival times (the number of days/1000) of 43 patients suffering from granulocytic leukemia:

$e_{(i)}$ = 0.007, 0.047, 0.058, 0.074, 0.177, 0.232, 0.273, 0.285, 0.317, 0.429,
0.440, 0.445, 0.455, 0.468, 0.495, 0.497, 0.532, 0.571, 0.579, 0.581,
0.650, 0.702, 0.715, 0.779, 0.881, 0.900, 0.930, 0.968, 1.077, 1.109,
1.314, 1.334, 1.367, 1.534, 1.712, 1.784, 1.877, 1.886, 2.045, 2.056,
2.260, 2.429, 2.509

It is reasonable to assume that the data comes from $W(p, \sigma)$, with $p = 2.5$; this will be illustrated in Chapter 9.

Islam et al. (2001) introduced a design variable x_i by taking

$$y_i = x_i + e_i, \quad 1 \le i \le 43, \tag{3.5.11}$$

with the following 43 values of x_i (generated from a uniform distribution):

0.00, 0.08, 0.60, 0.89, 0.97, 0.19, 0.52, 0.40, 0.26, 0.74, 0.09, 0.56, 0.58,
0.81, 0.59, 0.51, 0.88, 0.99, 0.73, 0.97, 0.30, 0.43, 0.90, 0.65, 0.90, 0.96,
0.16, 0.86, 0.91, 0.29, 0.94, 0.42, 0.31, 0.52, 0.40, 0.79, 0.69, 0.54, 0.59,
0.09, 0.61, 0.43, 0.60;

the model here is the linear model (3.2.1) with $\theta_0 = 0$ and $\theta_1 = 1$, σ being unknown. We pretend that θ_0 and θ_1 are not known and proceed to calculate the MML estimate of θ_1 and its standard error. From the equations (3.4.8)–(3.4.9) and (3.5.1)–(3.5.2), we get the following estimates:

	Estimate θ_1	Standard Error θ_1
LS	1.000	± 0.323
MML	1.025	± 0.211

The two estimates of θ_1 are both close to the true value ($\theta_1 = 1$) but the standard error of the MML estimate is 65% of that of the LS estimate. This illustrates the enormous advantage that the modified likelihood has over the least squares methodology.

The methodology above readily extends to situations where the errors in (3.2.1) have any other skew distribution, for example, Generalized Logistic. This is illustrated in Example 3.4. See also Chapter 11.

3.6 SHORT-TAILED SYMMETRIC FAMILY

Short-tailed symmetric distributions are not only important on their own but are particulary useful for modelling inliers (Tiku et al., 2001; Akkaya and Tiku, 2004). A coherent family of short-tailed symmetric distributions was recently introduced by Tiku and Vaughan (1999), namely,

$$f(y) \propto \frac{1}{\sigma}\left\{1+\frac{\lambda}{2r}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^r \left[1+\frac{1}{2k_1}\left(\frac{y-\mu}{\sigma}\right)^2\right]^{-p}, \quad -\infty < y < \infty; \tag{3.6.1}$$

r is an integer, $\lambda = r/(r-d)$, $d < r$, $k_1 = p - 3/2$ and $p > r + 3/2$. This family has kurtosis μ_4/μ_2^2 less than 3 for all values of p. Here, we will focus on a special form of f(y) when $p = \infty$, i.e.,

$$f(z) = C_1\left(1+\frac{\lambda}{2r}z^2\right)^r \phi(z), \quad -\infty < z < \infty, \tag{3.6.2}$$

where $z = (y - \mu)/\sigma$, $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$, and C_1 is a constant to make

$$\int_{-\infty}^{\infty} f(z)\,dz = 1.$$

Realizing that
$$\int_{-\infty}^{\infty} z^{2j}\,\phi(z)\,dz = (2j)!/(2^j\,j!), \tag{3.6.3}$$

it is easy to work out the value of C_1;

$$C_1 = 1\Bigg/\left\{\sum_{j=0}^{r}\binom{r}{j}\left(\frac{\lambda}{2r}\right)^j\frac{(2j)!}{2^j\,(j)!}\right\}. \tag{3.6.4}$$

It is also easy to work out the even moments of Z from (3.6.3), all odd moments being zero because of symmetry. In particular, we have the following values of the standard deviation and the kurtosis; $d < r$ ($\lambda > 0$):

r	d =	– 0.5	0.0	0.5	1.0	1.5	2.5	3.5
2	$\sqrt{\mu_2}/\sigma$	1.3572	1.4272	1.5275	1.6787	1.9149	—	—
	Kurtosis	2.559	2.437	2.265	2.026	1.711	—	—
4	$\sqrt{\mu_2}/\sigma$	1.5361	1.6051	1.6910	1.7990	1.9358	2.3189	2.7944
	Kurtosis	2.464	2.370	2.255	2.118	1.957	1.591	1.297

It is seen that the family (3.6.2) represents a wide variety of short-tailed symmetric distributions. For a given r, the kurtosis decreases as d increases. The minimum and maximum values the kurtosis assumes are 1 and 3, respectively. For $d \le 0$, the distributions (3.6.1) are unimodal. For $d > 0$, however, they are generally multimodal.

Suppose that e_i ($1 \le i \le n$) in the linear model (3.2.1) have one of the distributions in (3.6.2). The likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^n\left[\prod_{i=1}^{n}\left\{1+\frac{\lambda}{2r}z_i^2\right\}^r\right]e^{-\sum_{i=1}^{n} z_i^2/2}; \tag{3.6.5}$$

$$z_i = (w_i - \theta_0)/\sigma, \quad w_i = y_i - \theta_1 x_i \quad (1 \le i \le n).$$

Again, the ML estimators are elusive since the likelihood equations have no explicit solutions and solving them by iteration is very problematic. To apply the modified likelihood methodology, we first express the likelihood equations in terms of the ordered variates (for a given θ_1)

$$w_{(i)} = y_{[i]} - \theta_1 x_{[i]}, \quad 1 \le i \le n \tag{3.6.6}$$

that gives
$$\frac{\partial \ln L}{\partial\theta_0} = \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)} - \frac{\lambda}{\sigma}\sum_{i=1}^{n} g\{z_{(i)}\} = 0, \tag{3.6.7}$$

$$\frac{\partial \ln L}{\partial\theta_1} = \frac{1}{\sigma}\sum_{i=1}^{n} x_{[i]}\,z_{(i)} - \frac{\lambda}{\sigma}\sum_{i=1}^{n} x_{[i]}\,g\{z_{(i)}\} = 0 \tag{3.6.8}$$

and
$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_{(i)}^2 - \frac{\lambda}{\sigma}\sum_{i=1}^{n} z_{(i)}\, g\{z_{(i)}\} = 0; \tag{3.6.9}$$
$$g(z) = z/\{1 + (\lambda/2r)z^2\}. \tag{3.6.10}$$

Let $t_{(i)} = E\{z_{(i)}\}$ $(1 \le i \le n)$ be the expected values of the ordered variates $z_{(i)}$. The exact values of $t_{(i)}$ are not available at the moment. For $n \ge 10$, however, their approximate values obtained from the equations

$$\int_{-\infty}^{t_{(i)}} f(z)\, dz = \frac{i}{n+1}, \quad 1 \le i \le n, \tag{3.6.11}$$

are used (Tiku et al., 2001). The values of $t_{(i)}$ for r = 4 and d = 0, 1, and n = 10, 12, 15, 20, 30, 50 and 100, are given in Appendix 3B; the values for r = 2 and d = 0 are given in Tiku et al. (2001). The values might have an error of up to 5 units in the last decimal place.

As said earlier, solving (3.6.7) – (3.6.9) by iteration is problematic. From the first two terms of a Taylor series expansion, however, we obtain the linear functional

$$g\{z_{(i)}\} \cong \alpha_i + \gamma_i z_{(i)}, \quad 1 \le i \le n, \tag{3.6.12}$$

where $\alpha_i = (\lambda/r)t_{(i)}^3/\{1 + (\lambda/2r)t_{(i)}^2\}^2$

and $\gamma_i = \{1 - (\lambda/2r)t_{(i)}^2\}/\{1 + (\lambda/2r)t_{(i)}^2\}^2.$ (3.6.13)

We also write $\beta_i = 1 - \lambda\gamma_i, \quad 1 \le i \le n.$ (3.6.14)

It may be noted that $\beta_i \ge 0$ $(1 \le i \le n)$ for $d \le 0$ (i.e., $\lambda \le 1$). For $d > 0$, we will give an alternative linear functional which can be used if $\beta_i < 0$ (for some i). This ensures that the MML estimator of σ is always real and positive.

The coefficients β_i $(1 \le i \le n)$ have inverted umbrella ordering, i.e., they constitute a decreasing sequence until the middle value and increase in a symmetric fashion. For n = 20, r = 2 and d = 0, for example, the first ten coefficients are, $\beta_{n-i+1} = \beta_i$:

1.06 0.97 0.87 0.74 0.61 0.46 0.31 0.17 0.070 0.008

Thus, the order statistics in the middle receive small weights. This is instrumental in achieving robustness to short tails and to inliers (Chapter 8).

3.7 MML ESTIMATORS FOR SHORT-TAILED FAMILY

Incorporating (3.6.12) in (3.6.7)-(3.6.9), we obtain the modified likelihood equations. The solutions of these equations are the MML estimators:

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1\bar{x}_{[.]}, \quad \hat{\theta}_1 = K - \lambda D\hat{\sigma} \tag{3.7.1}$$

and
$$\hat{\sigma} = \left\{-\lambda B + \sqrt{((\lambda B)^2 + 4nC)}\right\}/2\sqrt{\{n(n-2)\}}; \tag{3.7.2}$$

$\bar{y}_{[.]}$, $\bar{x}_{[.]}$, m and K have exactly the same expressions as in (3.4.11) with δ_i replaced by β_i $(1 \le i \le n)$, and

$$D = \sum_{i=1}^{n} \alpha_i x_{[i]} \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2, \; B = \sum_{i=1}^{n} \alpha_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\}, \tag{3.7.3}$$

and
$$C = \sum_{i=1}^{n} \beta_i (y_{[i]} - \bar{y}_{[.]})^2 - K\sum_{i=1}^{n} \beta_i \,(x_{[i]} - \bar{x}_{[.]})y_{[i]};$$

α_i and β_i are given in (3.6.13)–(3.6.14). For $d \le 0$ ($\lambda \le 1$), β_i is positive for all i = 1, 2,...,n. Consequently, $\hat{\sigma}$ is always real and positive.

The estimators above are calculated as explained in Section 3.4, and the estimates stabilize sufficiently enough in two iterations.

Alternative expression: As said earlier, the coefficients β_i ($1 \le i \le n$) form a sequence of values which decrease until the middle value and then increase in a symmetric fashion. For $d > 0$ ($\lambda > 1$), however, a few β_i coefficients in the middle can assume negative values. Consequently, the MML estimator $\hat{\sigma}$ can cease to be real. To rectify this situation, we use an alternative linear approximation (Tiku et al., 2001) as follows.

Since the function g(z) is bounded, and so are the coefficients α_i and γ_i, we are in a position to recast the linear functional (3.6.12). Now,

$$g\{z_{(i)}\} \cong \alpha_i^* + \beta_i^* z_{(i)}, \quad 1 \le i \le n, \tag{3.7.4}$$

where $$\alpha_1^* = \frac{(\lambda/r)t_{(i)}^3 + (1 - 1/\lambda)t_{(i)}}{\{1 + (\lambda/2r)t_{(i)}^2\}^2} \quad \text{and} \quad \gamma_1^* = \frac{(1/\lambda) - (\lambda/2r)t_{(i)}^2}{\{1 + (\lambda/2r)t_{(i)}^2\}^2} \quad (\lambda > 1). \tag{3.7.5}$$

Realize that $\alpha_i^* + \gamma_i^* z_{(i)} \cong \alpha_i + \gamma_i z_{(i)}$ since $z_{(i)} - t_{(i)} \cong 0$ (asymptotically). We also write $\beta_i^* = 1 - \lambda\gamma_i^*$ ($1 \le i \le n$), which is always positive if $\lambda > 1$. Incidentally, note that

$$\alpha_i = -\alpha_{n-i+1}, \quad \alpha_i^* = -\alpha_{n-i+1}, \quad \sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \alpha_i^* = 0; \tag{3.7.6}$$

$$\beta_i = \beta_{n-i+1}, \; \beta_i^* = \beta_{n-i+1};$$

this follows from the fact that for any symmetric distribution $t_{(i)} = -t_{(n-i+1)}$. To compute the MML estimators, we proceed as follows:

For $\lambda = r/(r - d) > 1$, we compute the MML estimators from (3.7.1)–(3.7.3) with α_i and β_i (equations 3.6.13–3.6.14) replaced by α_i^* and β_i^* respectively. Otherwise, they are computed from (3.7.1)–(3.7.3). For d = 0 ($\lambda = 1$), $\alpha_i = \alpha_i^*$ and $\beta_i = \beta_i^*$ ($1 \le i \le n$). Realize that, irrespective of the value of d, the MML estimators are unique and explicit. This is a very desirable property from a theoretical as well as a practical point of view. As for the Weibull W(p, σ) distribution considered earlier, the MML estimators (3.7.1)–(3.7.2) are asymptotically fully efficient. Their asymptotic variances and covariances are given by $I^{-1}(\theta_0, \theta_1, \sigma)$, where I is the Fisher information matrix which exists for all r and d < r:

$$I = \frac{D}{\sigma^2}\begin{pmatrix} n & \sum_{i=1}^{n} x_i & 0 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & 0 \\ 0 & 0 & nD^*/D \end{pmatrix} \tag{3.7.7}$$

$$D = 1 - \lambda E_{(0,1)} + (\lambda^2/r)E_{(1,2)} \quad \text{and} \quad D^* = 2 - 3\lambda E_{(1,1)} + (\lambda^2/r)E_{(2,2)},$$

$$E_{(u,v)} = C_1 \sum_{j=1}^{r-v} \binom{r-v}{j}\left(\frac{\lambda}{2r}\right)^j \frac{\{2(u+j)\}!}{2^{u+j}(u+j)!} \tag{3.7.8}$$

In particular, $\text{Var}(\hat{\theta}_0) \cong \frac{\sigma^2}{nD}\left\{1 + \frac{n\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right\}$, $\text{Var}(\hat{\theta}_1) \cong \frac{\sigma^2}{D\sum_{i=1}^{n}(x_i - \bar{x})^2}$ (3.7.9)

and $\text{Var}(\hat{\sigma}) \cong \sigma^2/nD^*$.

To have an idea about the constants D and D*, we give their values below for r = 2; see also equation (11.4.4):

$$D = 1 - \lambda\left[\left(1 - \frac{\lambda}{4}\right)\Big/\left\{1 + \frac{\lambda}{2} + 3\left(\frac{\lambda}{4}\right)^2\right\}\right]$$

and

$$D^* = -1 + 3\left[\left\{1 + \frac{\lambda}{2} + 11\left(\frac{\lambda}{4}\right)^2\right\}\Big/\left\{1 + \frac{\lambda}{2} + 3\left(\frac{\lambda}{4}\right)^2\right\}\right]. \quad (3.7.10)$$

It may be noted that $\hat{\theta}_0$ and $\hat{\theta}_1$ are both uncorrelated with $\hat{\sigma}$ for large n. This is due to the symmetry of the distributions in the family (3.6.2).

We do not give details, but the equations (3.7.9) give accurate values for n ≥ 50. Note that the information matrix I does not exists if d gets close to r, i.e., λ tends to infinity.

3.8 LS ESTIMATORS FOR SHORT-TAILED FAMILY

The LS estimators for the short-tailed family above are

$$\tilde{\theta}_0 = \bar{y} - \tilde{\theta}_1\bar{x}, \quad \tilde{\theta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \Big/ \sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (3.8.1)$$

and

$$\tilde{\sigma} = \sqrt{\left[\sum_{i=1}^{n}(y_i - \bar{y} - \tilde{\theta}_1(x_i - \bar{x}))^2/(n-2)\mu_2\right]};$$

$\mu_2 = E(Z^2)$ is the variance of the distribution (3.6.2). In fact,

$$\mu_2 = \sum_{j=0}^{r}\binom{r}{j}\left(\frac{\lambda}{2r}\right)^j\left[\frac{\{2(i+j)\}!}{2^{i+j}(i+j)!}\right] \div \sum_{j=0}^{r}\binom{r}{j}\left(\frac{\lambda}{2r}\right)^j\left[\frac{(2j)!}{2^j(j)!}\right]. \quad (3.8.2)$$

Relative efficiency: The asymptotic relative efficiency (ARE) of $\tilde{\theta}_1$, in particular, is

$$100\,\{\text{Var}(\hat{\theta}_1)/\text{Var}(\tilde{\theta}_1)\} = 100(1/D\mu_2). \quad (3.8.3)$$

Its values can easily be computed from (3.7.8) and (3.8.2). They are slightly smaller than the values given in Table 3.5 for n = 100. It can be seen that the LS estimators $\tilde{\theta}_0$, particulary $\tilde{\theta}_1$, have low efficiencies as compared to the MML estimators $\hat{\theta}_0$ and $\hat{\theta}_1$.

We give in Table 3.5 the simulated relative efficiencies E_1, E_2 and E_3 (defined in 3.5.5) of the LS estimators for a few representative values of r and d (Tiku et al., 2001). It is seen that the MML estimators are considerably more efficient. We show in Chapter 8 that the MML estimators are robust to short tails and inliers, but not the LS estimators. It is, therefore, advantageous to use the MML estimators in place of the LS estimators.

Because of the symmetry of (3.6.2) and the following identity

Table 3.5: Relative efficiencies of the LS estimators for the short-tailed symmetric family

N	E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3
		r = 2, a = – 0.5			r = 2, a = 0			r = 2, a = 1.0	
10	97	99	92	96	99	91	93	98	97
20	96	96	94	91	93	93	80	76	94
30	94	96	94	88	92	93	73	70	92
50	93	95	94	89	89	93	69	64	92
100	93	90	95	85	87	93	61	56	92
		r = 4, a = – 1.0			r = 4, a = 0			r = 4, a = 2	
10	97	98	91	94	96	89	89	95	95
20	95	95	93	89	91	92	68	62	88
30	93	95	93	86	89	91	57	53	86
50	93	94	93	84	86	91	52	45	86
100	93	90	93	82	81	91	44	37	86

$$\sum_{i=}^{n} \beta_i (y_{[i]} - \bar{y}_{[.]})^2 \equiv \hat{\theta}_1 Q + \sum_{i=1}^{n} \beta_i \{y_{[i]} - \bar{y}_{[.]} - \hat{\theta}_1 (x_{[.]} - \bar{x}_{[.]})\}^2 , \tag{3.8.4}$$

$$Q = \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} ,$$

and factorizations exactly similar to (3.2.7), the following results are true (asymptotically);

(i) $\hat{\theta}_1$ is normally distributed with mean θ_1 and variance $\sigma^2 \Big/ \left\{ D \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\}$,

(ii) $(n - 2)\hat{\sigma}^2/\sigma^2$ is distributed as chi-square (more accurately a multiple of chi-square) with (n – 2) degrees of freedom and

(iii) $\hat{\theta}_1$ and $\hat{\sigma}$ are independently distributed.

The results (i)-(iii) will be used later in Section 3.13 for hypothesis testing.

3. 9 LONG-TAILED SYMMETRIC FAMILY

A rich family of long-tailed symmetric distributions is given by

$$f(e) \propto \frac{1}{\sigma} \left\{ 1 + \frac{e^2}{k\sigma^2} \right\}^{-p}, \quad -\infty < e < \infty; \tag{3.9.1}$$

$k = 2p - 3$ and $p \ge 2$, the same as (1.2.4) with $(y - \mu)/\sigma = e/\sigma = z$. It is easy to show that $E(e) = 0$, $Var(e) = \sigma^2$ and kurtosis

$$\mu_4/\mu_2^2 = 3(p - 3/2)/(p - 5/2) \tag{3.9.2}$$

which assumes values ∞, 9, 4.2, 3.4 and 3 for p = 2.5, 3.5, 5, 10 and ∞, respectively. Realize that the kurtosis is always greater than 3 (equal to 3 for $p = \infty$ in which case the distribution reduces to a normal). It may be noted that $t = \sqrt{(\nu/k)}\,(e/\sigma)$ has the Student t distribution with $\nu = 2p - 1$ degrees of freedom. The family (3.9.1) is particularly useful in modelling samples which contain outliers.

Assuming that the random errors in the linear model (3.2.1) have one of the distributions in the family (3.9.1), the likelihood equations are expressions in terms of the intractable functions

$$g\{z_{(i)}\} = z_{(i)}/\{1 + (1+k)z_{(i)}^{\;2}\}, \quad 1 \le i \le n, \tag{3.9.3}$$

where $z_{(i)} = \{w_{(i)} - \theta_0\}/\sigma$, and $w_{(i)} = y_{[i]} - \theta_1 x_{[i]}$ as in (3.6.6). The likelihood equations are, therefore, problematic (Tiku and Suresh, 1992a; Vaughan, 1992; Islam et al., 2001). Moreover, Puthenpura and Sinha (1986) have shown that iterations with likelihood equations might never converge if the data contains outliers.

To formulate the modified likelihood equations, we use the first two terms of a Taylor series expansion of $g\{z_{(i)}\}$ around $t_{(i)} = E\{z_{(i)}\}$ to have

$$g\{z_{(i)}\} \cong \alpha_i + \beta_i\, z_{(i)}, \quad 1 \le i \le n; \tag{3.9.4}$$

the coefficients α_i and β_i are given in (2.3.14).

MML estimators: In the likelihood equations if we replace $g\{z_{(i)}\}$ by the linear functions (3.9.4), we obtain the modified likelihood equations. Their solutions are the following MML estimators:

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{[.]}, \quad \hat{\theta}_1 = K + D\hat{\sigma} \tag{3.9.5}$$

and

$$\hat{\sigma} = \left\{B + \sqrt{(B^2 + 4nC)}\right\}/2\sqrt{\{n(n-2)\}}; \tag{3.9.6}$$

$$m = \sum_{i=1}^{n} \beta_i,\ \bar{y}_{[.]} = (1/m)\sum_{i=1}^{n} \beta_i x_{[i]},\ \bar{x}_{[.]} = (1/m)\sum_{i=1}^{n} \beta_i x_{[i]},$$

$$K = \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]}) y_{[i]} \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2,$$

$$D = \sum_{i=1}^{n} \alpha_i x_{[i]} \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2,$$

$$B = (2p/k)\sum_{i=1}^{n} \alpha_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\} \tag{3.9.7}$$

and

$$C = (2p/k)\sum_{i=1}^{n} \beta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\}^2$$

$$= (2p/k)\left\{\sum_{i=1}^{n} \beta_i (y_{[i]} - \bar{y}_{[.]})^2 - K\sum_{i=1}^{n} \beta_i\,(x_{[i]} - \bar{x}_{[.]})\, y_{[i]}\right\}.$$

The estimators are computed in two iterations as explained earlier. If for a sample C is negative, $\hat{\sigma}$ is computed from the sample with α_i replaced by zero and β_i replaced by $\beta_i^* = \{1 + (1/k)t_{(i)}^2\}$ as explained in Section 2.4 (Chapter 2). Thus, $\hat{\sigma}$ is always real and positive.

For the same reasons as for other distributions, the MML estimators above are asymptotically fully efficient. Their asymptotic variance-covariance matrix is given by I^{-1}, where I is the Fisher information matrix:

$$I(\theta_0, \theta_1, \sigma) = \frac{p(p-1/2)}{(p+1)(p-3/2)\sigma^2}\begin{pmatrix} n & \sum_{i=1}^{n} x_i & 0 \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & 0 \\ 0 & 0 & \dfrac{2n(p-3/2)}{p} \end{pmatrix} \quad (p \geq 2), \qquad (3.9.8)$$

which is exactly of the same form as (3.7.7) and $\hat{\theta}_0$ and $\hat{\theta}_1$ are uncorrelated with $\hat{\sigma}$ for large n, irrespective of the design ponts x_i $(1 \leq i \leq n)$.

Table 3.6: Values of (a) (n/σ^2) (MVB), and (b) (n/σ^2)Variance

	$\hat{\theta}_0$		$\hat{\theta}_1$		$\hat{\sigma}$	
n	(a)	(b)	(a)	(b)	(a)	(b)
			p = 2			
20	1.71	2.31	4.21	5.52	1.00	4.19
50	1.73	1.98	4.40	4.91	1.00	2.20
100	2.05	2.26	5.67	6.27	1.00	1.58
			p = 3			
20	2.73	3.00	6.74	7.32	0.80	1.30
50	2.76	2.93	7.03	7.45	0.80	1.05
100	3.28	3.42	9.07	9.72	0.80	1.05
			p = 5			
20	3.16	3.19	7.75	7.86	0.67	0.79
50	3.22	3.37	8.20	8.38	0.67	0.74
100	3.82	3.97	10.58	11.01	0.67	0.70

The asymptotic variances of the MML estimators are

$$\mathrm{Var}(\hat{\theta}_0) \cong \frac{\sigma^2}{n}\frac{(p+1)(p-3/2)}{p(p-1/2)}\left\{1 + \frac{\bar{x}^2}{s_x^2}\right\}, \; s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad (3.9.9)$$

$$\mathrm{Var}(\hat{\theta}_1) \cong \frac{\sigma^2}{n}\frac{(p+1)(p-3/2)}{p(p-1/2)}\frac{1}{s_x^2} \quad \text{and} \quad \mathrm{Var}(\hat{\sigma}) \cong \frac{(p+1)}{p(p-1/2)}\frac{\sigma^2}{2n}.$$

The asymptotic relative efficiency of $\tilde{\theta}_1$ is

$$100\{(p+1)(p-3/2)/p(p-1/2)\} \qquad (3.9.10)$$

which assumes the values 50, 80, 93 and 100 for p = 2, 3, 5 and ∞, respectively.

Efficiency: To have an idea about the efficiency of the MML estimators, we give in Table 3.6 the simulated values of the variances of $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$. Their means are not given since the bias is negligible (Tiku et al., 2001). Also given are the values of the minimum variance bounds. The variances of $\hat{\theta}_0$ and $\hat{\theta}_1$ are only marginally bigger than the corresponding asymptotic values, for all $p \geq 2$ and $n \geq 20$. It can be concluded that both $\hat{\theta}_0$ and $\hat{\theta}_1$ are highly efficient.

For p = 2, the variance of $\hat{\sigma}$ is considerably bigger than the corresponding minimum variance bound. But then for p = 2 the population kurtosis is infinite which is a very awkward situation for the estimation of σ (Chapter 1, equation 1.2.9). For $p \geq 3$, however, the difference between the variance of $\hat{\sigma}$ and the corresponding minimum variance bound is small which implies that, like $\hat{\theta}_0$ and $\hat{\theta}_1$, $\hat{\sigma}$ is highly efficient as expected.

Real-life applications of the results above are discussed in the last chapter (Chapter 11) and in Example 3.3.

Distributions: The following asymptotic results are true because of the asymptotic equivalence of the modified likelihood and likelihood equations and a decomposition exactly similar to (3.8.4), and factorizations exactly similar to (3.2.6) and (3.5.10), and the symmetry of the distributions in the family (3.9.1); $p \geq 2$:

(i) The distribution of $\hat{\theta}_1$ is normal with mean θ_1 and variance

$$(p + 1)\,(p - 3/2)\sigma^2 \Big/ \left\{ p(p - 1/2) \sum_{i=1}^{n} (x_i - \bar{x})^2 \right\},$$

(ii) $(n - 2)\hat{\sigma}^2/\sigma^2$ is distributed as chi-square (more accurately a multiple of chi-square) with n – 2 degrees of freedom and

(iii) $\hat{\theta}_1$ and $\hat{\sigma}^2$ are independently distributed.

These results will be used later for hypothesis testing.

3.10 GENERAL LINEAR MODEL

The results above readily genaralize to the multiple linear regression model (Islam et al., 2001; Islam and Tiku, 2004)

$$y_i = \theta_0 + \theta_1 x_{1i} + \; .... + \theta_k x_{ki} + e_i, \qquad 1 \leq i \leq n. \tag{3.10.1}$$

Suppose that e_i are iid and have the Weibull distribution W(p, σ). Let $(y_{[i]}, x_{1[i]}, ...x_{k[i]})$ be the concomitants of $e_{(i)}$, $1 \leq i \leq n$. The MML estimator of θ_0 is

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1 \bar{x}_{1[.]} - ... - \hat{\theta}_k \bar{x}_{k[.]} - (\Delta/m)\hat{\sigma}. \tag{3.10.2}$$

If we write $\quad Y_{[i]} = y_{[i]} - \bar{y}_{[.]}, X_{1[i]} = x_{1[.]} - \bar{x}_{1[.]}, ..., X_{k[i]} = x_{k[i]} - \bar{x}_{k[.]},$ (3.10.3)

then the MML estimators of θ_i $(1 \leq i \leq k)$ and σ are

$$\hat{\theta} = \{X'\Gamma_\delta X)^{-1} \{X'\Gamma_\delta Y - \hat{\sigma} X'\Delta 1) \tag{3.10.4}$$

and

$$\hat{\sigma} = \{-B + \sqrt{(B^2 + 4nC)}\} \Big/ 2\sqrt{\{n(n - k - 1)\}}. \tag{3.10.5}$$

Here,
$$Y = \begin{bmatrix} Y_{[1]} \\ Y_{[2]} \\ \vdots \\ Y_{[n]} \end{bmatrix}, \quad X = \begin{bmatrix} X_{1[1]} & \cdots\cdots & X_{k[1]} \\ X_{1[2]} & \cdots\cdots & X_{k[2]} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ X_{1[n]} & \cdots\cdots & X_{k[n]} \end{bmatrix}, \tag{3.10.6}$$

$$\Gamma_\delta = \begin{bmatrix} \delta_1 & 0 \cdots\cdots & 0 \\ 0 & \delta_2 \cdots\cdots & 0 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 0 & \cdots\cdots & \delta_n \end{bmatrix}, \quad \Delta = \begin{bmatrix} \Delta_1 & 0 \cdots\cdots & 0 \\ 0 & \Delta_2 \cdots\cdots & 0 \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 0 & \cdots\cdots & \Delta_n \end{bmatrix}, \tag{3.10.7}$$

$$B = \sum_{i=1}^{n} \Delta_i \{Y_{[i]} - K_1 X_{1[i]} - \ldots - K_k X_{k[i]}\}, \tag{3.10.8}$$

and
$$C = \sum_{i=1}^{n} \delta_i Y_{[i]}^2 - K_1 Q_1 - \ldots - K_k Q_k ; \tag{3.10.9}$$

$$Q_j = \sum_{i=1}^{n} \delta_i (x_{j[i]} - \bar{x}_{[.]}) y_{[i]} \tag{3.10.10}$$

and
$$K_j = \sum_{i=1}^{n} \delta_i (x_{j[i]} - x_{j[.]}) y_{[i]} \Bigg/ \sum_{i=1}^{n} \delta_i (x_{j[i]} - \bar{x}_{j[.]})^2, \quad 1 \le i \le n.$$

The coefficients δ_i and Δ_i $(1 \le i \le n)$ are defined in (3.4.11). The estimators in (3.10.2) – (3.10.5) have essentially the same efficiency properties as for the case k = 1.

The results (3.10.2) – (3.10.10) readily extend to other distributions, e.g., the families of distributions in (3.6.1) and (3.9.1); see Chapter 11 (Appendix). For symmetric distributions, there are no terms involving $\hat{\sigma}$ in (3.10.2) and (3.10.4); the distribution theory is similar to the case when k = 1. For example, $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ have a k-variate normal distribution and $(n - k - 1)\ \hat{\sigma}^2/\sigma^2$ has a chi-square distribution with $n - k - 1$ degrees of freedom and the two are independently distributed (asymptotically), for symmetric distributions.

3.11 STOCHASTIC LINEAR REGRESSION

In the model (3.2.1), suppose that the variable X is also stochastic (Chakraborty and Srinivasan, 1992) and has a normal distribution $N(\mu_1, \sigma_1^2)$. Assume that the random error e is normal $N(0, \sigma_e^2)$, $\sigma_e^2 = \sigma_2^2 (1 - \rho^2)$. In other words, the conditional distribution of Y given X = x is normal with mean $\theta_0 + \theta_1 x = \mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1)$ and variance σ_e^2; $\theta_0 = \mu_2 - \theta_1\mu_1$ and $\theta_1 = \rho(\sigma_2/\sigma_1)$. Realize that the joint distribution of X and Y is bivariate normal. The likelihood function is

$$L \propto \prod_{i=1}^{n} \left\{\frac{1}{\sigma_1^2\ \sigma_2^2\ (1-\rho^2)}\right\}^{1/2} \exp\left\{-\frac{1}{2\sigma_1^2}(x_i - \mu_1)^2\right\}$$
$$\times \exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}[y_i - \mu_2 - \rho\,(\sigma_2/\sigma_1)(x_i - \mu_i)]^2\right\} \tag{3.11.1}$$

The solutions of the equations $\partial \ln L/\partial\mu_1 = 0$, $\partial \ln L/\partial\mu_2 = 0$ etc., are the ML estimators of μ_1, μ_2, σ_1^2, σ_1^2 and ρ, respectively:

$$\bar{x} = (1/n)\sum_{i=1}^{n} x_i, \quad \bar{y} = (1/n)\sum_{i=1}^{n} y_i, \quad s_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n-1), \tag{3.11.2}$$

$$s_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1) \quad \text{and} \quad \hat{\hat{\rho}} = s_{xy}/s_x s_y;$$

$$s_{xy} = \sum_{i=1}^{n} (x_i - \bar{x}) y_i/(n-1).$$

The estimators can also be obtained from (3.2.2)–(3.2.3) by invoking the invariance property of the ML estimators. In other words, if (X, Y) has a bivariate normal distribution

$$f(X,Y) = g(X)\,h(Y\,|\,X) \tag{3.11.3}$$

then the estimators of θ_0, θ_1 and σ_e^2 obtained from the complete likelihood function (3.11.1) are exactly the same as those obtained from the conditional likelihood function

$$L_{y|x} \propto \left(\frac{1}{\sigma_e}\right)^n \exp\left\{-\frac{1}{2\sigma_e^2}\sum_{i=1}^{n} (y_i - \theta_0 - \theta_1 x_i)^2\right\} \tag{3.11.4}$$

with μ_1 and σ_1^2 replaced by $\bar{x}$ and s_x^2, respectively. In case of a bivariate normal, therefore, it suffices to use the conditional likelihood (3.11.4) for statistical inferences about θ_0, θ_1 and σ_e^2. The estimators of μ_1, μ_2, σ_1^2, σ_2^2 and ρ are very different if the marginal distribution of X is non-normal. We illustrate this by considering an important data set as follows.

EXAMPLE 3.2: Consider the following data where X = ln U, and U represents 100 times the white blood counts and Y represents the survival times (in weeks) of patients who died of acute myelogenous leukemia (Gross and Clark, 1975);

X: 3.135 2.015 3.761 3.258 4.094 4.654 4.605 5.136 3.989 4.248 4.543 5.768
5.858 6.908 6.908 6.254

Y: 65 156 100 134 16 108 121 4 39 143 3 26
22 1 1 5

Using graphical techniques (Q-Q plots) and goodness-of-fit tests (Chapter 9), Vaughan and Tiku (2000) identified a plausible model for this data: the marginal distribution of X is the extreme-value distribution (2.6.1), (ii) the regression of Y on X is linear, i.e., $E(Y \mid X = x) = \theta_0 + \theta_1 x$, and (iii) the conditional distribution of Y given X = x is normal $N(\theta_0 + \theta_1 x, \sigma_e^2)$; $\theta_0 = \mu - \theta_1 \delta$, $\theta_1 = \rho(\sigma/\eta)$ and $\sigma_e^2 = \sigma^2(1-\rho^2)$. Here, the likelihood function is

$$L \propto \left(\frac{1}{\eta\sigma(1-\rho^2)}\right)^n \exp\left\{-\sum_{i=1}^{n} (z_i + e^{-z_i}) - \frac{1}{2\sigma^2(1-\rho^2)}\sum_{i=1}^{n} e_i^2\right\} \tag{3.11.5}$$

where $z_i = (x_i - \delta)/\eta$ and $e_i = y_i - \mu - \rho(\sigma/\eta)(x_i - \delta)$, $1 \le i \le n$. The likelihood equations for estimating the parameters δ, η, μ, σ and ρ (and θ_1) are

$$\frac{\partial \ln L}{\partial \delta} = \frac{n}{\eta} - \frac{1}{\eta}\sum_{i=1}^{n} \exp(-z_i) - \frac{\rho}{\eta\sigma(1-\rho^2)}\sum_{i=1}^{n} e_i = 0, \tag{3.11.6}$$

$$\frac{\partial \ln L}{\partial \eta} = -\frac{n}{\eta} - \frac{1}{\eta}\sum_{i=1}^{n} z_i \exp(-z_i) + \frac{1}{\eta}\sum_{i=1}^{n} z_i - \frac{\rho}{\eta\sigma(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0 \tag{3.11.7}$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2(1-\rho^2)}\sum_{i=1}^{n} e_i = 0 \tag{3.11.8}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3(1-\rho^2)}\sum_{i=1}^{n} e_i^2 + \frac{\rho}{\sigma^2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0 \tag{3.11.9}$$

and
$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma^2(1-\rho^2)^2}\sum_{i=1}^{n} e_i^2 + \frac{1}{\sigma(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0. \tag{3.11.10}$$

Also, $$\frac{\partial \ln L}{\partial \theta_1} = \frac{\eta}{\sigma^2(1-\rho^2)}\sum_{i=1}^{n} z_i e_i = 0. \tag{3.11.11}$$

The equations are intractable. To formulate modified likelihood equations, we write

$$z_{(i)} = (x_{(i)} - \delta)/\eta \quad \text{and} \quad e_{[i]} = y_{[i]} - \mu - \rho(\sigma/\eta)(x_{(i)} - \delta), \quad 1 \le i \le n; \tag{3.11.12}$$

$x_{(i)}$ are the order statistics of x_i. The fact that complete sums are invariant to ordering implies that $\sum_{i=1}^{n} e_{[i]} = 0$ and $\sum_{i=1}^{n} z_{(i)} e_{[i]} = 0$, from equations (3.11.8) and (3.11.11). The equations (3.11.6) – (3.11.10) when expressed in terms of $z_{(i)}$ and $e_{[i]}$ are given by

$$\frac{\partial \ln L}{\partial \delta} = \frac{n}{\eta} - \frac{1}{\eta}\sum_{i=1}^{n} \exp(-z_i) = 0,$$

$$\frac{\partial \ln L}{\partial \eta} = -\frac{n}{\eta} + \frac{1}{\eta}\sum_{i=1}^{n} z_i - \frac{1}{\eta}\sum_{i=1}^{n} z_{(i)} \exp(-z_{(i)}) = 0,$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2(1-\sigma^2)}\sum_{i=1}^{n} e_{[i]} = 0, \tag{3.11.13}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3(1-\rho^2)}\sum_{i=1}^{n} e_{[i]}^2 = 0$$

and
$$\frac{\partial \ln L}{\partial \rho} = \frac{n\rho}{(1-\rho^2)} - \frac{\rho}{\sigma^2(1-\rho^2)^2}\sum_{i=1}^{n} e_{[i]}^2 = 0.$$

The first two equations are intractable because of the terms $\exp(-z_{(i)})$, $1 \le i \le n$.

3.12 MML ESTIMATORS FOR THE BIVARIATE DISTRIBUTION

To work out the MML estimators, we replace $\exp(-z_{(i)})$ by the linear functional

$$e^{-z_{(i)}} \cong \alpha_i - \beta_i z_{(i)}, \quad 1 \le i \le n, \tag{3.12.1}$$

and obtain the modified likelihood equations $\partial \ln L^*/\partial\delta = 0, \partial \ln L^*/\partial\eta = 0$, etc. The coefficients α_i and β_i are given in (2.6.6). The solutions of the modified likelihood equations are the following MML estimators:

$$\hat{\delta} = K + D\hat{\eta}, \quad \hat{\eta} = \{B + \sqrt{(B^2 + 4nC)}\}/2n, \tag{3.12.2}$$

$$\hat{\mu} = \bar{y} - \hat{\rho}(\hat{\sigma}/\hat{\eta})(\bar{x} - \hat{\delta}) \tag{3.12.3}$$

and
$$\hat{\sigma} = \left[s_y^2 + \frac{s_{xy}^2}{s_x^2}\left(\frac{\hat{\eta}^2}{s_x^2} - 1\right)\right]^{1/2} \quad \text{and} \quad \hat{\rho} = \left(\frac{s_{xy}}{s_x^2}\right)\frac{\hat{\eta}}{\hat{\sigma}}; \tag{3.12.4}$$

$\bar{x}$, $\bar{y}$, s_x^2, s_y^2 and s_{xy} are the sample means, variances and the covariance as defined in (3.11.2). The expressions K, D, B and C are the same as in (2.6.10). A bias correction in $\hat{\delta}$ and $\hat{\eta}$ is instituted as in Chapter 2 (equation (2.6.10)). The MML estimator of θ_1 is obtained from (3.11.11),

$$\hat{\theta}_1 = \sum_{i=1}^{n} (x_i - \hat{\delta})(y_i - \hat{\mu}) \Big/ \sum_{i=1}^{n} (x_i - \hat{\delta})^2 \tag{3.12.5}$$

which simplifies to $\hat{\theta}_1 = s_{xy}/s_x^2$, since $\partial \ln L^*/\partial\theta_1 = 0$.

Similarly, $$\hat{\theta}_0 = \hat{\mu} - \hat{\theta}_1\hat{\delta} \quad \text{and} \quad \hat{\sigma}_e^2 = \sum_{i=1}^{n} \{y_i - \hat{\mu} - \hat{\theta}_1(x_i - \hat{\delta})\}^2/(n-2). \tag{3.12.6}$$

It is very important to realize that $\hat{\delta}$, $\hat{\eta}$, $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\rho}$ are entirely different than those obtained under the assumption of bivariate normality. Under bivariate normality, all the estimators are functions of the sample means, the sample variances s_x^2 and s_y^2, and the sample covariance s_{xy}. That ceases to be true if the marginal distribution is non-normal. The MMLE $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}_e^2$ are, however, exactly the same as under bivariate normality. This is due to the fact that the conditional distribution in (3.11.3) is normal.

Remark: The estimator $\hat{\sigma}$ is always positive. This follows from the fact that $s_{xy}^2 \le s_x^2 s_y^2$ so that $s_y^2 - (s_{xy}^2/s_x^2) \ge s_y^2 - s_y^2 = 0$. Since $s_{xy}^2\, \hat{\eta}^2 / s_x^2$ is always positive, the result follows.

Remark: The estimator $\hat{\rho}^2$ is bounded between 0 and 1. This follows from the fact that

$$\hat{\rho}^2 = 1 \Big/ \left\{1 + \frac{s_x^4 s_y^2}{s_{xy}^2\, \hat{\eta}^2}\left(1 - \frac{s_{xy}^2}{s_x^2 s_y^2}\right)\right\} \tag{3.12.7}$$

and $0 \le s_{xy}^2 \le s_x^2 s_y^2$.

Variance-Covariance Matrix: The asymptotic variance-covariance matrix of the MML estimators (3.12.2)–(3.12.4) is given by $I^{-1}(\delta, \eta, \mu, \sigma, \rho)$, I being the Fisher information matrix consisting of the elements $-E(\partial^2 \ln L/\partial\delta^2)$, $-E(\partial^2 \ln L/\partial\eta^2)$ etc. After some laborious algebra, we have the following expressions for the asymptotic variances (Vaughan and Tiku, 2000); $c = \int_0^\infty \ln u \exp(-u)\, du \cong 0.57722$ is the Euler constant and

$$\begin{aligned}
&\kappa = (\pi^2 + 6c^2 + 6 - 12c)/\pi^2 \cong 1.10866: \\
&V(\hat{\delta}) \cong \kappa\eta^2/n, \qquad V((\hat{\eta}) \cong 6\eta^2/(n\pi^2), \\
&V((\hat{\mu}) \cong \{\kappa + (1 - \rho^2)(12c - 6)/\pi^2\}\sigma^2/n, \\
&V(\hat{\sigma}) \cong \{(1 - \rho^2)^2\pi^2 + 12\rho^2\}\sigma^2/(2n\pi^2) \\
\text{and} \quad &V(\hat{\rho}) \cong \{(1 - \rho^2)^2 (12 + \pi^2\rho^2)\}/(2n\pi^2);
\end{aligned} \tag{3.12.8}$$

(3.12.8) are entirely different than the corresponding expressions under bivariate normality.

For the data of Example 3.2, we have the following estimates with their standard errors given in brackets:

$$\hat{\delta} = 4.065(\pm 0.332),\ \hat{\eta} = 1.260(\pm 0.246),\ \hat{\mu} = 81.429(\pm 14.370),$$
$$\hat{\sigma} = 53.485(\pm 8.763) \text{ and } \hat{\rho} = -0.7121(\pm 0.114).$$

The standard error of $\hat{\delta}$ is the smallest (8% of the estimate) and that of $\hat{\eta}$ is the largest (20% of the estimate). With only 16 observations, this is quite reasonable. The estimate $\hat{\rho} = -0.7121$ is large in magnitude as compared to zero and is a clear indication that a higher white blood count tends to shorten the survival time of a patient. A formal test of H_0: $\rho = 0$ is given in Section 3.13.

Tiku and Kambo (1992) consider the situation when the marginal distribution of X is long-tailed symmetric and the conditional distribution of Y given X = x is normal. They work out the MML estimators and their asymptotic variances and covariances. The situation when the conditional distribution of Y is also non-normal is difficult. No solutions are presently available. The topic is, however, under investigation at the present time (Sazak, 2003).

3.13 HYPOTHESIS TESTING

Testing the null hypothesis H_0: $\theta_1=0$ in the linear model (3.2.1) is of enormous practical interest. This can be done as follows.

Skew distributions: Suppose that the error distribution is Weibull (3.3.1). We assume that the shape parameter p is known. If not, we determine its value by using one of the techniques given in Chapters 9 and 11.

To test H_0: $\theta_1=0$, we define the statistic

$$T = \sqrt{\left\{\sum_{i=1}^{n} \delta_i \left(x_{[i]} - \bar{x}_{[.]}\right)^2\right\}}(\hat{\theta}_1 / \hat{\sigma}); \tag{3.13.1}$$

$\hat{\theta}_1$ and $\hat{\sigma}$ are the MML estimators (3.4.9)–(3.4.10). Large values of T lead to the rejection of H_0 in favour of H_1: $\theta_1>0$, and large values of $|T|$ lead to the rejection of H_0 in favour of H_1: $\theta_1 \neq 0$.

If the design is symmetric in which case $\sum_{i=1}^{n} x_i = 0$ or the design points are taken to be $x_i - \bar{x}_i$ with $\sum_{i=1}^{n} (x_i - \bar{x}_i) = 0$, then $\hat{\theta}_1$ and σ are uncorrelated at any rate for large n (matrix 3.4.12). In view of Theorem 3.1, therefore, the null distribution of T is referred to the Student t with $\nu = n - 2$ d.f. Islam et al. (2001, Table 4) show that the Student t distribution gives acccurate aproximations to the percentage points of T, for all $p \geq 1.4$.

The statistic based on the LS estimators (3.5.1)-(3.5.2) is

$$G = \sqrt{(ns_x^{\,2})}\{\tilde{\theta}_1 / (\Gamma(1+2/p) - \Gamma^2(1+1/p))\tilde{\sigma}^2\}^{1/2}, \tag{3.13.2}$$

$$ns_x^{\,2} = \sum_{i=1}^{n} (\bar{x}_i - \bar{x})^2.$$

Islam et al. (2001, Table 5) show that the null distribution of G is not well approximated by the Student t or the normal N(0, 1). However, the Student t (and the standard normal) provide more accurate approximations for the percentage points of $|T|$ and $|G|$ than for T and G. This is due to the fact that $\hat{\theta}_1$ and $\hat{\sigma}$, and $\tilde{\theta}_1$and $\tilde{\sigma}$ are not independent of one another if the underlying distribution is skew. Consequently, the null distributions of T and G are not symmetric unless n is large. See also Tiku (1971a) who gives Leguerre series expansions for the null and non-null distributions and shows that the distribution of G is very sensitive to non-normality, particularly to population skewness; see also Srivastava (1959). Islam et al. (2001, Table 5) show that the power of the G test is considerably lower than that of the T test. For example, Islam et al. give the following simulated values of the power of the T and G tests; σ taken to be equal to 1 without loss of generality, and presumed value of the type I error is 0.050:

Values of the power; p = 2.0, n = 20.

θ_1:	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4
T	0.033	0.12	0.32	0.57	0.79	0.92	0.97	1.00
G	0.008	0.03	0.10	0.24	0.42	0.65	0.81	0.92

The very low power of the G test is partly due to its type I error being much smaller than the presumed value and partly due to the inefficiency of the LS estimator $\tilde{\theta}_1$ (Islam et al., 2001, p. 1008).

The asymptotic power functions of the T and G tests are, respectively,

$$\text{Prob}(Z \ge z_\alpha - |\lambda_1|) \quad \text{and} \quad \text{Prob}\,(Z \ge z_\alpha - |\lambda_2|); \tag{3.13.3}$$

Z is a normal N(0, 1) variate and z_α is its $100(1-\alpha)$ percent point. From equation (3.5.8),

$$\lambda_1^2 = n(p-1)^2\,\Gamma(1-2/p)s_x^2(\theta_1/\sigma)^2 \tag{3.13.4}$$

and

$$\lambda_2^2 = \frac{ns_x^2}{\Gamma(1+2/p) - \Gamma^2(1+1/p)}(\theta_1/\sigma)^2$$

are the noncentrality parameters. Since $\lambda_1^2/\lambda_2^2 > 1$ (equation 3.5.4), the T test is asymptotically more powerful than the G test.

Symmetric distributions: Suppose that the random errors have a distribution in the family (3.9.1). In view of the results (i)-(iii) stated in Section 3.9, we define the statistic ($p \ge 2$)

$$T = \sqrt{\left\{\frac{np(p-1/2)}{(p+1)(p-3/2)}s_x^2\right\}}(\hat{\theta}_1/\hat{\sigma}); \tag{3.13.5}$$

$\hat{\theta}_1$ and $\hat{\sigma}$ are the MML estimators (3.9.5) – (3.9.6). The null distribution of T is referred to the Student t with $\nu = n-2$ d.f.

The statistic based on the LS estimators is

$$G = \sqrt{(ns_x^2)}(\tilde{\theta}_1/\tilde{\sigma}). \tag{3.13.6}$$

Like the statistic T, the null distribution of G is symmetric and is well approximated by the Student t with $\nu = n-2$ d.f. Since for symmetric distributions

$$\text{Prob}(|T| \ge t_{\alpha/2}(\nu)\,|\,H_0) = 2\,\text{Prob}(T \ge {}_{\alpha/2}(\nu)\,|\,H_0) \tag{3.13.7}$$

and similary for the statistic G, it suffices to test H_0 against the one-sided alternative H_1: $\theta_1 > 0$. Tiku et al. (2001) show that the T test is somewhat more powerful than the G test, as expected. For example, they give the following simulated values of the power; $\sigma = 1$ and presumed type I error is 0.050:

Values of the power; p = 3.0, n=20.

θ_1:	0.0	0.4	0.8	1.2	1.6	2.0
T	0.050	0.15	0.35	0.59	0.81	0.92
G	0.052	0.15	0.35	0.58	0.79	0.90

The asymptotic power functions of the T and G tests are the same as (3.13.3) with noncentrality parameters

$$\lambda_1^2 = \frac{np(p-1/2)}{(p+1)(p-3/2)} s_x^2 (\theta_1/\sigma)^2 \quad \text{and} \quad \lambda_2^2 = ns_x^2(\theta_1/\sigma)^2. \tag{3.13.8}$$

Since $\lambda_1^2/\lambda_2^2 = p(p-1/2)/(p+1)(p-3/2)$ is greater than 1 for all values of p, the T test is asymptotically more powerful than the G test. Moreover, the T test is remarkably robust but not the G test (Chapter 8); see also Islam et al. (2001, Table 7).

Comment: Statistics for testing H_0 can similarly be defined for other distributions. For the short-tailed symmetric family (3.6.2), the T statistic is (Tiku et al., 2001)

$$T = \sqrt{(nDs_x^2)}(\hat{\theta}_1/\hat{\sigma}); \tag{3.13.9}$$

$\hat{\theta}_1$ and $\hat{\sigma}$ are the MML estimators (3.7.1)–(3.7.2), and the constant D is the same as in (3.7.7). The Student t distribution with $\nu = n - 2$ degrees of freedom gives remarkably accurate approximations for the percentage points of the null distribution of T, and the T test is more powerful than the G test (Tiku et al., 2001, Table 4):

$$G = \sqrt{(ns_x^2)}(\tilde{\theta}_1/\tilde{\sigma}\sqrt{\mu_2}), \tag{3.13.10}$$

$\tilde{\theta}_1$ and $\tilde{\sigma}$ being the LS estimators (3.8.1). The T test is remarkably robust to short-tailed distributions and to inliers in a sample but not the G test (Tiku et al., 2001, Table 4). This is illustrated in Chapter 8.

Stochastic covariate: In situations where the design variable X is stochastic, and there are many such situations, testing H_0: $\rho = 0$ (ρ being the correlation coefficient between X and Y) is of primary interest. For the data in Example 3.2, for example, this can be done as follows (Vaughan and Tiku, 2000, p. 61).

Since the MML estimators are asymptotically equivalent to the ML estimators, the likelihood function (3.11.5) is maximized by the MML estimators (3.12.2) – (3.12.4). Asymptotically, therefore, the likelihood ratio statistic is

$$\hat{\Lambda} = \frac{\max(L|H_0)}{\max(L|H_1)} = \left(\frac{\hat{\sigma}^2}{s_y^2}\right)^{n/2} (1-\hat{\rho}^2)^{n/2} \times \exp\left[\frac{(n-1)s_y^2}{2(1-\hat{\rho}^2)\hat{\sigma}^2}(1-\hat{\rho}_0^2) - \frac{n-1}{2}\right], \tag{3.13.11}$$

where $\hat{\rho}_0 = s_{xy}/s_x s_y$ is the Pearson product moment correlation coefficient. Since $\hat{\rho}$ and $\hat{\rho}_0$ both converge to ρ and $\hat{\sigma}$ and s_y both converge to σ as n tends to infinity, the exponent is essentially zero for large n. Therefore, $\hat{\Lambda}$ is a monotonic function of $\hat{\rho}^2$ for large n. For testing H_0 against H_1: $\rho > 0$ (or $\rho < 0$), the test based on $\hat{\Lambda}$ is uniformly most powerful (asymptotically). To test H_0: $\rho = 0$, the proposed statistic is (Vaughan and Tiku, 2000, p.62)

$$W = \hat{\rho}\pi\sqrt{(n/6)}, \tag{3.13.12}$$

$6/n\pi^2$ being the asymptotic variance of $\hat{\rho}$ obtained from (3.12.8) by equating ρ to zero. Since $\hat{\rho}$ is asymptotically equivalent to the ML estimators, the asymptotic null distribution of W is normal N(0, 1).

For the data given in Example 3.2, $W = -3.654$. Since this value is less than -2.33, we reject H_0 in favour of H_1: $\rho < 0$ at 1% significance level. The conclusion is that a higher blood count tends to shorten the survival time of a patient.

3.14 NUMERICAL EXAMPLES

EXAMPLE 3.3: The following data is reproduced from Draper and Smith (1966, p.8). The response variable Y represents Pounds of Steam used monthly and the design variable X represents Average Atmospheric Temperature (degrees Fahrenheit) in a certain production process; n = 25:

y:	10.98	11.13	12.51	8.40	9.27	8.73	6.36	8.50	7.82	9.14
x:	35.3	29.7	30.80	58.8	61.4	71.3	74.4	76.7	70.7	57.5
y:	8.24	12.19	11.88	9.57	10.94	9.58	10.09	8.11	6.83	8.88
x:	46.4	28.9	28.1	39.1	46.8	48.5	59.3	70.0	70.0	74.5
y:	7.68	8.47	8.86	10.36	11.08					
x:	72.1	58.1	44.6	33.4	28.6					

To investigate the error distribution, we constructed Q-Q plots of $w_i = y_i - \tilde{\theta}_1 x_i$. The family (3.6.1) with r = 2 and d = 1.0 ($\lambda = 1$) provides the most plausible distribution; its standard deviation is $\sigma_1 = 1.679\sigma$. Here, we have the following estimates:

LS: $\tilde{\theta}_0 = 13.623,\ \tilde{\theta}_1 = -0.0798,\ s = 0.8901,\ \sum_{i=1}^{25} (x_i - \bar{x})^2 = 7154.420,$

$$SE(\tilde{\theta}_1) = \pm\sqrt{\frac{(0.8901)^2}{7154.420}} = \pm 0.0105 \qquad (3.14.1)$$

MML: $\hat{\theta}_0 = 13.665,\ \hat{\theta}_1 = -0.0805,\ \hat{\sigma} = 0.5430,\ \hat{\sigma}_1 = 0.912,$

$$SE(\hat{\theta}_1) \cong \pm\sqrt{\frac{(0.5430)^2}{0.6364(7154.420)}} = \pm 0.0080, \qquad (3.14.2)$$

since D = 0.6364 from (3.7.10). The MMLE $\hat{\theta}_1$ is clearly more precise than the LSE $\tilde{\theta}_1$.

The statistics $|T| = 10.08$ and $|G| = 7.60$; (3.14.3)

both are significantly different from zero at 1 percent significance level.

EXAMPLE 3.4: Consider the computer generated errors e_i $(1 \le i \le 30)$ from the Generalized Logistic GL(b, σ); b = 8. 0 and σ = 1. The model is

$$y_i = \theta_0 + \theta_1 x_i + e_i, \quad (1 \le i \le 30), \qquad (3.14.4)$$

where $\theta_1 = 1$ and x_i are uniform (0,1). The data is given below.

Y:	5.646	7.741	7.463	10.692	8.150	5.662	5.910	3.334	4.335	6.463
X:	0.863	0.334	0.311	0.747	0.466	0.796	0.144	0.448	0.359	0.820
Y:	5.335	6.515	5.978	3.568	4.892	8.535	7.035	6.004	4.087	4.926
X:	0.470	0.009	0.699	0.518	0.624	0.295	0.867	0.720	0.060	0.519
Y:	6.357	3.798	5.448	7.251	5.645	6.349	5.473	7.494	7.213	7.264
X:	0.766	0.669	0.134	0.762	0.823	0.412	0.168	0.722	0.659	0.553

Proceeding exactly along the same lines as in Sections 3.3-3.4, we obtain the MML estimators. They are exactly the same as (3.4.8)–(3.4.11) with δ_i replaced by β_i, $\Delta_i = \alpha_i - (b+1)^{-1}$, and B and C so obtained are both multiplied by (b + 1). The coefficients α_i and β_i are exactly the same as (2.5.5). The elements of the Fisher information matrix $I(\theta_0, \theta_1, \sigma)$ are

$$I_{11} = nb/(b+2)\sigma^2, \quad I_{12} = nb\bar{x}/(b+2)\sigma^2$$

$$I_{13} = nb\{\psi(b+1) - \psi(2)\}/(b+2)\sigma^2, \quad I_{22} = b\sum_{i=1}^{n} x_i^2/(b+2)\sigma^2, \tag{3.1.4.5}$$

$$I_{23} = nb\{\psi(b+1) - \psi(2)\}\bar{x}/(b+2)\sigma^2$$

and

$$I_{33} = nb[(b+2)/b + [\psi'(b+1) + \psi'(2)] + [\psi(b+1) - \psi(2)]^2]/(b+2)\sigma^2$$

It may be noted that

$$\frac{\partial \ln L^*}{\partial \theta_1} = \frac{(b+1)}{\sigma^2}\sum_{i=1}^{n}\beta_i\,(x_{[i]} - \bar{x}_{[.]})^2[K - D\sigma - \theta_1] \tag{3.14.6}$$

Thus, we obtain the following estimates and their standard errors.

LS: $\tilde{\theta}_1 = 0.983$ with standard error $\pm\sqrt{\dfrac{s_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$

$$= \pm\sqrt{\frac{(1.561)^2}{1.896}} = \pm 1.285$$

MML: $\hat{\theta}_1 = 1.087$ with standard error $\pm\sqrt{\dfrac{\hat{\sigma}^2}{(b+1)\sum_{i=1}^{n}\beta_i(x_{[i]} - \bar{x}_{[.]})^2}}$

$$= \pm\sqrt{\frac{(1.252)^2}{9(0.153)}} = \pm 1.066. \tag{3.14.7}$$

The estimated standard error of $\hat{\theta}_1$ calculated from the asymptotic formula

$$V(\hat{\theta}_1) \cong (b+2)\sigma^2 \Big/ b\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3.14.8}$$

is

$$\pm\sqrt{\frac{10}{8(1.896)}} = \pm 1.017$$

which is very close to the value given in (3.14.7).

Again, the MML estimate is numerically close to the LS estimate and is clearly more precise.

SUMMARY

In this Chapter, we consider linear regression models. Traditionally, the errors in the models are assumed to be normally distributed. The least square estimators (LSE) of the parameters are identical to the MLE under normality and are fully efficient. In practice, however, non-normal distributions are more prevalent. The MLE are intractable under non-normality and the LSE are inefficient. We derive MMLE for the three families of distributions considered in the previous chapter. The estimators are obtained by ordering the errors e_i $(1 \le i \le n)$. We show that the MMLE are asymptotically fully efficient. For small n, their variances are only marginally bigger than the minimum variance bounds. We investigate the relative efficiencies of the LSE: they are much less efficient than the MMLE. A disconcerting feature of the LSE is that their relative efficiencies descrease as the sample size n increases and stabilize at values considerably less than 100 percent. We develop hypothesis testing procedures. We give numerical examples to illustrate the superiority of the MMLE over the LSE. We extend the methodology of modified likelihood to the important situation when the design variable X is also stochastic.

APPENDIX 3A

INFORMATION MATRIX

To derive the elements of the Fisher information matrix (3.4.12), consider the equation (3.3.5), i.e.,

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{-(p-1)}{\sigma}\sum_{i=1}^{n} x_i z_i^{-1} + \frac{p}{\sigma}\sum_{i=1}^{n} x_i z_i^{p-1}, \quad z_i = (y_i - \theta_0 - \theta_1 x_i)/\sigma. \tag{3A.1}$$

First, realize that $E(\partial \ln L/\partial\theta_1) = 0$. This follows from the fact that

$$E\left(\frac{1}{z}\right) = p\int_0^\infty \exp(-z^p)z^{p-2}\,dz = \Gamma\left(1-\frac{1}{p}\right)$$

and

$$E(z^{p-1}) = p\int_0^\infty \exp(-z^p)z^{2(p-1)}\,dz = \Gamma\left(2-\frac{1}{p}\right) = \frac{p-1}{p}\Gamma\left(1-\frac{1}{p}\right).$$

Now,

$$-E\left(\frac{\partial^2 \ln L}{\partial\theta_1^2}\right) = \frac{p-1}{\sigma^2}\sum_{i=1}^{n} x_i^2 E(z_i^{-2}) + \frac{p(p-1)}{\sigma^2}\sum_{i=1}^{n} x_i^2 E(z_i^{p-2})$$

$$= \frac{p-1}{\sigma^2}\Gamma\left(1-\frac{2}{p}\right)\sum_{i=1}^{n} x_i^2 \frac{p(p-1)}{\sigma^2}\Gamma\left(2-\frac{2}{p}\right)\sum_{i=1}^{n} x_i^2 \tag{3A.2}$$

$$= \frac{(p-1)^2}{\sigma^2}\Gamma\left(1-\frac{2}{p}\right)\sum_{i=1}^{n} x_i^2.$$

All other elements in (3.4.12) can, similarly, be obtained. Alternatively, one can work with $\partial \ln L^*/\partial\theta_1$ and $\partial^2 \ln L^*/\partial\theta_1^2$ and use the asymptotic arguments of Appendix 2A to arrive at exactly the same results as (3A.1)–(3A.2). The information matrices (3.7.7) and (3.9.8), and others given in this book, are obtained exactly along the same lines.

APPENDIX 3B

VALUES OF $E\{Z_{(i)}\}$

The approximate values of $t_{(i)}$ for r = 4, a = 0.0

n	i	$t_{(I)}$	n	i	$t_{(i)}$	n	i	$t_{(i)}$	n	i	$t_{(i)}$
10	1	– 2.1868	30	6	– 1.5209	50	6	– 1.9866	100	16	– 1.7236
10	2	– 1.5868	30	7	– 1.3466	50	7	– 1.8553	100	17	– 1.6648
10	3	– 1.1053	30	8	– 1.1795	50	8	– 1.7330	100	18	– 1.6074
10	4	– 0.6580	30	9	– 1.0173	50	9	– 1.6174	100	19	– 1.5512
10	5	– 0.2187	30	10	– 0.8581	50	10	– 1.5069	100	20	– 1.4962
12	1	– 2.3057	30	11	– 0.7007	50	11	– 1.4003	100	21	– 1.4421
12	2	– 1.7513	30	12	– 0.5444	50	12	– 1.2968	100	22	– 1.3889
12	3	– 1.3205	30	13	– 0.3885	50	13	– 1.1957	100	23	– 1.3365
12	4	– 0.9313	30	14	– 0.2329	50	14	– 1.0964	100	24	– 1.2847
12	5	– 0.5564	30	15	– 0.0773	50	15	– 0.9984	100	25	– 1.2335
12	6	– 0.1850									
			40	1	– 2.9727						
15	1	– 2.4438	40	2	– 2.5968	50	16	– 0.9016	100	26	– 1.1828
15	2	– 1.9361	40	3	– 2.3399	50	17	– 0.8055	100	27	– 1.1326
15	3	– 1.5547	40	4	– 2.1341	50	18	– 0.7100	100	28	– 1.0827
15	4	– 1.2208	40	5	– 1.9569	50	19	– 0.6148	100	29	– 1.0332
15	5	– 0.9076				50	20	– 0.5199	100	30	– 0.9840
15	6	– 0.6029	40	6	– 1.79771						
15	7	– 0.3010	40	7	– 1.65068	50	21	– 0.4252	100	31	– 0.9350
15	8	0.0004	40	8	– 1.51222	50	22	– 0.3305	100	32	– 0.8863
			40	9	– 1.37996	50	23	– 0.2359	100	33	– 0.8378
20	1	– 2.6111	40	10	– 1.25220	50	24	– 0.1413	100	34	– 0.7894
20	2	– 2.1523				50	25	– 0.0468			
20	3	– 1.8196	40	11	– 1.12772				100	35	– 0.7411
20	4	– 1.5380	40	12	– 1.00561	100	1	– 3.3826	100	36	– 0.6930
20	5	– 1.2822	40	13	– 0.88516	100	2	– 3.0740	100	37	– 0.6449
20	6	– 1.0403	40	14	– 0.76587	100	3	– 2.8724	100	38	– 0.5970
20	7	– 0.8055	40	15	– 0.64732	100	4	– 2.7170	100	39	– 0.5490
20	8	– 0.5741	40	16	– 0.52922	100	5	– 2.5880	100	40	– 0.5012
20	9	– 0.3441	40	17	– 0.41139	100	6	– 2.4761	100	41	– 0.4533
20	10	– 0.1143	40	18	– 0.29366	100	7	– 2.3763	100	42	– 0.4055
			40	19	– 0.17600	100	8	– 2.2854	100	43	– 0.3577
30	1	– 2.8290	40	20	– 0.05835	100	9	– 2.2014	100	44	– 0.3099
30	2	– 2.4233				100	10	– 2.1229	100	45	– 0.2622
30	3	– 2.1402	50	1	– 3.07877						
30	4	– 1.9091	50	2	– 2.72259	100	11	– 2.0489	100	46	– 0.2144
30	5	– 1.7064	50	3	– 2.48240	100	12	– 1.9785	100	47	– 0.1666
			50	4	– 2.29231	100	13	– 1.9112	100	48	– 0.1189
			50	5	– 2.13046	100	14	–1.8466	100	49	– 0.0711
						100	15	– 1.7841	100	50	– 0.0233

The approximate values of $t_{(i)}$ for r = 4, a = 1.0

n	i	$t_{(I)}$	n	i	$t_{(i)}$	n	i	$t_{(i)}$	n	i	$t_{(i)}$
10	1	– 2.4376	30	6	– 1.7772	50	6	– 2.2420	100	16	– 1.9816
10	2	– 1.8441	30	7	– 1.5976	50	7	– 2.1125	100	17	– 1.9226
10	3	– 1.3415	30	8	– 1.4214	50	8	– 1.9910	100	18	– 1.8648
10	4	– 0.8339	30	9	– 1.2453	50	9	– 1.8750	100	18	– 1.8080
10	5	– 0.2858	30	10	– 1.0670	50	10	– 1.7629	100	20	– 1.7519
12	1	– 2.5532	30	11	– 0.8846	50	11	– 1.6534	100	21	– 1.6965
12	2	– 2.0093	30	12	– 0.6968	50	12	– 1.5456	100	22	– 1.6416
12	3	– 1.5704	30	13	– 0.5031	50	13	– 1.4386	100	23	– 1.5871
12	4	– 1.1497	30	14	– 0.3041	50	14	– 1.3318	100	24	– 1.5329
12	5	– 0.7114	30	15	– 0.1014	50	15	– 1.2245	100	25	– 1.4788
12	6	– 0.2421									
			40	1	– 3.1967	50	16	– 1.1163	100	26	– 1.4249
15	1	– 2.6869	40	2	– 2.8346	50	17	– 1.0067	100	27	– 1.3709
15	2	– 2.1923	40	3	– 2.5863	50	18	– 0.8955	100	28	– 1.3170
15	3	– 1.8115	40	4	– 2.3863	50	19	– 0.7822	100	29	– 1.2628
15	4	– 1.4653	40	5	– 2.2127	50	20	– 0.6668	100	30	– 1.2085
15	5	– 1.1231									
15	6	– 0.7679	40	6	– 2.0554	50	21	– 0.5492	100	31	– 1.1539
15	7	– 0.3918	40	7	– 1.9084	50	22	– 0.4296	100	32	– 1.0991
15	8	0.0001	40	8	– 1.7683	50	23	– 0.3081	100	33	– 1.0438
			40	9	– 1.6323	50	24	– 0.1852	100	34	– 0.9881
20	1	– 2.8484	40	10	– 1.4986	50	25	– 0.0614	100	35	– 0.9320
20	2	– 2.4040									
20	3	– 2.0771	40	11	– 1.3657	100	1	– 3.5914	100	36	– 0.8754
20	4	– 1.7946	40	12	– 1.2324	100	2	– 3.2943	100	37	– 0.8183
20	5	– 1.5303	40	13	– 1.0977	100	3	– 3.1002	100	38	– 0.7606
			40	14	– 0.9608	100	4	– 2.9505	100	39	– 0.7024
20	6	– 1.2706	40	15	– 0.8211	100	5	– 2.8261	100	40	– 0.6437
20	7	– 1.0067									
20	8	– 0.7330	40	16	– 0.6781	100	6	– 2.7181	100	41	–0.5844
20	9	– 0.4468	40	17	– 0.5318	100	7	– 2.6216	100	42	– 0.5245
20	10	– 0.1499	40	18	– 0.3824	100	8	– 2.5334	100	43	– 0.4641
			40	19	– 0.2303	100	9	– 2.4519	100	44	– 0.4032
30	1	– 3.0584	40	20	– 0.0765	100	10	– 2.3754	100	45	–0.3419
30	2	– 2.6671									
30	3	– 2.3922	50	1	– 3.2988	100	11	– 2.3030	100	46	– 2.2802
30	4	– 2.1657	50	2	– 2.9558	100	12	– 2.2340	100	47	– 0.2182
30	5	– 1.9644	50	3	– 2.7241	100	13	– 2.1678	100	48	– 0.1558
			50	4	– 2.5401	100	14	– 2.1039	100	49	– 0.0933
			50	5	– 2.3827	100	15	– 2.0419	100	50	– 0.0307

CHAPTER 4

Binary Regression with Logistic and Nonlogistic Density Functions

4.1 INTRODUCTION

In the previous chapter, we assumed a linear relationship between a response variable Y and a design variable X. For the random error component in the model, we considered three families of distributions: (i) skew, (ii) short-tailed symmetric, and (iii) long-tailed symmetric. By using the method of modified likelihood, we estimated the unknown parameters. We extended the methodology to k (≥ 1) design variables. We showed that for all these three families of distributions, the likelihood equations are intractable. As a consequence, the ML estimators are elusive. We formulated modified likelihood equations which have explicit solutions, the MML estimators. We showed that the MML estimators are asymptotically fully efficient, and are highly efficient for small sample sizes. We also showed that the MML estimators are considerably more efficient than the LS (least squares) estimators. In fact, the LS estimators have a disconcerting feature, namely, their relative efficiencies decrease with increasing sample size n and stabilize at values considerably less than 100 percent. We also developed hypothesis testing procedures. In numerous biomedical and industrial applications, however, the response variable Y can assume only two values 0 and 1 and depends on one or more design variables called risk factors or covariates. Since Y is a binary random variable, its expected value is between 0 and 1. A model which quantifies a relationship between Y and a risk factor(s) X has to be structured in such a way that E(Y) assumes a value between 0 and 1. This is accomplished as follows.

4.2 LINK FUNCTIONS

In the first place, suppose that Y depends on a single covariate X through the relationship

$$\pi(x) = E(Y \mid X = x) = \int_{-\infty}^{z} f(u)\,du = F(z) \tag{4.2.1}$$

where $z = \eta(x)$ (a function of x) and f(u) is a completely specified density function, and Y is presumed to increase with X. In the first place, we assume $\eta(x)$ to be linear so that $z = \gamma_0 + \gamma_1 x$, $\gamma_1 > 0$ a priori. The function F(u) is the cumulative density function of u and, in the context of binary regression, F^{-1} is called the link function. The problem is to estimate γ_0 and γ_1 and to verify whether there is statistical evidence to support the presumption that γ_1 is positive.

The density function that has been used very extensively in binary regression is the logistic:

$$f(u) = e^{-u}/(1 + e^{-u})^2, \quad -\infty < u < \infty. \tag{4.2.2}$$

The logistic density has a beautiful property, namely,

$$\pi(x) = F(z) = e^z/(1 + e^z) \tag{4.2.3}$$

which gives the very interesting result that

$$\text{In}[\pi(x)/(1 - \pi(x))] = \gamma_0 + \gamma_1 x. \tag{4.2.4}$$

The value $\pi(x)/(1 - \pi(x))$ = OD is called odds. Thus, the logistic density (4.2.2) has the property that ln(OD) is a linear function of the covariate X (Hosmer and Lemeshow, 1989; Collett, 1991; Kleinbaum, 1994; Agresti, 1996). Realize that for the logistic density, the link function (also called logit) is

$$F^{-1} = \text{In}[\pi(x)/(1 - \pi(x))]. \tag{4.2.5}$$

In binary regression, the assumption of the logistic density is very predominant. There is, however, a need to open up this area to nonlogistic density functions (Tiku and Vaughan, 1997). Robinson et al. (1998) give link functions for several nonlogistic density functions.

Given a random sample (y_i, x_i), $1 \le i \le n$, we want to estimate γ_0 and γ_1. Assuming that X is nonstochastic and is measurable without error, the likelihood function is

$$L = \prod_{i=1}^{n} \{F(z_i)\}^{y_i} \{1 - F(z_i)\}^{1-y_i} \tag{4.2.6}$$

where $F(z) = \int_{-\infty}^{z} f(u)\,du$ and f(u) is any density function not necessarily logistic. An alternative expression of L in terms of the ordered variates $z_{(i)}$ is obtained as follows.

Since γ_1 is a priori positive, $z_{(i)}$ are determined by the ordered values (Tiku and Vaughan, 1997)

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}. \tag{4.2.7}$$

Clearly, the ordered variates $z_{(i)}$ are given by

$$z_{(i)} = \gamma_0 + \gamma_1 x_{(i)}, \quad 1 \le i \le n. \tag{4.2.8}$$

Let $y_{[i]}$ be the y-observation (concomitant) which corresponds to $x_{(i)}$. Then

$$L = \prod_{i=1}^{n} \{F(z_{(i)})\}^{w_i} \{1 - F(z_{(i)})\}^{1-w_i} \quad (w = y_{[i]}). \tag{4.2.9}$$

The likelihood equations for estimating γ_0 and γ_1 are

$$\frac{\partial \ln L}{\partial \gamma_0} = \sum_{i=1}^{n} \{w_i g_1(z_{(i)}) - (1 - w_i) g_2(z_{(i)})\} = 0 \quad \text{and} \tag{4.2.10}$$

$$\frac{\partial \ln L}{\partial \gamma_1} = \sum_{i=1}^{n} x_{(i)} \{w_i g_1(z_{(i)}) - (1 - w_i) g_2(z_{(i)})\} = 0; \tag{4.2.11}$$

$$g_1(z) = f(z)/F(z) \quad \text{and} \quad g_2(z) = f(z)/(1 - F(z)). \tag{4.2.12}$$

Due to the intractrable nature of $g_1(z)$ and $g_2(z)$, the equations (4.2.10) – (4.2.11) have no explicit solutions. However, when the density function is logistic, computer software is available to solve these equations by iteration. The software also gives the approximate values of the standard errors which it computes from the Fisher information matrix. The elements of this matrix will be given later in Section 4.4.

4.3 MODIFIED LIKELIHOOD ESTIMATORS

We now derive the MML estimators of γ_0 and γ_1 and calculate their approximate standard errors as follows.

We linearize the g_1 and g_2 functions and write

$$g_1(z_{(i)}) \cong \alpha_{1i} - \beta_{1i} z_{(i)} \quad \text{and} \quad g_2(z_{(i)}) \cong \alpha_{2i} + \beta_{2i} z_{(i)}. \tag{4.3.1}$$

As in earlier chapters, the coefficients $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ are obtained from the first two terms of a Taylor series expansion around $t_{(i)}$ $(1 \le i \le n)$:

$$\beta_{1i} = \{f^2(t_{(i)}) - F(t_{(i)}) f'(t_{(i)})\}/F^2(t_{(i)}) \quad \text{and} \quad \alpha_{1i} = g_1(t_{(i)}) + \beta_{1i} t_{(i)}, \tag{4.3.2}$$

and

$$\beta_{2i} = \{f^2(t_{(i)}) + (1 - F(t_{(i)}))f'(t_{(i)})\}/(1 - F(t_{(i)}))^2 \quad \text{and} \tag{4.3.3}$$

$$\alpha_{2i} = g_2(t_{(i)}) - \beta_{2i} t_{(i)};$$

$f'(z)$ is the differential coefficient of $f(z)$ with respect to z which exists for all $z = t_{(i)}$ $(1 \le i \le n)$ for the wide variety of densities listed in Appendix 4A. The values of $t_{(i)}$ are obtained from the equations

$$\int_{-\infty}^{t_{(i)}} f(u)\,du = \frac{i}{n+1}, \quad 1 \le i \le n; \tag{4.3.4}$$

$t_{(i)}$ could be called "population" quantiles. The random variates $z_{(i)}$ $(1 \le i \le n)$ are regarded as the order statistics of a random sample of size n from a population with density function $f(u)$. This would imply that $z_{(i)} - t_{(i)}$ $(1 \le i \le n)$ tend to zero as n tends to infinity.

Incorporating the linear functions (4.3.1) in (4.2.10) – (4.2.11), we obtain the modified likelihood equations, namely,

$$\frac{\partial \ln L}{\partial \gamma_0} \cong \frac{\partial \ln L^*}{\partial \gamma_0} = \sum_{i=1}^{n} m_i\{(\delta_i / m_i) - \gamma_0 - \gamma x_{(i)}\} = 0 \quad \text{and} \tag{4.3.5}$$

$$\frac{\partial \ln L}{\partial \gamma_1} \cong \frac{\partial \ln L^*}{\partial \gamma_1} = \sum_{i=1}^{n} m_i x_{(i)}\{(\delta_i / m_i) - \gamma_0 - \gamma x_{(i)}\} = 0; \tag{4.3.6}$$

$$\delta_i = \alpha_{1i} w_i - \alpha_{2i}(1 - w_i) \quad \text{and} \quad m_i = \beta_{1i} w_i + \beta_{2i}(1 - w_i), \tag{4.3.7}$$

$$w_i = y_{[i]} \quad (1 \le i \le n).$$

The solutions of (4.3.5) – (4.3.6) are the following MML estimators:

$$\hat\gamma_1 = \sum_{i=1}^{n} \delta_i (x_{(i)} - \bar x_a) \Big/ \sum_{i=1}^{n} m_i (x_{(i)} - \bar x_a)^2 \quad \text{and} \quad \hat\gamma_0 = (\delta/m) - \hat\gamma_1 \bar x_a; \tag{4.3.8}$$

$$\delta = \sum_{i=1}^{n} \delta_i, \; m = \sum_{i=1}^{n} m_i \quad \text{and} \quad \bar x_a = (1/m) \sum_{i=1}^{n} m_i x_{(i)}.$$

If $\hat\gamma_1$ turns out to be negative, equate it to 0.000001, since γ_1 is positive.

It may be noted that

$$\sum_{i=1}^{n} m_i (x_{(i)} - \bar x_a)^2 = \sum_{i=1}^{n} m_i x_{(i)}^2 - m \bar x_a^2.$$

Revised estimates: Following a suggestion by Lee et al. (1980), we now calculate $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ from (4.3.2) – (4.3.3) by replacing $t_{(i)}$ by

$$t_{(i)} = \hat\gamma_0 + \hat\gamma_1 x_{(i)}, \quad 1 \le i \le n, \tag{4.3.9}$$

and calculate the revised estimates $\hat{\gamma}_0$ and $\hat{\gamma}_1$ from equations (4.3.8). If need be, this process is repeated a few times until the coefficients (4.3.2) – (4.3.3) stabilize sufficiently enough. This process needs to be repeated only a few times as calculations reveal that the coefficients α_i and β_i in (4.3.2) – (4.3.3) stabilize very quickly; see also Lee et al. (1980).

Comment: The procedure above for calculating the MML estimates is straightforward. The procedure for calculating the ML estimates is to first find the link F^{-1} = inverse of the cumulative distribution function $\pi(x) = F(z)$, and use the initial aproximations $\pi_{i0}=(y_i+0.5)/(n_i+1)$ (often $n_i=1$). The π_{i0} values are then regressed on the covariate(s) to get initial approximations to the γ-coefficients (using a weighted least square approach, e.g. equations (3.2.9) – (3.2.10)) and then re-evaluating the π estimates and repeating (Collett 1991, Appendix B). The weights used are obtained from either Hessian or information matrix, and their inverses. For the logistic density, the estimates of the variances of the ML estimators of γ_0 and γ_1 are the same regardless of the use of the Hessian versus the information matrix. This is, however, not true if the density function f(u) is nonlogistic. This creates difficulties in computing the ML estimates from nonlogistic density functions. The MML estimates have no such computational difficulty. Besides, the MML estimates are the same (almost) as the ML estimates (Examples 4.1 – 4.5); see also Appendix 2B (Chapter 2).

4.4 VARIANCES AND COVARIANCES

Since the MML estimators are asymptotically equivalent to the ML estimators, their asymptotic variances and the covariance are given by $I^{-1}(\gamma_0, \gamma_1)$ where I is the Fisher information matrix. Alternatively, we proceed as follows to obtain the asymptotic variances and the covariance.

The equations (4.3.5) – (4.3.7) can be written as

$$\frac{\partial \ln L^*}{\partial \gamma_0} = \sum_{i=1}^{n} m_i\{(\delta_i/m_i) - \gamma_0 - \gamma_1 x_i\} = 0 \quad \text{and} \tag{4.4.1}$$

$$\frac{\partial \ln L^*}{\partial \gamma_1} = \sum_{i=1}^{n} m_i x_i\{(\delta_i/m_i) - \gamma_0 - \gamma_1 x_i\} = 0; \tag{4.4.2}$$

$$\delta_i = \alpha_{1i} y_i - \alpha_{2i}(1 - y_i) \quad \text{and} \quad m_i = \beta_{1i} y_i + \beta_{2i}(1 - y_i), \tag{4.4.3}$$

and $t_{(i)} = \gamma_0 + \gamma_1 x_{(i)}$ is replaced by $t_i = \gamma_0 + \gamma_1 x_i$ in (4.3.2) – (4.3.3). The initial values are taken as

$$t_i = \tilde{\gamma}_0 + \tilde{\gamma} x_i,$$

where $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ are the least squares estimators, namely,

$$\tilde{\gamma}_1 = \sum_{i=1}^{n} (x_i - \bar{x}) y_i \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{and} \quad \tilde{\gamma}_0 = \bar{y} - \tilde{\gamma}_1 \bar{x}. \tag{4.4.4}$$

The solutions of (4.4.1) – (4.4.3) are exactly the same as those of (4.3.5) – (4.3.7), and are obtained by iteration. For long-tailed density functions, however, (4.4.1) – (4.4.3) need a considerably larger number of iterations to stabilize than do the equations (4.3.5) – (4.3.7). The latter benefit from the advantages of ordering which yields the MML estimates quickly in a very few iterations. This is indeed a basic characteristic of the modified likelihood estimation based on order statistics as shown in Chapter 3; see also Lee et al. (1980).

Noting that $t_i = \hat{\gamma}_0 + \hat{\gamma}_i x_i$ and $z_i = \gamma_0 + \gamma_1 x_i$ and since $\hat{\gamma}_0$ and $\hat{\gamma}_1$ converge to γ_0 and γ_1, respectively, $z_i - t_i$ $(1 \le i \le n)$ converge to zero as n tends to infinity. Consequently, α_i's and β_i's are treated as constant coefficients for large n, in which case (since Y_i is binary with expected value $F(z_i)$)

$$E(\delta_i) \cong Q_i z_i \quad \text{and} \quad E(m_i) \cong Q_i; \quad Q_i = f^2(z_i)/\{F(z_i)[1 - F(z_i)]\}. \tag{4.4.5}$$

The MML estimators $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are asymptotically unbiased. This follows from the fact that $E(\hat{\theta}_1 / \hat{\theta}_2) \cong E(\hat{\theta}_1) / E(\hat{\theta}_2)$ for large n (Stuart and Ord, 1987, p. 325) and δ/n, m/n and $(1/n)\sum_{i=1}^{n} m_i x_i$ converge to their expected values as n tends to infinity. Moreover, the asymptotic variance-covariance matrix of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ is given by (since asymptotically $E(\partial^2 \ln L/\partial\theta^2) \cong E(\partial^2 \ln L^*/\partial\theta^2)$)

$$V = \begin{bmatrix} -E\left(\dfrac{\partial^2 \ln L}{\partial\gamma_0{}^2}\right) & -E\left(\dfrac{\partial^2 \ln L}{\partial\gamma_0\partial\gamma_1}\right) \\ -E\left(\dfrac{\partial^2 \ln L}{\partial\gamma_0\partial\gamma_1}\right) & -E\left(\dfrac{\partial^2 \ln L}{\partial\gamma_1{}^2}\right) \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma Q_i & \Sigma Q_i x_i \\ \Sigma Q_i x_i & \Sigma Q_i x_i{}^2 \end{bmatrix}^{-1}. \tag{4.4.6}$$

Realize that y_i is a binary random variable and its variance is $F(z_i)[1 - F(z_i)]$. An estimate of V is obtained by replacing Q_i by

$$\hat{Q}_i = f^2(\hat{z}_i) / [F(\hat{z}_i)(1 - F(\hat{z}_i))] \tag{4.4.7}$$

where $\hat{z}_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_i \quad (1 \le i \le n).$

4.5 HYPOTHESIS TESTING

Testing the null hypothesis $H_0 : \gamma_1 = 0$ is of great practical interest. Rejection of H_0 clearly implies that the risk factor (covariate) X has an effect on the response Y. Under H_0, the estimated probability of observing n_1 number of 1's and n_0 number of 0's is

$$\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}, \; n_1 + n_0 = n. \tag{4.5.1}$$

Assuming that H_0 is not true, the estimated probability is

$$\prod_{i=1}^{n} \{\hat{\pi}_i{}^{y_i}(1 - \hat{\pi}_i)^{1-y_i}\}, \; \hat{\pi}_i = F(\hat{z}_i). \tag{4.5.2}$$

The likelihood ratio statistic is, therefore,

$$G = -2\ln\left[\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0} \Big/ \prod_{i=1}^{n} \hat{\pi}_i{}^{y_i} (1 - \hat{\pi}_i)^{1-y_i}\right]. \tag{4.5.3}$$

Large values of G lead to the rejection of H_0 in favour of $H_1: \gamma_1 > 0$. For large n, the null distribution of G is chi-square with 1 degree of freedom. Alternatively, one calculates the Wald statistic

$$W = \hat{\gamma}_1/\sqrt{(SE)^2} \tag{4.5.4}$$

and rejects H_0 if W is large, $(SE)^2$ being the estimated variance of $\hat{\gamma}_1$. For large n, the null distribution of W is referred to N(0, 1).

4.6 LOGISTIC DENSITY

The modified likelihood estimation can be carried out for any density function without much difficulty. For the logistic density, however, it simplifies, i.e., the coefficients (4.3.2) – (4.3.3) are given by

$$\beta_{1i} = \beta_{2i} = \beta_i,\ \alpha_{1i} = \alpha_i \quad \text{and} \quad \alpha_{2i} = 1 - \alpha_i \quad (1 \le i \le n) \tag{4.6.1}$$

where
$$\beta_i = e^{t_{(i)}}/\{1 + e^{t_{(i)}}\}^2 \quad \text{and} \quad \alpha_i = \{1 + e^{t_{(i)}}\}^{-1} + \beta_i t_{(i)}, \tag{4.6.2}$$

and with $w_i = y_{[i]}$,

$$m_i = \beta_i \quad \text{and} \quad \delta_i = \alpha_i w_i - (1 - \alpha_i)(1 - w_i). \tag{4.6.3}$$

And, $t_{(i)}$ is calculated from the equation

$$t_{(i)} = -\ln(q_i^{-1} - 1),\ q_i = i/(n + 1), \quad 1 \le i \le n, \tag{4.6.4}$$

which is the solution of (4.3.4) if f(u) is the logistic density.

Remark: There is enormous difficulty in calculating the ML estimates for nonlogistic density functions. Most of the work reported is, therefore, based on the logistic density. Since the MML estimators are computationally straight forward, they provide an opportunity to open up the area to nonlogistic density functions, and to stochastic covariates (Section 4.11).

4.7 COMPARISON OF THE ML AND MML ESTIMATES

We now consider real-life data sets where the ML estimates and their standard errors are available based, of course, on the logistic density function (4.2.2). We report the ML and the MML estimates and their standard errors. The two are seen to be the same (almost).

Example 4.1: Consider the widely reported CHD data given on page 3 of Hosmer and Lemeshow (1989). We reproduce this data in Appendix 4B for ready reference. The data represents the values of Y (coronary heart disease status) and the corresponding values of the covariate X (age) of 100 subjects; y = 0 and 1 represent the absence and presence of the disease, respectively. The problem is to determine the relationship between Y and X.

The ML and the MML estimates and their standard errors are given in Table 4.1. Only two iterations were needed for the MML estimates to stabilize. Also given are the values of the Wald statistic W and the likelihood ratio statistic G . The values in brackets are the results of the first calculation, i.e., the MML estimates (and their standard errors) based on $t_{(i)}$, obtained from (4.3.4).

Table 4.1: Estimates and standard errors for the CHD data.

	Coefficient	Estimate	Standard error	W	G
ML	γ_0	– 5.310	1.134		
	γ_1	0.111	0.024	4.61	29.31
MML	γ_0	– 5.309(– 4.754)	1.134(1.073)		
	γ_1	0.111 (0.098)	0.024 (0.023)	4.61(4.26)	29.31(28.93)

The ML estimates and their standard errors quoted above are from Hosmer and Lemeshow (1989, p.11). It can be seen that the method of modified likelihood gives essentially the same results as the Fisher maximum likelihood. Both methods overwhelmingly reject H_0 : $\gamma_1 = 0$. The conclusion is that age has a definite effect on the CHD. It may be noted that the computations with the MML estimators are straightforward and, unlike the ML estimators, no software package is needed to compute them.

Example 4.2: The data set used in previous example was large with sample size n=100. Consider now the following data (y_i, x_i) with n=10 observations taken from Hosmer and Lemeshow (1989, p. 132):

y_i = 0, 1, 0, 0, 0, 0, 0, 1, 0, 1

x_i = 0.225, 0.487, – 1.080, – 0.870, – 0.580, – 0.640, 1.614, 0.352, – 1.025, 0.929

In fact, y_i are randomly generated from a uniform(0, 1) distribution and x_i are generated from a standard normal. Here, we have the following values, based on the logistic density. The ML estimates and their standard errors are from Hosmer and Lemeshow (1989).

Table 4.2: The ML and MML estimates, n=10.

	Coefficient	Estimate	Standard error
ML	γ_0	– 1.0	0.83
	γ_1	1.4	1.00
MML	γ_0	– 1.00	0.829
	γ_1	1.38	1.010

Even for this small data set, the modified likelihood methodology gives essentially the same results as the maximum likelihood. This is typical of modified likelihood estimation as will be illustrated from time to time.

Example 4.3: Another data set of n = 27 observations is given in Agresti (1996, p.88). For this data, X quantifies the proliferative activity of cells after a patient receives an injection of tritiated thymidine. The response variable Y is 1 if the patient achieves remission and Y=0 otherwise. Here, we have the results given in Table 4.3. The ML estimates and their standard errors are from Agresti (1996, p. 87). Only two iterations were needed for the MML estimates to stabilize sufficiently enough.

Table 4.3: The ML and MML estimates for Agresti's data.

	Coefficient	Estimate	Standard error	G
ML	γ_0	– 3.777	*	
	γ_1	0.145	0.059	*
MML	γ_0	– 3.777(– 3.319)	1.379	
	γ_1	0.145 (0.122)	0.059	8.299
*Are not given in Agresti (1996). The values in brackets are the results of the first iteration, as in Table 4.1.				

Assuming the logistic density (4.2.2) in every situation is, however, too restrictive. Tiku and Vaughan (1997) extend the modified likelihood methodology to non-logistic density functions as follows.

4. 8 NON LOGISTIC DENSITY FUNCTIONS

Let f(u) be any completely specified density function with $F(z) = \int_{-\infty}^{z} f(u)\,du$ as the cumulative density function. For example, Tiku and Vaughan (1997) consider the following densities:

(1) normal, (2) logistic, (3) Student t,
(4) extreme value I, (5) extreme value II, and (6) Generalized Logistic, (4.8.1)

which represent a very wide range of symmetric and skew distributions. The functional forms of the densities (1)-(6) above and their cumulative density functions are given in Appendix 4A. The modified likelihood equations are exactly of the same form as (4.3.5) – (4.3.7), and their solutions are exactly of the same form as (4.3.8). The only difference is that for the logistic, $t_{(i)}$ $(1 \le i \le n)$ are obtained from (4.6.4) and the α_i and β_i coefficients are obtained from (4.6.1) – (4.6.2). For a nonlogistic density function, $t_{(i)}=F^{-1}(i/(n+1))$, $1 \le i \le n$, and the coefficients $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ used in the computation of the MML estimators are obtained from (4.3.2) – (4.3.3). The cumulative density functions are given in Appendix 4A and it is easy to find $t_{(i)}$ and the coefficients needed to calculate the MML estimators. For a given data set (y_i, x_i), $1 \le i \le n$, we proceed as follows (Tiku and Vaughan, 1997).

Table 4.4: The MML estimates and their standard errors, and the values of the W and G statistics for the CHD data; n = 100, SE = Standard error.

			Statistic				Statistic	
Coeff	MMLE	SE	W	G	MMLE	SE	W	G
	Normal				Logistic			
γ_0	– 3.146	0.625			– 2.927	0.625		
γ_1	0.0658	0.0134	4.91	29.16	0.0611	0.0133	4.59	29.31
	Student t (ν = 7)				Student t (ν = 5)			
γ_0	– 2.897	0.634			– 2.747	0.610		
γ_1	0.0605	0.0134	4.51	29.20	0.0574	0.0129	4.45	29.25
	Student t (ν = 3)				Extreme value II			
γ_0	– 2.215	0.514			– 3.313	0.648		
γ_1	0.0462	0.0109	4.24	29.42	0.0617	0.0127	4.86	29.12
	Generalized Logistic (b = 2)				Generalized Logistic (b = 0.5)			
γ_0	– 2.951	0.764			– 2.733	0.532		
γ_1	0.0787	0.0169	4.66	29.02	0.0490	0.0107	4.58	29.36

Graphical technique: A Q–Q plot is obtained by plotting the ordered x-values (4.2.7) against the " population" quantiles $t_{(i)}$ determined by (4.3.4). A density function f(u) which yields a straight line (or closest to such) is one which is apparently most appropriate (Chapter 9). The Q–Q plots of the data considered in Example 4.1 above are given in Tiku and Vaughan (1997) for the normal, logistic, extreme value, Generalized Logistic and Student t. The Q–Q plot based on the Generalized Logistic with b=0.5 most closely approximates a straight line over the entire data set. Since the method of modified likelihood automatically assigns very small weights to the extreme order statistics in a sample from long-tailed symmetric distributions (Chapter 2), Student t with ν = 3 is also a reasonable model. For illustration, the MML estimates of γ_0 and γ_1 and their standard errors for various density functions are given in Table 4.4, reproduced from Tiku and Vaughan (1997, p. 892). For a comparison to be meaningful, the scale of the x-values has to be the same for all density functions. This is accomplished by dividing each x_i $(1 \le i \le n)$ by the square root of the dispersion

$$\sigma_x^2 = \int_{-\infty}^{\infty} u^2 f(du)\, du - \left[\int_{-\infty}^{\infty} uf(u)\, du\right]^2. \qquad (4.8.2)$$

For the logistic density, for example, $\sigma_x^2 = 3.2898$.

It can be seen that considerable improvements over logistic are realized by utilizing either the Student t ($\nu = 3$) or the generalized logistic ($b = 0.5$). For example, the relative efficiency

$$E = 100\,(\text{Variance under nonlogistic density}) / (\text{Variance under logistic}) \qquad (4.8.3)$$

is given below for the CHD data:

Density	Relative Efficiency	
	$\hat{\gamma}_0$	$\hat{\gamma}_1$
Student t ($\nu = 3$)	67	67
Generalized logistic ($b = 0.5$)	72	64

Moreover, the two density functions yield almost the highest values of the W and G statistics (Table 4.4). It is, therefore, very important to process a given data set through various density functions and determine the best fit. The method of modified likelihood estimation makes this task very easy.

For the data in Example 4.2 with $n = 10$, it is the extreme value I density function which provides the best fit. For example, we have the following MML estimates and their standard errors, and the corresponding W and G statistics:

Coeff.	MMLE	SE	W	G	MMLE	SE	W	G
	Logistic				Extreme Value I			
γ_0	– 0.552	0.457			– 0.152	0.272		
γ_1	0.761	0.554	1.37	2.47	0.622	0.376	1.65	3.39

Here the relative efficiencies E are 35 and 46 percent for $\hat{\gamma}_0$ and $\hat{\gamma}_1$, respectively. The logistic density is clearly inadequate.

Remark: Treating the covariate X as nonstochastic is, in numerous situations, an oversimplification. In general, X will be stochastic in nature and its distribution should be incorporated in the likelihood function. This will be done later in Section 4.11.

4.9 QUADRATIC MODEL

In prevous chapters, $z = \eta(x)$ was taken to be a linear function so that $z = \gamma_0 + \gamma_1 x$. The modified likelihood methodology readily extends to higher order relationships, e.g.,

$$z = \gamma_0 + \gamma_1 x + \ldots + \gamma_j x^j\,. \qquad (4.9.1)$$

Consider $j = 2$ (Oral, 2002) so that,

$$z_{(i)} = \gamma_0 + \gamma_1 x_{(i)} + \gamma_2 x_{(i)}^2; \qquad (4.9.2)$$

γ_1 and γ_2 are a priori positive. Although the method works for any density, but we will consider the logistic for illustration. Writing,

$$\delta_i = \alpha_i w_i - (1-\alpha_i)(1-w_i) \quad \text{and} \quad m_i = \beta_i,\ w_i = y_{[i]} \quad (1 \le i \le n), \tag{4.9.3}$$

where α_i and β_i are given by (4.6.2), the modified likelihood equations are

$$\frac{\partial \ln L}{\partial \gamma_0} \cong \frac{\partial \ln L^*}{\partial \gamma_0} = \delta - m\gamma_0 - (\Sigma m_i x_{(i)})\gamma_1 - (\Sigma m_i x_{(i)}^2)\gamma_2 = 0, \tag{4.9.4}$$

$$\frac{\partial \ln L}{\partial \gamma_1} \cong \frac{\partial \ln L^*}{\partial \gamma_1} = \Sigma\delta_i x_{(i)} - (\Sigma m_i x_{(i)})\gamma_0 - (\Sigma m_i x_{(i)}^2)\gamma_1 - (\Sigma m_i x_{(i)}^3)\gamma_2 = 0 \quad \text{and} \tag{4.9.5}$$

$$\frac{\partial \ln L}{\partial \gamma_2} \cong \frac{\partial \ln L^*}{\partial \gamma_2} = \Sigma\delta_i x_{(i)}^2 - (\Sigma m_i x_{(i)}^2)\gamma_0 - (\Sigma m_i x_{(i)}^3)\gamma_1 - (\Sigma m_i x_{(i)}^4)\gamma_2 = 0, \tag{4.9.6}$$

each sum carried over i = 1, 2, . . . , n; $m = \Sigma m_i$. The solutions of (4.9.4) – (4.9.6) are the following MML estimators:

$$\begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} m & \Sigma m_i x_{(i)} & \Sigma m_i x_{(i)}^2 \\ \Sigma m_i x_{(i)} & \Sigma m_i x_{(i)}^2 & \Sigma m_i x_{(i)}^3 \\ \Sigma m_i x_{(i)}^2 & \Sigma m_i x_{(i)}^3 & \Sigma m_i x_{(i)}^4 \end{bmatrix}^{-1} \begin{bmatrix} \delta \\ \Sigma\delta_i x_{(i)} \\ \Sigma\delta_i x_{(i)}^2 \end{bmatrix} \quad (d = \Sigma\delta_i). \tag{4.9.7}$$

The estimators $\hat{\gamma}_0$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are computed in a few iterations, exactly the same way as in Section 4.3. If $\hat{\gamma}_i$ (i = 1, 2) turns out to be negative, equate it to 0.000001 (a small number).

Asymptotic covariance matrix: Exactly for the same reasons as before, the asymptotic covariance matrix of the estimators in (4.9.7) is

$$V = \begin{bmatrix} \Sigma Q_i & \Sigma Q_i x_i & \Sigma Q_i x_i^2 \\ \Sigma Q_i x_i & \Sigma Q_i x_i^2 & \Sigma Q_i x_i^3 \\ \Sigma Q_i x_i^2 & \Sigma Q_i x_i^3 & \Sigma Q_i x_i^4 \end{bmatrix}^{-1}, \tag{4.9.8}$$

Q_i is the same as in (4.4.5); V is estimated by $\hat{V}$ obtained by replacing Q_i by

$$\hat{Q}_i = f^2(\hat{z}_i)/[F(\hat{z}_i)(1 - F(\hat{z}_i))],\ \hat{z}_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_i + \hat{\gamma}_2 x_i^2. \tag{4.9.9}$$

It may be noted that the matrix V above and its estimator $\hat{V}$ are positive definite. This follows from the fact that their minors are all positive. Consider, for example,

$$M_{11} = \begin{bmatrix} \Sigma Q_i x_i^2 & \Sigma Q_i x_i^3 \\ \Sigma Q_i x_i^3 & \Sigma Q_i x_i^4 \end{bmatrix} = (\Sigma Q_i x_i^2)(\Sigma Q_i x_i^4) - (\Sigma Q_i x_i^3)^2.$$

Realizing that Q_i is positive, we define $a_i = \sqrt{Q_i}\, x_i$ and $b_i = \sqrt{Q_i}\, x_i^2$.

By Cauchy-Schwarz inequality

$$M_{11} = (\Sigma a_i^2)(\Sigma b_i^2) - (\Sigma a_i b_i)^2 > 0.$$

Since $x_i^2 \ne cx_i$ (c being a constant), $M_{11} \ne 0$. Similarly, all other minors are positive.

Example 4.4: Consider the CHD data with n = 100 observations considered in Example 4.1. Assuming the logistic density and the quadratic relationship $z = \gamma_0 + \gamma_1 x + \gamma_2 x^2$, we have the following MML estimates and their standard errors. Three iterations were needed for the estimates to stabilize sufficiently enough.

Coefficient	MML estimate*	Standard error	W	G
γ_0	– 0.440 (– 0.408)	0.318		
γ_1	1.283 (1.160)	0.281	4.561	
γ_2	0.075 (0.000)	0.292	0.256	29.38

* The values in brackets are the results of the first iteration.

For this data, the linear relationship $z = \gamma_0 + \gamma_1 x$ is clearly adequate since $\hat{\gamma}_2$ is not significantly different from zero.

Consider the data referred to in Example 4.3 with n = 27 observations. Here the results are as follows; $z = \gamma_0 + \gamma_1 x + \gamma_2 x^2$:

Coefficient	MML Estimate	Standard Error	W
γ_0	– 0.222	0.648	
γ_1	2.940	1.330	2.210
γ_2	0.000	—	—

Since $\hat{\gamma}_2 = 0.000$, the G statistic is the same as before (G = 8.299). Clearly, the quadratic model does not give a better fit than a linear model. The null hypothesis $H_0 : \gamma_1 = 0$ is rejected in favour of $H_1 : \gamma_1 > 0$, at 5% significance level. The estimated model $\pi(x) = F(z)$ is

$$z = -0.222 + 2.940\, x. \qquad (4.9.10)$$

4.10 MULTIPLE COVARIATES

The method of modified likelihood estimation readily generalizes to k (≥ 2) covariates. Here, we have

$$z_i = \gamma_0 + \sum_{j=1}^{k} \gamma_j x_{ji} \quad (1 \le i \le n) \qquad (4.10.1)$$

and assume that

$$E(Y \mid X_1 = x_{1i}, \ldots\ldots, X_k = x_{ki}) = F(z_i) = \int_{-\infty}^{z} f(u)\, du\,. \qquad (4.10.2)$$

A priori we have no reason to believe that all the covariates X_j, $1 \le j \le k$, are not equally effective in increasing the response Y. Therefore, we initially take γ_j's all equal and positive in which case the ordered z-values $z_{(i)}$ correspond to the ordered x-values $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$, where

$$x_i = x_{1i} + x_{2i} + \ldots + x_{ki}\,,\ 1 \le i \le n. \qquad (4.10.3)$$

Let $y_{[i]}$ be the y-observation which corresponds to $x_{(i)}$, $1 \le i \le n$. The modified likelihood equations are exactly of the same form as (4.3.5) – (4.3.7). Their solutions are the following MML estimators:

$$\hat{\Gamma} = (X'MX))^{-1} X'\Delta, \qquad (4.10.4)$$

where $\Gamma' = (\gamma_0, \gamma_1, \ldots\ldots, \gamma_k)$ and

$$X = \begin{bmatrix} 1 & x_{11}^* & x_{21}^* & . & . & . & x_{k1}^* \\ 1 & x_{12}^* & x_{22}^* & . & . & . & x_{k2}^* \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & x_{1n}^* & x_{2n}^* & . & . & . & x_{kn}^* \end{bmatrix}$$

is the matrix

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & . & . & . & x_{k1} \\ 1 & x_{12} & x_{22} & . & . & . & x_{k2} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & x_{1n} & x_{2n} & . & . & . & x_{kn} \end{bmatrix}$$

with rows arranged so as to correspond to the ordered variates $x_{(i)}$, $1 \le i \le n$,

$$M = \begin{pmatrix} m_1 & 0 & . & . & 0 \\ 0 & m_2 & . & . & 0 \\ . & . & . & . & 0 \\ . & . & . & . & 0 \\ 0 & . & . & . & m_n \end{pmatrix} \quad \text{and} \quad \Delta = \begin{pmatrix} \delta_1 \\ \delta_2 \\ . \\ . \\ \delta_n \end{pmatrix}; \tag{4.10.5}$$

m_i and δ_i are exactly the same as in (4.3.7).

The MML estimates are obtained exactly the same way as before, initially calculating $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ from (4.3.2) – (4.3.3) with $t_{(i)}$ obtained from (4.3.4). Subsequently, $t_{(i)}$ is replaced by

$$t_i = \hat{\gamma}_0 + \sum_{j=1}^{k} \hat{\gamma}_j x_{ji}^* , \quad 1 \le i \le n, \tag{4.10.6}$$

and the process repeated. The coefficients $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ stabilize very quickly in a few iterations.

Exactly along the same lines as before, it is not difficult to prove that the MML estimators (4.10.4) are asymptotically unbiased and their covariance matrix (asymptotic) is

$$V = \begin{bmatrix} \Sigma Q_i & \Sigma Q_i x_{1i} & . & . & \Sigma Q_i x_{ki} \\ \Sigma Q_i x_{1i} & \Sigma Q_i x_{1i}^2 & . & . & \Sigma Q_i x_{1i} x_{ki} \\ . & . & . & . & . \\ . & . & . & . & . \\ \Sigma Q_i x_{ki} & \Sigma Q_i x_{1i} x_{ki} & . & . & \Sigma Q_i x_{ki}^2 \end{bmatrix}^{-1}, \tag{4.10.7}$$

each sum carried over $1 \le i \le n$; Q_i is the same as (4.4.5) but z_i is given by (4.10.1) and is estimated by

$$\hat{z}_1 = \hat{\gamma}_0 + \sum_{j=1}^{k} \hat{\gamma}_j x_{ji} , \quad 1 \le i \le n. \tag{4.10.8}$$

The matrix V and its estimate $\hat{V}$ are both positive definite.

Example 4.5: Consider the data from Finney (1947), reproduced in Appendix 4C for ready reference. The data is also given in Aitken et al. (1989, Appendix 4, p. 346). There are n = 39 observations on three variables: VOL= Volume of air inspired, RATE = Rate of air inspired, and RESP = presence of vaso-constriction. RESP is the response variable, with value 1 if no vaso-constriction exists and 0 if there is.

Assuming the logistic density for f(u) as in (4.2.1) – (4.2.2) and

$$z = \gamma_0 + \gamma_1 \text{ RATE} + \gamma_2 \text{ VOL}, \tag{4.10.9}$$

Aitken et al. (1989) give the MML estimates and their standard errors. The values are given in Table 4.5. Also given are the MML estimates and their standard errors (Tiku and Vaughan, 1997, p. 895); five iterations were needed for the MML estimates to stabilize sufficiently enough. It is seen that the MML estimates are essentially the same as the ML estimates and are equally efficient.

Table 4.5: The ML and MML estimates for the Finney data, under logistic density.

	Coefficient	Estimate*	Standard Error	W	G
ML	γ_0	– 9.530	3.224		
	γ_1	2.649	0.912	2.90	
	γ_2	3.882	1.425	2.72	24.27
MML	γ_0	– 9.530 (– 7.730)	3.233 (2.605)		
	γ_1	2.649 (2.232)	0.914 (0.771)	2.90	
	γ_2	3.882 (3.013)	1.429 (1.138)	2.72	24.27 (23.83)

* The values in brackets are the results of the first iteration based on $t_{(i)}$ given in (4.6.4).

Tiku and Vaughan (1997) processed this data through the density functions (4.8.1). It is the Generalized Logistic (b = 0.5) which provides a better model with the following results. Each covariate was divided by a suitable constant 1.4142 so as to have the same dispersion σ_x^2 as that of the logistic (i.e., 3.2898).

Generalized Logistic (b = 0.5)				
Coefficient	MML Estimate	Standard Error	W	G
γ_0	– 9.577	3.134		
γ_1	2.398	0.831	2.89	
γ_2	3.684	1.369	2.69	26.06

It can be seen that the standard errors are smaller, W statistic is almost the same, and the G statistic is larger. This is a substantial improvement.

Remark: In the context of binary regression it clearly seems advantageous to process a given data set through various density functions to locate the one which yields best results. The method of modified likelihood makes this task possible. The method also makes it possible to study the effect of outliers on the MML estimators as we will see later in Chapter 8.

4.11 STOCHASTIC COVARIATES

Treating the covariate X as nonstochastic in all situations is an oversimplification. In Example 4.5, for example, the rate of air inspired is not measureable without error and should not, therefore, be treated as a nonstochastic variable. In Example 4.2, x-values were generated from a normal distribution N(0, 1) but treated as nonstochastic values in subsequent analysis (Hosmer and Lemeshow, 1989). We will now consider the more realistic situation when X is also a stochastic variable and is treated as such. We will show that the resulting estimators $\hat{\gamma}_0$ and $\hat{\gamma}_1$ have considerably smaller standard errors than those obtained by imposing the unnecessary restriction that X is nonstochastic.

Generalized Logistic: Consider the situation when the covariate X in Section 4.2 has the Generalized Logistic distribution (Oral, 2002)

$$h(x) = b\gamma_1 e^{-(\gamma_0 + \gamma_1 x)} / \{1 + e^{-(\gamma_0 + \gamma_1 x)}\}^{b+1}, \; -\infty < x < \infty, \qquad (4.11.1)$$

with $\gamma_1 > 0$. In principle, f(u) in (4.2.1) can be any density function. For illustration, however, we take $f(u) = be^{-u}/(1 + e^{-u})^{b+1}$ $(-\infty < u < \infty)$, the same as h(z); $z = \gamma_0 + \gamma_1 x$.

Writing,

$$\text{🚔}(x_i) = F(z_i) = \int_{-\infty}^{z_i} f(u)\, du \qquad (4.11.2)$$

where $z_i = \gamma_0 + \gamma_1 x_i$ $(\gamma_1 > 0)$, the likelihood function is

$$L = L_x L_{y|x} = \left[\prod_{i=1}^{n} \gamma_1 h(z_i)\right]\left\{\prod_{i=1}^{n} [F(z_i)]^{y_i} [1 - F(z_i)]^{1-y_i}\right\}, \qquad (4.11.3)$$

$$h(z_i) = be^{-z_i}/\{1 + e^{-z_i}\}^{b+1}.$$

This gives

$$\ln L = n \ln \gamma_1 + \sum_{i=1}^{n} \ln h(z_i) + \sum_{i=1}^{n} \{y_i \ln F(z_i) + (1 - y_i) \ln (l - F(z_i))\}. \qquad (4.11.4)$$

The likelihood equations are

$$\frac{\partial \ln L}{\partial \gamma_0} = -n + (b+1)\sum_{i=1}^{n} g(z_i) + \sum_{i=1}^{n} \{y_i g_1(z_i) - (1 - y_i) g_2(z_i)\} = 0 \quad \text{and} \qquad (4.11.5)$$

$$\frac{\partial \ln L}{\partial \gamma_1} = \frac{n}{\gamma_1} - \sum_{i=1}^{n} x_i + (b+1)\sum_{i=1}^{n} x_i g(z_i) + \sum_{i=1}^{n} x_i \{y_i g_1(z_i) - (1 - y_i) g_2(z_i)\} = 0; \qquad (4.11.6)$$

$$g(z) = \frac{e^{-z}}{(1 + e^{-z})} = \frac{1}{1 + e^{z}}, \; g_1(z) = \frac{f(z)}{F(z)} \quad \text{and} \quad g_2(z) = \frac{f(z)}{1 - F(z)}. \qquad (4.11.7)$$

The functions $g_1(z)$ and $g_2(z)$ are exactly the same as (4.2.12). The equations (4.11.5)-(4.11.6) are intractable and solving them by iteration is problematic, as is generally true with likelihood equations.

4.12 MODIFIED LIKELIHOOD

Let

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \qquad (4.12.1)$$

be the order statistics of the n random observations available on the covariate X. Since $\gamma_1 > 0$, the corresponding ordered z_i variates are

$$z_{(i)} = \gamma_0 + \gamma_1 x_{(i)}, \; 1 \le i \le n. \qquad (4.12.2)$$

The likelihood equations (4.11.5)-(4.11.6) expressed in terms of $z_{(i)}$ are

$$\frac{\partial \ln L}{\partial \gamma_0} = -n + (b+1)\sum_{i=1}^{n} g(z_{(i)}) + \sum_{i=1}^{n} \{w_i g_1(z_{(i)}) - (1 - w_i) g_2(z_{(i)})\} = 0 \qquad (4.12.3)$$

and

$$\frac{\partial \ln L}{\partial \gamma_1} = \frac{n}{\gamma_1} - \sum_{i=1}^{n} x_{(i)} + (b+1)\sum_{i=1}^{n} x_{(i)} g(z_{(i)})$$

$$+ \sum_{i=1}^{n} x_{(i)} \{w_i g_1(z_{(i)}) - (1 - w_i) g_2(z_{(i)})\} = 0, \qquad (4.12.4)$$

$w_i = y_{[i]}$ being the concomitant of $x_{(i)}$ $(1 \le i \le n)$ as in (4.2.9). To obtain the modified likelihood equations which are asymptotically equivalent to the likelihood equations (4.11.5) – (4.11.6) above, we linearize the g-functions:

$$\gamma_1(z_{(i)}) \cong \alpha_{1i} - \beta_{1i} z_{(i)} \quad \text{and} \quad g_2(z_{(i)}) \cong \alpha_{2i} + \beta_{2i} z_{(i)}, \quad 1 \le i \le n; \tag{4.12.5}$$

the coefficients $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ are given in (4.3.2) and (4.3.3), respectively, with $t_{(i)}$ obtained from

$$\int_{-\infty}^{t} f(u)\,du = \frac{i}{n+1},\ t = t_{(i)}; \quad 1 \le i \le n. \tag{4.12.6}$$

From the first two terms of a Taylor series expansion of $g\{z_{(i)}\}$ around $t_{(i)}^* = E\{z_{(i)}\}$, we have

$$g(z_{(i)}) \cong \alpha_i^* - \beta_i^* z_{(i)},\ 1 \le i \le n. \tag{4.12.7}$$

The coefficients (α_i^*, β_i^*) are given by the equations

$$\alpha_i^* = \{1 + e^{a_i}\}^{-1} + \beta_i^* a_i \quad \text{and} \quad \beta_i^* = e^{a_i} / (1 + e^{a_i})^2, \quad a_i = t_{(i)}^*. \tag{4.12.8}$$

For $n \ge 10$, the approximate values of $t_{(i)}^* = a_i$ obtained from the following equations are used:

$$\int_{-\infty}^{a_i} h(z)dz = q_i = \frac{i}{n+1} \quad (1 \le i \le n) \tag{4.12.9}$$

which for the Generalized Logistic gives $a_i = -\ln(q_i^{-1/b} - 1)$ as in (2.5.6). Realize that $t_{(i)}$ is initially equal to a_i, $1 \le i \le n$.

4.13 THE MML ESTIMATORS

Incorporating (4.12.5) – (4.12.7) in (4.12.3) – (4.12.4) and solving the resulting equations, we obtain the MML estimators:

$$\hat{\gamma}_0 = (1/m)(\delta - n) - \hat{\gamma}_1 \bar{x}_{(i)} \quad \text{and} \quad \hat{\gamma}_1 = \{B + \sqrt{(B^2 + 4nC)}\}/2C \tag{4.13.1}$$

where $\delta = \sum_{i=1}^{n} \delta_i$, $m = \sum_{i=1}^{n} m_i$, $\bar{x}_{(.)} = (1/m)\sum_{i=1}^{n} m_i x_{(i)}$, $B = \sum_{i=1}^{n} (\delta_i - 1)(x_{(i)} - \bar{x}_{(.)})$

and

$$C = \sum_{i=1}^{n} m_i (x_{(i)} - x_{(.)})^2 = \sum_{i=1}^{n} m_i x_{(i)}^2 - (1/m)\left(\sum_{i=1}^{n} m_i x_{(i)}\right)^2; \tag{4.13.2}$$

$$\delta_i = w_i \alpha_{1i} - (1 - w_i)\alpha_{2i} + (b+1)\alpha_i^* \quad \text{and} \quad m_i = w_i \beta_{1i} + (1 - w_i)\beta_{2i} + (b+1)\beta_i^*.$$

Realize that m_i is positive. Consequently, $\hat{\gamma}_1$ is positive.

Revised estimates: The estimators above are sharpened by doing a few iterations as explained in Section 4.3 (equation (4.3.9)). On the second and subsequent iterations, $t_{(i)}$ in (4.3.2) – (4.3.3) are replaced by

$$t_{(i)} = \hat{\gamma}_0 + \hat{\gamma}_1 x_{(i)}, \quad 1 \le i \le n, \tag{4.13.3}$$

and the revised estimates calculated from (4.13.1) – (4.13.2) as explained in Section 4.3. No revision is needed in the coefficients α_i^* and β_i^* given in (4.12.8); they are computed only once.

Remark: The estimator $\hat{\gamma}_0$ is scale invariant and $\hat{\gamma}_1$ is location invariant, and $\hat{\gamma}_1$ is real and positive.

4.14 ASYMPTOTIC PROPERTIES

Since the MML estimators are asymptotically equivalent to the ML estimators (Chapter 2, Appendix 2A), they have all the asymptotic properties of the ML estimators. In particular, they are asymptotically unbiased and their variances and the covariance (asymptotic) are given by

$$V = \begin{bmatrix} -E\left(\frac{\partial^2 \ln L}{\partial \gamma_0^2}\right) & -E\left(\frac{\partial^2 \ln L}{\partial \gamma_0 \partial \gamma_1}\right) \\ -E\left(\frac{\partial^2 \ln L}{\partial \gamma_0 \partial \gamma_1}\right) & -E\left(\frac{\partial^2 \ln L}{\partial \gamma_1^2}\right) \end{bmatrix}^{-1} . \qquad (414.1)$$

Writing (Oral, 2002)

$$Q = \sum_{i=1}^{n} Q_i, \; P_1 = nb/(b+2) \quad \text{and} \quad Q_i = f^2(z_i)/[F(z_i)(1 - F(z_i))], \qquad (4.14.2)$$

the elements of V are (Appendix 4D)

$$-E\left(\frac{\partial^2 \ln L}{\partial \gamma_0^2}\right) = Q + P_1, \qquad (4.14.3)$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \gamma_0 \partial \gamma_1}\right) = \frac{1}{\gamma_1}\{-(Q + P_1)\gamma_0 + [\psi(b) - \psi(1)]Q \qquad (4.14.4)$$

$$+ [\psi(b+1) - \psi(2)]P_1\} \text{ and } -E\left(\frac{\partial^2 \ln L}{\partial \gamma_1^2}\right) = \frac{1}{\gamma_1^2}\{(Q + P_1)\gamma_0^2 - 2\gamma_0([\psi(b) - \psi(1)]Q$$
$$+ [\psi(b+1) - \psi(2)]P_1) + ([\psi'(b) + \psi'(1)] + [\psi(b) - \psi(1)]^2)Q$$
$$+ (\psi'(b+1) + \psi'(2) + [\psi(b+1) - \psi(2)]^2)P_1 + n\}; \qquad (4.14.5)$$

$\psi(x)$ is the psi-function and $\psi'(x)$ its derivative as said earlier. Its values are given in Appendix 2D (Chapter 2).

Comment: For b = 1 (logistic density), the expressions above simplify, that is, $P_1 = n/3$ and

$$-E\left(\frac{\partial^2 \ln L}{\partial \gamma_0^2}\right) = Q + \frac{n}{3}, \quad -E\left(\frac{\partial^2 \ln L}{\partial \gamma_0 \partial \gamma_1}\right) = -\frac{\gamma_0}{\gamma_1}\left(Q + \frac{n}{3}\right) \text{ and}$$

$$-E\left(\frac{\partial^2 \ln L}{\partial \gamma_1^2}\right) = \frac{1}{\gamma_1^2}\left\{\left(Q + \frac{n}{3}\right)\gamma_0^2 + 3.2898\,Q + 1.4299n\right\}. \qquad (4.14.6)$$

In particular,

$$\text{Var}(\hat{\gamma}_1) \cong \gamma_1^2/(3.2898\,Q + 1.4299n); \qquad (4.14.7)$$

its estimate is obtained by replacing γ_1 by $\hat{\gamma}_1$ and Q by $\hat{Q} = \sum_{i=1}^{n} \hat{Q}_i$,

where $$\hat{Q}_i = f^2(\hat{z}_i)/[F(\hat{z}_i)(1 - F(\hat{z}_i))], \; \hat{z}_i = \hat{\gamma}_0 + \hat{\gamma}_1 \bar{x}_i. \qquad (4.14.8)$$

Example 4.5: Consider again the CHD data with n=100 observations. Assume the logistic density (4.11.1). We reproduce below the previous estimates and their standard errors based only on the conditional likelihood $L_{y|x}$, and the new estimates and their standard errors based on the full likelihood $L = L_x L_{y|x}$:

Estimates for the CHD data						
X		$\hat{\gamma}_0$	SE($\hat{\gamma}_0$)	$\hat{\gamma}_1$	SE($\hat{\gamma}_1$)	W
Nonstochastic	ML	– 5.310	1.134	0.111	0.024	—
	MML	– 5.309	1.134	0.111	0.024	4.63
Stochastic	MML	– 6.181	0.463	0.136	0.010	13.6

There is not much change in the numerical values of the estimators but their standard errors are substantially smaller than those based on $L_{y|x}$. This was to be expected since the likelihood function is now used in full; see also Chapter 3 (Section 3.11).

We also did similar computations with Agresti (1996) data considered in Example 4.3. The results are given below for the logistic density:

The MML estimates and standard errors

	$\hat{\gamma}_0$	SE($\hat{\gamma}_0$)	$\hat{\gamma}_1$	SE($\hat{\gamma}_1$)	W
X nonstochastic	– 3.777	1.379	0.145	0.059	2.46
X stochastic	– 3.688	0.576	0.175	0.024	7.29

Again, there is not much change in the estimates but the standard errors are substantially smaller. The advantages of using the full likelihood are irrefutable**.** The method generalizes to (k > 1) risk factors (covariates).

The G statistic is exactly similar to (4.5.3) but calculated by using the new estimates (4.13.1). Its value does not change much. The statistics W and G will be shown to be robust to deviations from an assumed distribution and to outliers (Chapter 8).

4.15 SYMMETRIC FAMILY

Assume that the covariate X has a distribution in the symmetric family ($\gamma_1 > 0$)

$$h(x) = \frac{\gamma_1}{\sqrt{k}\,\beta(1/2, p-1/2)} \left\{1 + \frac{1}{k}(\gamma_0 + \gamma_1 x)^2\right\}^{-p}, \quad -\infty < x < \infty; \tag{4.15.1}$$

$k = 2p - 3$, $p \ge 2$. Here, the modified likelihood equations are

$$\frac{\partial \ln L}{\partial \gamma_0} \cong \frac{\partial \ln L^*}{\partial \gamma_0} = \sum_{i=1}^{n} (\delta_i - m_i z_{(i)}) = 0 \quad \text{and} \tag{4.15.2}$$

$$\frac{\partial \ln L}{\partial \gamma_1} \cong \frac{\partial \ln L^*}{\partial \gamma_1} = \frac{n}{\gamma_1} + \sum_{i=1}^{n} x_{(i)}(\delta_i - m_i z_{(i)}) = 0; \tag{4.15.3}$$

$$\gamma_i = w_i\alpha_{1i} - (1 - w_i)\alpha_{2i} - (2p/k)\alpha_i \quad \text{and} \tag{4.15.4}$$
$$m_i = w_i\beta_{1i} - (1 - w_i)\beta_{2i} + (2p/k)\beta_i \;(w_i = y_{[i]}).$$

The coefficients (α_{1i}, β_{1i}) and (α_{2i}, β_{2i}) are the same as in (4.3.2) – (4.3.3) with $t_{(i)}$ obtained from the equation

$$\int_{-\infty}^{t(i)} f(u)\,du = \frac{i}{n+1} \;(1 \le i \le n); \quad f(u) = \frac{1}{\sqrt{k}\,\beta(1/2, p-1/2)}\left(1 + \frac{1}{k}u^2\right)^{-p}. \tag{4.15.5}$$

The coefficients (α_i, β_i) are the same as in (2.3.14) with $t_{(i)}$ replaced by a_i and the latter obtained from $\int_{-\infty}^{a_i} f(z)\,dz = i/(n+1)$.

The solutions of (4.15.2) – (4.15.3) are the MML estimators (Oral, 2002):

$$\hat{\gamma}_0 = (\delta/m) - \hat{\gamma}_1 \bar{x}_{(.)} \quad \text{and} \quad \hat{\gamma}_1 = \{B + \sqrt{(B^2 + 4nC)}\}/2C \qquad (4.15.6)$$

where
$$\delta = \sum_{i=1}^{n} \delta_i, \quad m = \sum_{i=1}^{n} m_i, \bar{x}_{(.)} = (1/m)\sum_{i=1}^{n} m_i x_{(i)},$$

$$B = \sum_{i=1}^{n} \delta_i (x_{(i)} - \bar{x}_{(.)}) \quad \text{and} \quad C = \sum_{i=1}^{n} m_i (x_{(i)} - \bar{x}_{(.)})^2 .$$

The computation proceeds as in Section 4.13; $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ are revised in each iteration but not (α_i, β_i). The latter are computed only once. If for a sample, C in (4.15.6) assumes a negative value, α_i is replaced by 0 and β_i is replaced by $\beta_i^* = 1/\{(1 + (1 + (1/k)a_i^2\}$, and the estimators calculated.

Asymptotic covariance matrix: As in Section 4.14, the asymptotic variances and the covariance of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ in (4.15.6) are given by

$$V = \begin{bmatrix} (Q+R)^{-1}\gamma_o^2 + (Q+P)^{-1} & (Q+R)^{-1}\gamma_0\gamma_1 \\ (Q+R)^{-1}\gamma_0\gamma_1 & (Q+R)^{-1}\gamma_1^2 \end{bmatrix}; \qquad (4.15.7)$$

Q and Q_i are exactly the same as those in (4.14.2) and $f(z_i)$ is the same as in (4.15.5) with u replaced by $z_i = \gamma_0 + \gamma_1 x_i \quad (1 \le i \le n)$, and

$$P = np(p - 1/2)/(p+1)(p - 3/2) \quad \text{and} \quad R = 2n(p - 1/2)/(p+1). \qquad (4.15.8)$$

The estimators $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are considerably more efficient than those based on the conditional likelihood $L_{y|x}$, as in Example 4.5.

Remark: Oral and Tiku (2003) take

$$z_i = \gamma_0 + \gamma_1(x_i - \mu)/\sigma \quad (1 \le i \le n) \qquad (4.15.9)$$

in (4.11.1) – (4.11.3) and work out the MML estimators of μ, σ, γ_0 and γ_1. The results are interesting. For example, the estimator of γ_1 is

$$\hat{\gamma}_1 = \hat{\sigma} \sum_{i=1}^{n} \delta_i (x_{(i)} - \bar{x}_{(.)}) \Big/ \sum_{i=1}^{n} m_i (x_{(i)} - \bar{x}_{(.)})^2 \qquad (4.15.10)$$

where δ_i and $\bar{x}_{(.)}$ are exactly the same as in (4.3.8) and $\hat{\sigma}$ is exactly the same as in (2.5.11); if $\hat{\gamma}_1$ turns out to be negative, it is equated to 0.000001. It is interesting to see that $\hat{\gamma}_1$ is location and scale invariant. The MML estimators of γ_1 given in (4.3.8) and (4.13.1) are not scale invariant and, for a comparison of efficiencies for different distributions, they have to be painstakingly adjusted for the scale. No such adjustments is needed for the estimator in (4.15.10).

SUMMARY

In binary regression, the dependent variable Y assumes only two values 0 or 1 and is related to a design variable (called risk factor) X through a link function. Specifically, $E(Y|X = x) = F(z)$ where $z = \gamma_0 + \gamma_1 x$ and F(z) is the cumulative density function of a completely specified density function f(u); F^{-1} is called the link function. We consider the estimation of γ_0

and γ_1 in this chapter. Traditionally, f(u) has been assumed to be logistic and X regarded as nonstochastic. This is too restrictive for real-life applications. We give solutions in both situations: (i) f(u) is not necessarily logistic and X is nonstochastic, and (ii) f(u) is not necessarily logistic and X is stochastic. We derive the MMLE and show that, besides being easy to compute, they are highly efficient. In situations when the MLE are available, we show that the MMLE are numerically the same (almost) as the MLE. We give procedures for testing an assumed value of γ_1. In particular, we show that assuming X nonstochastic (conditional approach) results in enormous loss of efficiency if X is in fact stochastic. We also give solutions when X is nonstochastic and $z = \gamma_0 + \gamma_1 x_1 + \ldots + \gamma_k x_k$ $(k \geq 2)$. We generalize the results to two or more nonstochastic risk factors.

APPENDIX 4A

DENSITY AND CUMULATIVE DENSITY FUNCTIONS

Density $(-\infty \leq u \leq \infty)$

Normal*: $(2\pi)^{-1/2} \exp(-u^2/2)$

Student t*: $$\frac{1}{\sqrt{\nu}\,\beta(1/2, \nu/2)}\left(1 + \frac{1}{\nu}t^2\right)^{-(\nu+1)/2}$$

Extreme value I: $e^{-u} \exp(-e^{-u})$

Extreme value II: $e^{u} \exp(-e^{u})$

Generalized logistic: $be^{-u}/(1 + e^{-u})^{b+1}$

Cumulative density

Normal: Intractable

Student t: $$\begin{cases} \dfrac{1}{2} + \dfrac{1}{\pi}\tan^{-1}\left(\dfrac{u}{\sqrt{\nu}}\right) + \dfrac{\sqrt{\nu}u}{(u^2+\nu)}\displaystyle\sum_{i=0}^{l-1}\dfrac{(1+u^2/\nu)^{-i}}{(2i+1)\beta(1/2, i+1/2)} & \text{if } \nu = 2l+1 \\ \dfrac{1}{2} + \dfrac{u}{2\pi\sqrt{(u^2+\nu)}}\displaystyle\sum_{i=0}^{l-1}\dfrac{(1+u^2/\nu)^{-i}}{\beta(1/2, i+1/2)} & \text{if } \nu = 2l \end{cases}$$

Extreme Value I: $\exp(-e^{-u})$

Extreme Value II: $1 - \exp(-e^{u})$

Generalized logistic: $(1 + e^{-u})^{-b}$

*IMSL subroutine is available to give the value of the cumulative density function for a given u.

APPENDIX 4B

Y = CHD status, X = Age

y:	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
x:	20	23	24	25	25	26	26	28	28	29	30	30	30	30	30	30
y:	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
x:	32	32	33	33	34	34	34	34	34	35	35	36	36	36	37	37
y:	0	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1
x:	37	38	38	39	39	40	40	41	41	42	42	42	42	43	43	43
y:	0	0	1	1	0	1	0	1	0	0	1	0	1	1	0	0
x:	49	50	50	51	52	52	53	53	54	55	55	55	56	56	56	57
y:	0	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1
x:	57	57	57	57	57	58	58	58	59	59	60	60	61	62	62	63
y:	0	1	1	1												
x:	64	64	65	69												

APPENDIX 4C

Y = Existence of vaso-constriction
X_1 = Volume in litres
X_2 = Rate in litres

y:	1	1	1	1	1	1	0	0	0	0
x_1:	3.70	3.50	1.25	0.75	0.80	0.70	0.60	1.10	0.90	0.90
x_2:	0.825	1.090	2.500	1.500	3.200	3.500	0.750	1.700	0.750	0.450
y:	0	0	0	1	1	1	1	1	0	1
x_1:	0.80	0.55	0.60	1.40	0.75	2.30	3.20	0.85	1.70	1.80
x_2:	0.570	2.750	3.000	2.330	3.750	1.640	1.600	1415	1.060	1.800
y:	0	0	0	0	1	0	1	0	1	0
x_1:	0.40	0.95	1.35	1.50	1.60	0.60	1.80	0.95	1.90	1.60
x_2:	2.000	1.360	1.350	1.360	1.780	1.500	1.500	1.900	0.950	0.400
y:	1	0	0	1	1	1	0	0	1	
x_1:	2.70	2.35	1.10	1.10	1.20	0.80	0.95	0.75	1.30	
x_2:	0.750	0.030	1.830	2.200	2.000	3.330	1.900	1.900	1.625	

APPENDIX 4D

ELEMENTS OF THE INFORMATION MATRIX

To find the expected values $-E(\partial^2 \ln L/\partial\gamma_0^2)$, $-E(\partial^2 \ln L/\partial\gamma_0\partial\gamma_1)$ and $-E(\partial^2 \ln L/\partial\gamma_1^2)$, we use the well known result that for a bounded function in two random variables,

$$E\{\eta(z, y)\} = E_z\{E_{y|z}\, \eta(z, y)\}. \tag{4D.1}$$

Consider, for example,

$$-\frac{1}{n} E\left(\frac{\partial^2 \ln L}{\partial\gamma_0 \partial\gamma_1}\right) = E\Bigg[-\frac{1}{n}\sum_{i=1}^{n} x_i \Bigg\{ y_i \frac{f'(z_i)F(z_i) - f^2(z_i)}{F^2(z_i)}$$

$$- (1 - y_i) \frac{f'(z_i)(1 - F(z_i)) + f^2(z_i)}{(1 - F(z_i))^2}\Bigg\}$$

$$+ \frac{b+1}{n}\sum_{i=1}^{n} \frac{x_i e^{-z_i}}{(1 + e^{-z_i})^2}\Bigg]$$

$$= E_z\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{z_i - \gamma_0}{\gamma_1}\right) Q_i + \frac{b+1}{n}\sum_{i=1}^{n}\left(\frac{z_i - \gamma_0}{\gamma_1}\right)\frac{e^{-z_i}}{(1 + e^{-z_i})^2}\right]$$

$$= E\left(\frac{z - \gamma_0}{\gamma_1}\right)\frac{1}{n}\sum_{i=1}^{n} Q_i + \frac{b+1}{\gamma_1} E\left\{\frac{z e^{-z}}{(1 + e^{-z})^2} - \gamma_0 \frac{e^{-z}}{(1 + e^{-z})^2}\right\}$$

$$= \frac{1}{n\gamma_1}\,[-(Q + P_1)\gamma_0 + \{\psi(b) - \psi(1)\}Q + \{\psi(b+1) - \psi(2)\}P_1]$$

from the results given in Appendix 2D (Chapter 2): $Q = \sum_{i=1}^{n} Q_i$ and $P_1 = nb/(b+2)$.

CHAPTER 5

Autoregressive Models in Normal and Non-Normal Situations

5.1 INTRODUCTION

In Chapter 3, we considered linear regression models with normal as well as non-normal error distributions. We did, however, assume that the errors are iid (identically and independently distributed). We derived the MML estimators and showed that they are highly efficient. In fact, they are asymptotically the minimum variance bound estimators. In numerous applications in biology, biomedical and agricultural sciences, business and economics and hydrosciences, however, the errors are not independently distributed but are correlated with one another. This gives rise to autoregressive models. There is a very extensive literature on the subject but based primarily on the normality assumption (Anderson, 1949; Durbin, 1960; Tiao and Tan, 1966; Beach and Mackinnon, 1978; Kramer, 1980; Magee et al. 1987; Velu and Gregory, 1987; Maller, 1989; Cogger, 1990; Weiss, 1990; Schäffler, 1991; Nagaraja, et al. , 1992). As mentioned in previous chapters, there is now a realization that non-normal distributions occur so frequently in practice. It is, therefore, of paramount importance to consider non-normal distributions in autoregressive models. In an interesting paper, Tan and Lin (1993) assumed normality but based their estimators on censored samples (considered here in Chapter 7). They showed that their estimators have good efficiency and robustness properties for numerous non-normal distributions. In this chapter, we consider non-normal distributions and derive MML estimators from complete samples. We show that they are remarkably efficient. In Chapter 8, we show that the estimators have excellent robustness properties. For illustration, we consider three families of distributions: gamma, short-tailed symmetric (STS) and long-tailed symmetric (LTS) distributions (Tiku et al., 1999; Akkaya and Tiku, 2001a, b; 2002a, b). The latter two families have been considered in Chapter 3 in the context of linear regression. We also consider the LS estimators; they have low efficiencies as compared to the MML estimators. In fact, the LS estimators have a disconcerting feature, namely, their relative efficiencies decrease as the sample size n increases. This is not a good prospect from a theoretical as well as a practical point of view.

5.2 A SIMPLE AUTOREGRESSIVE MODEL

A simple autoregressive model is given by

$$y_t = \mu' + \delta x_t + e_t$$
$$e_t = \phi\, e_{t-1} + a_t \quad (1 \le t \le n) \tag{5.2.1}$$

where y_t = observed value of a random variable y at time t

x_t = pre-determined value of a nonstochastic design variable x at time t

ϕ = autoregressive coefficient ($|\phi| < 1$) ;

a_t are called innovations and assumed to be iid. An alternative form of (5.2.1) is

$$y_t - \phi y_{t-1} = \mu + \delta\,(x_t - \phi x_{t-1}) + a_t \quad (1 \le t \le n). \tag{5.2.2}$$

Of particular interest here is the estimation of δ and testing the null hypothesis H_0: $\delta = 0$ when $\phi \ne 0$. For $\phi = 0$, (5.2.2) reduces to a simple linear regression model considered in Chapter 3. For $\delta = 0$, it reduces to a time series AR(1) model considered in Section 5.13.

There are two models for y_0 (Vinod and Shenton, 1996), x_0 being a nonstochastic design value, see also Dickey and Fuller (1979) and Abadir (1995);

Model A : y_0 is constant, $y_0 = 0$ in particular,

Model B : y_0 is random.

We will primarily work with Model B for its flexibility. We assume that y_0 has the same variance as y_t (Appendix 5A). Also, the likelihood function (conditional to y_0) under Model B is exactly of the same form as that under Model A. The conditional likelihood of y_t ($1 \le t \le n$) makes the estimation of parameters easier (of course, at the expense of losing the information in the observation y_0), and there are other advantages (Hamilton, 1994, p.123). Since n is large in the context of autoregression, losing this information is not of much consequence. Realize that (5.2.2) is a nonlinear model because of the parameter $\delta\phi$ and the estimation of parameters is, therefore, involved.

5.3 GAMMA DISTRIBUTION

Assume that the innovations a_t in (5.2.2) are iid and have the gamma distribution ($k > 1$)

$$G(k, \sigma):\ f(a) = \frac{1}{\sigma^k \Gamma(k)}\, e^{-a/\sigma}\, a^{k-1}, \quad 0 < a < \infty. \tag{5.3.1}$$

We assume that the shape parameter k is known. Determination of shape parameters such as k is considered in Chapter 9 in case their values are not known. See also Chapter 11.

Denote the sample observations by (y_i, y_{i-1}; x_i, x_{i-1}), $1 \le i \le n$, where x_i are design values and y_i ($1 \le i \le n$) are random observations. Writing

$$z_i = \{y_i - \phi\, y_{i-1} - \mu - \delta(x_i - \phi\, x_{i-1})\}/\sigma, \quad 1 \le i \le n, \tag{5.3.2}$$

the likelihood function (conditional to y_0) is

$$L \propto \left(\frac{1}{\sigma}\right)^n e^{-\sum_{i=1}^{n} z_i} \prod_{i=1}^{n} z_i^{k-1}. \tag{5.3.3}$$

Let $z_{(i)} = \{w_{(i)} - \mu\}/\sigma$, where $w_{(i)}$ are the order statistics (for given δ and ϕ) of

$$w_i = y_i - \phi y_{i-1} - \delta(x_i - \phi x_{i-1}), \quad 1 \le i \le n. \tag{5.3.4}$$

In fact, $$w_{(i)} = y_{[i]} - \phi y_{[i]-1} - \delta(x_{[i]} - \phi x_{[i]-1}), \quad 1 \le i \le n. \tag{5.3.5}$$

The quadruple ($y_{[i]}$, $y_{[i]-1}$; $x_{[i]}$, $x_{[i]-1}$) may be called concomitant of $w_{(i)}$ and is that observation (y_i, y_{i-1}; x_i, x_{i-1}) which determines $w_{(i)}$. Since complete sums are invariant to ordering, the likelihood equations can be written in terms of $z_{(i)}$ as follows:

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma} - \frac{k-1}{\sigma} \sum_{i=1}^{n} z_{(i)}^{-1} = 0, \tag{5.3.6}$$

$$\frac{\partial \ln L}{\partial \phi} = \frac{1}{\sigma} \sum_{i=1}^{n} (y_{[i]-1} - \delta x_{[i]-1}) - \frac{k-1}{\sigma} \sum_{i=1}^{n} (y_{[i]-1} - \delta x_{[i]-1}) z_{(i)}^{-1} = 0, \tag{5.3.7}$$

$$\frac{\partial \ln L}{\partial \delta} = \frac{1}{\sigma} \sum_{i=1}^{n} (x_{[i]} - \phi x_{[i]-1}) - \frac{k-1}{\sigma} \sum_{i=1}^{n} (x_{[i]} - \phi x_{[i]-1}) z_{(i)}^{-1} = 0 \tag{5.3.8}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} z_{(i)} - \frac{k-1}{\sigma} \sum_{i=1}^{n} z_{(i)} z_{(i)}^{-1} = 0. \tag{5.3.9}$$

Solving these equations by iteration is very problematic realizing that (5.2.2) is a nonlinear model. There are other difficulties too, e.g., if $z_{(1)}$ converges to zero which is more likely for large n, the first three equations are not defined, and the simplified equation (5.3.9), i.e.,

$$\partial \ln L/\partial \sigma = -(nk/\sigma) + (1/\sigma) \sum_{i=1}^{n} z_{(i)} = 0 \tag{5.3.10}$$

may give negative estimates of σ (Akkaya and Tiku, 2001 b). We now obtain modified likelihood equations which have no such difficulties.

5.4 MODIFIED LIKELIHOOD

Write $t_{(i)} = E\{z_{(i)}\}$, $1 \le i \le n$. The values of $t_{(i)}$ are available for $n \le 20$ (Biometrika Tables Vol II, Table 20). For $n \ge 10$, their approximate values obtained from the following equations are used,

$$\frac{1}{\Gamma(k)} \int_0^{t_{(i)}} e^{-z} z^{k-1} dz = \frac{i}{n+1}, \; 1 \le i \le n. \tag{5.4.1}$$

The use of the approximate values in place of the exact values has essentially no effect on the efficiencies of the MML estimators, particularly for $n \ge 20$. To linearize $z_{(i)}^{-1}$, we have from the first two terms of a Taylor series expansion (Akkaya and Tiku, 2001a),

$$z_{(i)}^{-1} \cong \alpha_i - \beta_i z_{(i)}; \; \alpha_i = 2/t_{(i)}, \quad \text{and} \quad \beta_i = 1/t_{(i)}^2. \tag{5.4.2}$$

It may be noted that $\beta_i (1 \le i \le n)$ is a decreasing sequence of positive numbers. Substituting (5.4.2) in (5.3.6)-(5.3.9), we obtain the modified likelihood equations:

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma} - \frac{k-1}{\sigma} \sum_{i=1}^{n} (\alpha_i - \beta_i z_{(i)}) = 0, \tag{5.4.3}$$

$$\frac{\partial \ln L}{\partial \phi} \cong \frac{\partial \ln L^*}{\partial \phi} = \frac{1}{\sigma} \sum_{i=1}^{n} (y_{[i]-1} - \delta x_{[i]-1}) - \frac{k-1}{\sigma} \sum_{i=1}^{n} (y_{[i]-1} - \delta x_{[i]-1}) (\alpha_i - \beta_i z_{(i)}) = 0, \tag{5.4.4}$$

$$\frac{\partial \ln L}{\partial \delta} \cong \frac{\partial \ln L^*}{\partial \delta} = \frac{1}{\sigma} \sum_{i=1}^{n} (x_{[i]} - \phi x_{[i]-1}) - \frac{k-1}{\sigma} \sum_{i=1}^{n} (x_{[i]} - \phi x_{[i]-1}) (\alpha_i - \beta_i z_{(i)}) = 0 \tag{5.4.5}$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \, l \ln L^*}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^{n} z_{(i)} - \frac{k-1}{\sigma} \sum_{i=1}^{n} z_{(i)} (\alpha_i - \beta_i z_{(i)}) = 0. \tag{5.4.6}$$

The solutions of these equations are the MML estimators and have explicit algebraic forms:

$$\hat{\mu} = \bar{v}_{[.]} - \hat{\delta}\bar{u}_{[.]} - (\Delta/m)\hat{\sigma}, \quad \hat{\delta} = G - H\hat{\sigma}, \tag{5.4.7}$$

$$\hat{\phi} = \left[\sum_{i=1}^{n} \beta_i(y_{[i]-1} - \hat{\delta}x_{[i]-1})(y_{[i]} - \hat{\delta}x_{[i]}) - (1/m)\left\{\sum_{i=1}^{n}\beta_i(y_{[i]-1} - \hat{\delta}x_{[i]-1})\right\}\right.$$
$$\left. \times \left\{\sum_{i=1}^{n}\beta_i(y_{[i]} - \hat{\delta}x_{[i]})\right\} - \hat{\sigma}\sum_{i=1}^{n}\{(\Delta_i - (\Delta/m)\beta_i\}(y_{[i]-1} - \hat{\delta}x_{[i]-1})\right]$$
$$\div \left[\sum_{i=1}^{n}\beta_i(y_{[i]-1} - \hat{\delta}x_{[i]-1})^2 - (1/m)\{\sum_{i=1}^{n}\beta_i(y_{[i]-1} - \hat{\delta}x_{[i]-1})\}^2\right] \tag{5.4.8}$$

and

$$\hat{\sigma} = \{-B + \sqrt{(B^2 + 4nC)}\}/2\sqrt{\{n(n-3)\}}; \tag{5.4.9}$$

$$m = \sum_{i=1}^{n}\beta_i; \; \Delta_i = \alpha_i - (k-1)^{-1}, \; \Delta = \sum_{i=1}^{n}\Delta_i; \; \bar{v}_{[.]} = (1/m)\sum_{i=1}^{n}\beta_i v_{[i]},$$

$$v_{[i]} = y_{[i]} - \hat{\phi}y_{[i]-1}; \; \bar{u}_{[i]} = (1/m)\sum_{i=1}^{n}\beta_i u_{[i]}, \; u_{[i]} = x_{[i]} - \hat{\phi}x_{[i]-1},$$

$$G = \sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})v_{[i]} \Big/ \sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})^2,$$

$$H = \sum_{i=1}^{n}\Delta_i(u_{[i]} - \bar{u}_{[.]}) \Big/ \sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})^2,$$

$$B = (k-1)\sum_{i=1}^{n}\Delta_i\{v_{[i]} - \bar{v}_{[.]} - G(u_{[i]} - \bar{u}_{[.]})\} \tag{5.4.10}$$

and

$$C = (k-1)\sum_{i=1}^{n}\beta_i\{v_{[i]} - \bar{v}_{[.]} - G(u_{[i]} - \bar{u}_{[.]})\}^2$$
$$= (k-1)\left\{\sum_{i=1}^{n}\beta_i(v_{[i]} - \bar{v}_{[.]})^2 - G\sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})v_{[i]}\right\}.$$

Realize that $\hat{\delta}$, $\hat{\phi}$ and $\hat{\sigma}$ do not involve $\hat{\mu}$. The resemblance of the expressions above with those in (3.4.8)–(3.4.11) may be noted.

Remark: Since β_i ($1 \le i \le n$) are all positive, the MML estimator $\hat{\sigma}$ is always real and positive. The ML estimator of σ does not necessarily have this property and can assume negative values if, for example, some of the initial observations y_t in (5.2.2) are outliers (Akkaya and Tiku, 2001b).

Computation: Since $\sigma > 0$ and μ is a constant, the ordered variates $z_{(i)}$ are determined by $w_{(i)}$. The latter are obtained by ordering $w_i = y_i - \phi y_{i-1} - \delta(x_i - \phi x_{i-1})$, $1 \le i \le n$. To initialize ordering of w_i, we ignore the constraint $d = -\delta\phi$ (Durbin, 1960; Akkaya and Tiku, 2001a) and calculate the initial estimates from the equations (obtained by minimizing $\Sigma_i w_i^2$):

$$\begin{pmatrix}\hat{\phi}_0\\ \hat{\delta}_0\\ \hat{d}_0\end{pmatrix} = \begin{pmatrix}\Sigma y_{i-1}^2 & \Sigma y_{i-1}x_i & \Sigma y_{i-1}x_{i-1}\\ \Sigma y_{i-1}x_i & \Sigma x_i^2 & \Sigma x_{i-1}x_i\\ \Sigma y_{i-1}x_{i-1} & \Sigma x_i x_{i-1} & \Sigma x_{i-1}^2\end{pmatrix}^{-1}\begin{pmatrix}\Sigma y_{i-1}y_i\\ \Sigma x_i y_i\\ \Sigma x_{i-1}y_i\end{pmatrix}, \tag{5.4.11}$$

each sum carried over i = 1, 2,. . ., n. Initially, therefore,

$$w_{(i)} = y_{[i]} - \hat{\phi}_0 y_{[i]-1} - \hat{\delta}_0 x_{[i]} - \hat{d}_0 x_{[i]-1}, \quad 1 \le i \le n. \tag{5.4.12}$$

Using the concominants ($y_{[i]}$, $y_{[i]-1}$; $x_{[i]}$, $x_{[i]-1}$) determined by (5.4.12), the MML estimates $\hat{\delta}$ and $\hat{\sigma}$ are calculated from (5.4.7) and (5.4.9) with $\hat{\phi} = \hat{\phi}_0$. The MML estimate $\hat{\phi}$ is then calculated from (5.4.8). A second iteration is carried out with $\hat{\phi}_0$, $\hat{\delta}_0$ and $\hat{d}_0$ replaced by $\hat{\phi}$, $\hat{\delta}$ and $-\hat{\phi}\hat{\delta}$, respectively. The third iteration gives the final estimates $\hat{\phi}$, $\hat{\delta}$ and $\hat{\sigma}$. In all our computations, not more than three iterations were needed for the estimates to stabilize sufficiently enough (Akkaya and Tiku, 2001a). The estimate $\hat{\mu}$ is then computed from (5.4.7).

5.5 ASYMPTOTIC COVARIANCE MATRIX

Since the MML estimators are asymptotically equivalent to the ML estimators (Chapter 2), the asymptotic variance-covariance matrix of $\hat{\mu}$, $\hat{\delta}$ and $\hat{\sigma}$ (for given ϕ) is equal to $I^{-1}(\mu, \delta, \sigma)$, I being the Fisher information matrix. The elements of I, i.e. the values of $-E(\partial^2 \ln L/\partial\mu^2)$, $-E(\partial^2 \ln L/\partial\mu\partial\delta)$, $-E(\partial^2 \ln L/\partial\delta^2)$ etc., are not difficult to work out (Chapter 3). Thus, we have the following asymptotic variances (k > 2):

$$V(\hat{\mu}) \cong \frac{(k-2)\sigma^2}{n}\left[\frac{k}{2} + \frac{n\bar{u}^2}{\Sigma(u_i - \bar{u})^2}\right], \tag{5.5.1}$$

$$V(\hat{\delta}) \cong \frac{(k-2)\sigma^2}{\Sigma(u_i - \bar{u})^2} \quad \text{and} \quad V(\hat{\sigma}) \cong \frac{\sigma^2}{2(n-3)}. \tag{5.5.2}$$

Since $\hat{\phi}$ converges to ϕ quickly as n becomes large, u_i is replaced by $\hat{u}_i = x_i - \hat{\phi}x_{i-1}$, $1 \le i \le n$.

It is interesting to note that $V(\hat{\delta})/\sigma^2$ is free of μ, and $V(\hat{\sigma})/\sigma^2$ is free of μ, δ and ϕ. The equations (5.5.1)-(5.5.2) give close approximations to the true variances for large n. The simulated values of the variances (based on N = [100000/n] Monte Carlo runs) are given in Table 5.1 and compared with the theoretical values. The latter are obtained by averaging the values of $V(\hat{\delta})/\sigma^2$, calculated from (5.5.2) for the N random samples. There is close agreement between the two for large n.

The values given in Table 5.1 are slightly different than those of Akkaya and Tiku (2001a). Here, U_i and a_i $(1 \le i \le n)$ are generated simultaneously from the Uniform (0, 1) and gamma $G(a, \sigma)$ distributions, respectively. Then, the design points are obtained from the equation

Table 5.1: Values of 10^2 (Variance)/σ^2 ; k = 3.

	n = 200				n = 300			
	ϕ = 0.5		ϕ = 0.9		ϕ = 0.5		ϕ = 0.9	
	δ	σ	δ	σ	δ	σ	δ	σ
Asymp.	0.302 (0.300)	0.254	0.0524 (0.0524)	0.254	0.200 (0.201)	0.168	0.0350 (0.0351)	0.168
Simul.	0.367	0.288	0.0590	0.250	0.231	0.201	0.0440	0.174

$$x_i = \sqrt{12}\,(U_i - 0.5)/\sqrt{1-\phi^2}\,,\ 1 \le i < n. \tag{5.5.3}$$

The divisor $\sqrt{1-\phi^2}$ gives a wider spread to the design points realizing that the variance $V(y_i) = \sigma^2/(1-\phi^2)$. The values in brackets are obtained by replacing u_i by $\hat{u}_i$ $(1 \le i \le n)$.

The design point x_0 is generated as x_i in (5.5.3) from an independent observation U_0, and $y_0 = a_0/\sqrt{1-\phi^2}$ where a_0 is an independent innovation and has the same distribution as that of $a_i (1 \le i \le n)$.

Akkaya and Tiku (2001 a) generate the design points x_i only once to be common to all the N number of y-samples. We generate x_i and y_i $(1 \le i \le n)$ simultaneously but regard x_i as nonstochastic values. This procedure is perhaps more realistic in practice particularly in business and economics. It must be said, however, that the two procedures yield results not much different from one another.

Since the simulated values are close to the asymptotic values, it follows that the MML estimators are highly efficient for large n. These results are true for other designs, e.g., x_i generated from normal N(0, 1); see Table 5.5, for example.

5.6 LEAST SQUARES

Since the ML estimators are intractable in the context of autoregression, the LS estimators have been used very extensively. They are obtained by minimizing $\Sigma\, a_t^2$ and, if necessary, corrected for bias. The LS estimators are computed by iteration exactly the same way as the MML estimators. In fact, the LS estimators can be obtained from (5.4.7)-(5.4.10) simply by equating Δ_i and β_i $(1 \le i \le n)$ to zero and 1, respectively, and replacing the multiplier k – 1 by 1 in the expression for B and C. Denote the LS estimators by $\tilde{\mu}$, $\tilde{\delta}$, $\tilde{\phi}$ and $\tilde{\sigma}$; $\tilde{\delta}$ and $\tilde{\phi}$ are exactly the same as $\hat{\delta}$ and $\hat{\phi}$ with $\Delta_i = 0$ and $\beta_i = 1$ $(1 \le i \le n)$. While the MML estimators $\hat{\mu}$ and $\hat{\sigma}$ are self bias-correcting, the LS estimators $\tilde{\mu}$ and $\tilde{\sigma}$ need to be corrected for bias. The bias-corrected LS estimators of μ and σ are, as in (3.2.9)-(3.2.12),

$$\tilde{\mu} = \bar{v} - \tilde{\delta}\bar{u} - k\tilde{\sigma} \quad \text{and} \tag{5.6.1}$$

and

$$\tilde{\sigma} = \sqrt{\sum_{i=1}^{n} \{v_i - \bar{v} - \tilde{\delta}(u_i - \bar{u})\}^2} \bigg/ \sqrt{\{k(n-3)\}}\,; \tag{5.6.2}$$

$$v_i = y_i - \tilde{\phi}\, y_{i-1}, \quad u_i = x_i - \tilde{\phi}\, x_{i-1}\ (1 \le i \le n),$$

$$\bar{v} = (1/n)\sum_{i=1}^{n} v_i \quad \text{and} \quad \bar{u} = (1/n)\sum_{i=1}^{n} v_i\,.$$

It is, however, very difficult to work out the expected values and the variances and covariances of the LS estimators. Their means and variances have to be obtained by simulation.

Relative efficiency: The LS estimators have very low efficiencies as compared to the MML estimators. Also, the relative efficiency

$$\text{RE} = 100\ (\text{Variance of the MMLE /Variance of the LSE}) \tag{5.6.3}$$

of the LS estimators decreases with increasing n. To illustrate this, we give in Table 5.2 the simulated values of the means and variances. We give values only for $\phi = 0.5$ and $k = 2$. The relative efficiencies are more or less the same for other values of ϕ. The mean of $\tilde{\sigma}$ is almost the same as that of $\hat{\sigma}$ and is not, therefore, reported. For n = 50, we have the following values of the mean and variance of the MMLE, and the mean and RE of the LSE, for $\phi = 0.9$; $k = 2$, $\delta = 1$ and $\sigma = 1$ (without loss of generality):

	δ		ϕ		σ	
	MML	LS	MML	LS	MML	LS
Mean	1.00	1.00	0.88	0.87	0.97	0.99
n(Variance)	0.090	42	0.059	38	0.688	56

Table 5.2: Values of (1) Mean, (2) n(Variance), and the relative efficiency of the LS estimators; k = 2, $\delta = 1$, $\phi = 0.5$, $\sigma = 1$.

	δ		ϕ		σ	δ		ϕ		σ
	MML	LS	MML	LS	MML	MML	LS	MML	LS	MML
		n = 30					n = 50			
(1)	1.00	1.00	0.47	0.43	0.96	1.00	1.00	0.49	0.46	0.97
(2)	0.58	1.28	0.38	0.77	0.70	0.47	1.25	0.31	0.78	0.61
RE		45		49		58		38	40	52
		n = 100					n = 200			
(1)	1.00	1.00	0.50	0.48	0.98	1.00	1.00	0.50	0.49	0.99
(2)	0.37	1.23	0.23	0.74	0.62	0.29	1.29	0.18	0.89	0.59
RE		30		31	51		22		20	41

It can be seen that the use of the LS estimators results in substantial loss of efficiency. We show in Chapter 8 that, unlike the MML estimators, the LS estimators are not robust to deviations from the true value of the shape parameter k and other data anomalies.

The parameter μ in the model (5.2.2) is of much less importance, but the MML estimator $\hat{\mu}$ is substantially more efficient than the LS estimator $\tilde{\mu}$ and its bias is more or less of the same magnitude. The efficiency of the LSE $\tilde{\mu}$ as compared to the MMLE $\hat{\mu}$ decreases as n increases.

5.7 HYPOTHESIS TESTING FOR GAMMA

Testing the null hypothesis $H_0 : \delta = 0$ is of primary importance in the context of autoregression. To that end, we have the following lemma.

Lemma 5.1: The conditional (ϕ and σ known) distribution of $\hat{\delta}$ is asymptotically normal with mean δ and variance (k > 2)

$$\sigma^2 \Big/ \left\{(k-1)\sum_{i=1}^{n}\beta_i\,(u_{[i]}-\bar{u}_{[.]})^2\right\};\ u_i = x_{[i]} - \phi x_{[i]-1},\ \bar{u}_{[.]} = (1/m)\sum_{i=1}^{n}\beta_i u_{[i]}\,. \qquad (5.7.1)$$

Proof: In view of $\partial \ln L^*/\partial\mu = 0$, the modified likelihood equation when re-organized assumes the form

$$\frac{\partial \ln L}{\partial \delta} \cong \frac{\partial \ln L^*}{\partial \delta} = \frac{(k-1)}{\sigma^2}\left(\sum_{i=1}^{n}\beta_i(u_{[i]}-\bar{u}_{[.]})^2\right)\{\hat{\delta}(\phi,\sigma)-\delta\} = 0, \qquad (5.7.2)$$

$\hat{\delta} = G - H\sigma$; the parameter ϕ occurs in (5.7.1). The result then follows from the fact that the modified likelihood equation $\partial \ln L^*/\partial\delta = 0$ is asymptotically equivalent to $\partial \ln L/\partial\delta = 0$, and $E(\partial^r \ln L^*/\partial\delta^r) = 0$ for all $r \geq 3$.

Comment: In practice, ϕ and σ in (5.7.1)–(5.7.2) are not known and are replaced by their MML estimators $\hat{\phi}$ and $\hat{\sigma}$. Since $\hat{\phi}$ and $\hat{\sigma}$ converge to ϕ and σ, respectively, as n tends to infinity, the distribution of $\sqrt{n}\,(\hat{\delta} - \delta)$ is normal for large n.

Testing $H_0 : \delta = 0$. To test the null hypothesis H_0, we define the statistic (Akkaya and Tiku, 2001a)

$$T_1 = \sqrt{\left\{(k-1)\sum_{i=1}^{n}\beta_i\,(u_{[i]} - \bar{u}_{[.]})^2\right\}}\left(\frac{\hat{\delta}}{\hat{\sigma}}\right),\quad u_{[i]} = x_{[i]} - \hat{\phi}x_{[i]-1}. \tag{5.7.3}$$

Large values of T_1 lead to the rejection of H_0 in favour of H_1: $\delta > 0$. The asymptotic null distribution of T_1 is normal N(0, 1). In fact, the normal distribution N(0, 1) provides accurate approximations to the null distribution of T_1, even for small sample size n. For example, we have the following simulated values of the probability $P(T_1 \geq 1.645 \mid H_0)$; k = 3:

$\phi = 0.0$			$\phi = 0.5$			$\phi = 0.9$		
n = 30	50	100	30	50	100	30	50	100
0.053	0.048	0.040	0.050	0.052	0.054	0.050	0.053	0.047

It can be seen that the normal approximation is adequate.

Power function: The asymptotic power function of the T_1 test is

$$P\{Z \geq z_\alpha - |\lambda_1|\},\quad \lambda_1^2 = (\delta/\sigma)^2\,(k-1)\sum_{i=1}^{n}\beta_i\,(u_{[i]} - \bar{u}_{[.]})^2,\quad u_{[i]} = x_{[i]} - \phi x_{[i]-1}; \tag{5.7.4}$$

Z is a normal N(0, 1) variate and z_α is its $100(1-\alpha)\%$ point, and λ_1^2 is the noncentrality parameter. It may be noted that $\bar{u}_{[.]} - \bar{u} \cong 0$ for large n, and (Appendix 2A)

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})^2 = \frac{1}{n}\,E\left(\frac{1}{z^2}\right)\sum_{i=1}^{n}(u_i - \bar{u})^2$$

$$= \frac{1}{n(k-1)(k-2)}\sum_{i=1}^{n}(u_i - \bar{u})^2\ (k > 2) \tag{5.7.5}$$

so that asymptotically $\lambda_1^2 \cong (k-2)^{-1}\,(\delta/\sigma)^2\sum_{i=1}^{n}(u_i - \bar{u})^2$.

The corresponding test statistic based on the LS estimators is

$$G_1 = \sqrt{\sum_{i=1}^{n}(u_i - \bar{u})^2}\ (\hat{\delta}/\sqrt{k}\hat{\sigma}). \tag{5.7.6}$$

The asymptotic null distribution of G_1 is N(0, 1). Its asymptotic power function is

$$P\{Z \geq z_\alpha - |\lambda_o|\},\ \lambda_o^2 = k^{-1}\,(\delta/\sigma)^2\sum_{i=1}^{n}(u_i - \bar{u})^2. \tag{5.7.7}$$

Since $\lambda_1^2/\lambda_o^2 = k/(k-2)$ is greater than 1, the T_1 test is asymptotically more powerful than the G_1 test.

Comment: For the values given in Table 5.2, the design points x_i are generated as in (5.5.4). The relative efficiencies of the LS estimators are essentially the same if the design points x_0 and x_i $(1 \le i \le n)$ are generated from a normal distribution N(0, 1). As expected, only the variances of the MML and LS estimators $\hat{\delta}$ and $\tilde{\delta}$ are affected substantially by a change in the design; see, for example, Table 5.5.

5.8 SHORT-TAILED SYMMETRIC DISTRIBUTIONS

Assume that the distribution of the innovations a_t is STS (short-tailed symmetric)

$$f(a) \propto \left(\frac{1}{\sigma}\right)\left\{1+\frac{\lambda}{2r}\left(\frac{a}{\sigma}\right)^2\right\}^r e^{-a^2/2\sigma^2}, \quad -\infty < a < \infty; \tag{5.8.1}$$

$\lambda = r/(r-d)$, $d < r$. Here, the likelihood equations $\partial \ln L/\partial\mu = 0$, $\partial \ln L/\partial\delta = 0$, etc., are expressions in terms of

$$g(z_i) = z_i/\{1 + (\lambda/2r)z_i^2\},\ z_i = \{y_i - \phi y_{i-1} - \mu - \delta(x_i - \phi x_i)\}/\sigma. \tag{5.8.2}$$

Solving these equations is almost impossible (Akkaya and Tiku, 2002b).

Modified likelihood: We express the likelihood equations in terms of the ordered variates

$$z_{(i)} = (w_{(i)} - \mu)/\sigma, \quad w_{(i)} = y_{[i]} - \phi y_{[i]-1} - \delta(x_{[i]} - \phi x_{[i]-1}),$$

as in (5.3.6)-(5.3.9). We then linearize (as in Sections 3.6-3.7) the function

$$g(z_{(i)}) = z_{(i)}/\{1 + (\lambda/2r)z_{(i)}^2\}, \quad 1 \le i \le n. \tag{5.8.3}$$

For $\lambda \le 1$, we use the linear functional (3.6.12); for $\lambda > 1$, we use (3.7.4). This ensures that the MML estimator of σ is always real and positive. The resulting MML estimators are

$$\hat{\mu} = \bar{v}_{[.]} - \hat{\delta}\bar{u}_{[.]}, \quad \hat{\delta} = G - \lambda H\hat{\sigma}, \quad \hat{\phi} = K - \lambda D\hat{\sigma} \tag{5.8.4}$$

and

$$\hat{\sigma} = \{-\lambda B + \sqrt{(\lambda B)^2 + 4nC}\}/2\sqrt{n(n-3)}; \tag{5.8.5}$$

$$v_{[i]} = y_{[i]} - \hat{\phi}y_{[i]-1}, \quad u_{[i]} = x_{[i]} - \hat{\phi}x_{[i]-1}, \quad \bar{v}_{[.]} = (1/m)\sum_{i=1}^{n}\beta_i v_{[i]},$$

$$\bar{u}_{[.]} = (1/m)\sum_{i=1}^{n}\beta_i u_{[i]}, \quad m = \sum_{i=1}^{n}\beta_i,$$

$$G = \frac{\sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})v_{[i]}}{\sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})^2}, \quad H = \frac{\sum_{i=1}^{n}\alpha_i(u_{[i]} - \bar{u}_{[.]})}{\sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})^2}, \tag{5.8.6}$$

$$K = \frac{1}{\Delta}\left[\sum_{i=1}^{n}\beta_i(y_{[i]} - \hat{\delta}x_{[i]})(y_{[i]-1} - \hat{\delta}x_{[i]-1}) - \frac{1}{m}\left\{\sum_{i=1}^{n}\beta_i(y_{[i]} - \hat{\delta}x_{[i]})\right\}\left\{\sum_{i=1}^{n}\beta_i(y_{[i]-1} - \hat{\delta}x_{[i]-1}\right\}\right],$$

$$D = \frac{1}{\Delta}\sum_{i=1}^{n}\alpha_i(y_{[i]-1} - \hat{\delta}x_{[i]-1}),$$

$$\Delta = \sum_{i=1}^{n}\beta_i(y_{[i]-1} - \hat{\delta}x_{[i]-1})^2 - \frac{1}{m}\left\{\sum_{i=1}^{n}\beta_i(y_{[i]-1} - \hat{\delta}x_{[i]-1}\right\}^2; \tag{5.8.7}$$

$$B = \sum_{i=1}^{n} \alpha_i \{v_{[i]} - \bar{v}_{[.]} - G(u_{[i]} - \bar{u}_{[.]})\} \quad \text{and}$$

$$C = \sum_{i=1}^{n} \beta_i \{v_{[i]} - \bar{v}_{[.]} - G(u_{[i]} - \bar{u}_{[.]})\}^2 \tag{5.8.8}$$

$$= \sum_{i=1}^{n} \beta_i \{v_{[i]} - \bar{v}_{[.]})^2 - G \sum_{i=1}^{n} \beta_i (u_{[i]} - \bar{u}_{[.]}) v_{[i]}.$$

It is interesting to note that the expressions (5.8.4)-(5.8.8) are exactly similar to (5.4.7)-(5.4.10) in spite of the fact that the distributions (5.3.1) and (5.8.1) are entirely different from one another.

Least squares: The LS estimators $\tilde{\mu}$, $\tilde{\delta}$ and $\tilde{\phi}$ are exactly the same as $\hat{\mu}$, $\hat{\delta}$ and $\hat{\phi}$ above with α_i and β_i $(1 \le i \le n)$ equated to 0 and 1, respectively. The bias corrected LS estimator of σ is

$$\tilde{\sigma} = \sqrt{\sum_{i=1}^{n} \{(v_i - \bar{v}) - \tilde{\delta}(u_i - \bar{u})\}^2} \Big/ \sqrt{(n-3)\mu_2} \tag{5.8.9}$$

where $v_i = y_i - \tilde{\phi} y_{i-1}$, $u_i = x_i - \tilde{\phi} x_{i-1}$, $\bar{v} = (1/n)\sum_{i=1}^{n} v_i$ and $\bar{u} = (1/n)\sum_{i=1}^{n} u_i$;

μ_2 is the variance of the distribution (5.8.1). Realize that $s_e = \tilde{\sigma}\sqrt{\mu_2}$ is exactly the same as $\hat{\sigma}$ with α_i and β_i equated to 0 and 1, respectively, and is the LSE of the population standard deviation.

Remark: The MML estimators (5.8.4)-(5.8.8) are computed exactly the same way as those in Section 5.4. The LS estimators are computed in a similar fashion.

5.9 ASYMPTOTIC COVARIANCE MATRIX

The asymptotic variance-covariance matrix of $\hat{\mu}$, $\hat{\delta}$ and $\hat{\sigma}$ (for a given ϕ) is equal to $I^{-1}(\mu, \delta, \sigma)$, I being the Fisher Information matrix. The non-zero elements of I are given below:

$$I_{11} = -E(\partial^2 \ln L/\partial\mu^2) = (n/\sigma^2)Q, \tag{5.9.1}$$

$$I_{12} = -E(\partial^2 \ln L/\partial\mu\partial\delta) = (Q/\sigma^2) \sum_{i=1}^{n} u_i,$$

$$I_{22} = -E(\partial^2 \ln L/\partial\delta^2) = (Q/\sigma^2) \sum_{i=1}^{n} u_i^2 \, (u_i = x_i - \phi x_{i-1}) \tag{5.9.2}$$

and

$$I_{33} = -E(\partial^2 \ln L/\partial\sigma^2) \tag{5.9.3}$$

$$= (n/\sigma^2)[-1 + 3\mu_2 - 3\lambda E(r-1, j+1) + (\lambda^2/r)\, E(r-2, j+2)];$$

$$Q = 1 - \lambda E(r-1, j) + (\lambda^2/r)\, E(r-2, j+1). \tag{5.9.4}$$

The variance μ_2 is given in (3.8.2) and for $r \ge i$,

$$E(r-i, j+k) = \left[\sum_{j=0}^{r-i} \binom{r-i}{j} \left(\frac{\lambda}{2r}\right)^j\right] \frac{\{2(j+k)!\}}{2^{j+k}\,(j+k)!} \Bigg/ \left[\sum_{j=0}^{r} \binom{r}{j} \left(\frac{\lambda}{2r}\right)^j \frac{(2j)!}{2^j\,(j)!}\right]. \tag{5.9.5}$$

The beauty of the expressions (5.9.1) – (5.9.3) may be noted. Realize that for large n, $V(\hat{\delta}) \cong \sigma^2/Q \sum_{i=1}^{n} (u_i - \bar{u})^2$.

It is easy to find the asymptotic variances of the estimators from (5.9.1) – (5.9.5). The simulated variances for n = 100 given in Table 5.3 were found to be only marginally bigger than the asymptotic variances. It can, therefore, be concluded that the MML estimators are highly efficient. We show in Chapter 8 that they are also remarkably robust to STS distributions and to inliers.

Remark: For the STS distributions, the convergence of the variances to their asymptotic values is much faster than for the gamma family (skew distributions). This is partly due to the fact that for symmetric distributions, the bias in $\hat{\mu}$ (and $\tilde{\mu}$) is negligible even for small n.

Efficiency of LS Estimators: We simulated values of the relative efficiencies of the LS estimators defined in (5.6.3), the design points being the same as in (5.5.4). The values are given in Table 5.3 for $\phi = 0.5$. The relative efficiencies for $\phi = 0.0$ and 0.9 are essentially the same as for $\phi = 0.5$ and are not, therefore, reported; E_1, E_2, E_3 and E_4 are the relative efficiencies of the LS estimators $\tilde{\mu}$, $\tilde{\delta}$, $\tilde{\phi}$ and $\tilde{\sigma}$, respectively; $\mu = 0$ and $\delta = 1$:

Table 5.3: Variances of the MML estimators and the relative efficiencies of the LS estimators: (1) (r = 2, d = 1) and (2) (r = 4, d = 2); ϕ =0.5.

	(n/σ^2) Variance				Relative efficiency			
n	$\hat{\mu}$	$\hat{\delta}$	$\hat{\phi}$	$\hat{\sigma}$	E_1	E_2	E_3	E_4
				Distribution (1)				
30	2.69	1.52	0.681	0.238	67	73	74	91
50	2.38	1.02	0.515	0.251	70	63	66	93
100	2.07	0.89	0.486	0.234	70	57	57	94
				Distribution (2)				
30	3.48	1.52	0.505	0.187	57	55	57	87
50	3.21	1.20	0.364	0.171	58	46	44	87
100	2.80	0.87	0.335	0.162	59	40	41	88

It can be seen that the MML estimators are enormously more efficient than the LS estimators.

5.10 HYPOTHESIS TESTING FOR STS DISTRIBUTIONS

To develop a test for H_0: $\delta = 0$, we have the following result.

Lemma 5.2: Conditionally (ϕ and σ known), the asymptotic distribution of $\hat{\delta}(\phi, \sigma)$ is normal with mean δ and variance

$$\sigma^2/Q \sum_{i=1}^{n} (u_i - \bar{u})^2; \qquad (5.10.1)$$

u_i and $\bar{u}$ are exactly the same as in (5.7.5).

Proof: Follows exactly along the same lines as Lemma 5.1 and the fact that

$$\lim_{n\to\infty} (1/n) \sum_{i=1}^{n} \beta_i (u_{[i]} - \bar{u}_{[.]})^2 = (1/n) \sum_{i=1}^{n} (u_i - \bar{u})^2 \; E\left[1 - \frac{\lambda - (\lambda^2/2r)z^2}{\{1 + (\lambda/2r)z^2\}^2}\right] \qquad (5.10.2)$$

$$= (1/n)Q\sum_{i=1}^{n}(u-\bar{u})^2\,; \tag{5.10.3}$$

Q is given in (5.9.4) and $\bar{u} = (1/n)\sum_{i=1}^{n} u_i$.

To test H_0, we define the statistic ($u_i = x_i - \hat{\phi}x_{i-1}$)

$$T_2 = \sqrt{Q\sum_{i=1}^{n}(u_i-\bar{u})^2}\left(\frac{\hat{\delta}}{\hat{\sigma}}\right) \tag{5.10.4}$$

as in Akkaya and Tiku (2001b). Large values of T_2 lead to the rejection of H_0 in favour of H_1: $\delta > 0$. The null distribution of T_2 is asymptotically normal N(0, 1). It may be noted that the variances of $\hat{\phi}$ and $\hat{\sigma}$ are both considerably smaller than that of $\hat{\delta}$ (Table 5.3). That speeds up the convergence of the distribution of T_2 to normality.

The test based on the LS estimators is ($u_i = x_i - \tilde{\phi}\,x_{i-1}$)

$$G_2 = \sqrt{\sum_{i=1}^{n}(u_i-\bar{u})^2}\;(\tilde{\delta}/\tilde{\sigma}\sqrt{\mu_2}). \tag{5.10.5}$$

The null distribution of G_2 is asymptotically normal N(0, 1).

The asymptotic power functions of the T_2 and G_2 tests are, respectively,

$$P\{Z \ge z_\alpha - |\lambda_1|\} \quad\text{and}\quad P\{Z \ge z_\alpha - |\lambda_o|\}\,; \tag{5.10.6}$$

$$\lambda_1^2 = \left[Q\sum_{i=1}^{n}(u_i-\bar{u})^2\right](\delta/\sigma)^2 \quad\text{and}\quad \lambda_0^2 = \left[\sum_{i=1}^{n}(u_i-\bar{u})^2\right](\delta/\sigma\sqrt{\mu_2})^2 \tag{5.10.7}$$

are the noncentrality parameters and $u_i = x_i - \phi x_{i-1}$. The ratio

$$\lambda_1^2/\lambda_0^2 = Q\mu_2 \tag{5.10.8}$$

is always greater than 1. For example, we have the following values:

	(r = 2, d = 0)	(r = 2, d = 1)	(r = 4, d = 0)	(r = 4, d = 2)
λ_1^2/λ_0^2:	1.12	1.80	1.17	2.79

The T_2 test is, therefore, asymptotically more powerful than the G_2 test.

The null distributions of T_2 and G_2 are closely approximated by normal N(0, 1) for all $n \ge 60$. The T_2 test is, however, considerably more powerful than the G_2 test. For n = 70 and n = 100, for example, we have the simulated values of the type I error and power given in Table 5.4.

Comment: For $n \le 60$, the normal N(0, 1) gives values of the type I errors

$$P(T_2 \ge z_\alpha \mid H_0) \quad\text{and}\quad P(G_2 \ge z_\alpha \mid H_0) \tag{5.10.9}$$

quite a bit larger than its presumed value. But, that is due to the fact that the asymptotic variances used in (5.10.4) and (5.10.5) are a little too small than their exact values. Unfortunately, it is very difficult to work out the exact variances of both the T_2 and the G_2 test statistics.

Table 5.4: Power of the T and G tests; $\mu = 0$, $\phi = 0.5$.

Short-tailed symmetric distributions		$\delta =$	0.0	0.1	0.2	0.3	0.4	0.5
(r = 2, d = 0)	n = 70	T	0.059	0.21	0.54	0.87	0.97	1.00
		G	0.064	0.20	0.49	0.83	0.96	1.00
	n = 100	T	0.049	0.27	0.60	0.86	0.98	1.00
		G	0.048	0.24	0.54	0.83	0.97	1.00
(r = 2, d = 1)	n = 70	T	0.055	0.25	0.62	0.71	0.97	1.00
		G	0.058	0.17	0.42	0.53	0.88	0.95
	n = 100	T	0.054	0.28	0.77	0.90	0.98	1.00
		G	0.051	0.20	0.52	0.73	0.89	1.00

For $n \le 60$, therefore, the simulated variances may be used. Using the variances given in Table 5.3, we have $T_2 = \hat{\delta}/(0.1425)$ and $G_2 = \tilde{\delta}/(0.1797)$ for n = 50 and $\phi = 0.5$ (r = 2 and d = 1). The values of the type 1 error and power of the two tests are given below:

Simulated power values (r = 2, d = 1): $\phi = 0.5$, n = 50.

Statistic	$\delta =$	0.0	0.1	0.2	0.3	0.4	0.5
T		0.055	0.20	0.41	0.67	0.86	0.95
G		0.049	0.16	0.30	0.52	0.70	0.85

The normal N(0, 1) provides accurate approximations for the percentage points of the variance-adjusted T_2 and G_2 statistics. The T_2 test is, however, considerably more powerful. We show in Chapter 8 that the T_2 test is also remarkably robust. This is due to the inverted umbrella ordering of the coefficients β_i as explained in Chapter 3.

5.11 LONG-TAILED SYMMETRIC DISTRIBUTIONS

Assume that the innovations a_t in the model (5.2.2) are iid and have a distribution in the long-tailed symmetric (LTS) family

$$f(a) \propto \frac{1}{\sigma}\left\{1 + \frac{a^2}{k\sigma^2}\right\}^{-p}, \quad -\infty < a < \infty; \tag{5.11.1}$$

$k = 2p - 3$, $p \ge 2$. Here, the likelihood equations are

$$\begin{aligned} &\partial \ln L/\partial\mu = (2p/k\sigma)\, \Sigma_i g(z_i) = 0 \\ &\partial \ln L/\partial\delta = (2p/k\sigma)\, \Sigma_i (x_i - \phi x_{i-1})\, g(z_i) = 0 \\ &\partial \ln L/\partial\phi = (2p/k\sigma)\, \Sigma_i (y_{i-1} - \delta x_{i-1})\, g(z_i) = 0 \text{ and} \\ &\partial \ln L/\partial\sigma = -(n/\sigma) + (2p/k\sigma)\, \Sigma_i z_i\, g(z_i) = 0; \end{aligned} \tag{5.11.2}$$

$$z_i = (w_i - \mu)/\sigma,\ w_i = y_i - \phi y_{i-1} - \delta(x_i - \phi x_{i-1}),$$
$$g(z) = z/\{1 + (1/k)z^2\}. \tag{5.11.3}$$

It is almost impossible to solve the equations (5.11.2). Hence, the ML estimators of μ, δ, ϕ and σ are not available. The modified likelihood equations are obtained first by expressing (5.11.2) in terms of the ordered variates

$$z_{(i)} = \{w_{(i)} - \mu\}/\sigma, \quad w_{(i)} = y_{[i]} - \phi y_{[i]-1} - \delta(x_{[i]} - \phi x_{[i]-1}), \quad 1 \le i \le n,$$

as before. Then, $g(z_{(i)})$ is replaced by the linear functional

$$g(z_{(i)}) \cong \alpha_i + \beta_i z_{(i)}, \quad 1 \le i \le n; \tag{5.11.4}$$

the coefficients α_1 and β_i are given in (2.3.14).

The solutions of the modified likelihood equations are the MML estimators:

$$\hat{\mu} = \bar{v}_{[.]} - \hat{\delta}\bar{u}_{[.]}, \quad \hat{\delta} = G + H\hat{\sigma}, \quad \hat{\phi} = K + D\hat{\sigma} \text{ and} \tag{5.11.5}$$

$$\hat{\sigma} = \{B + \sqrt{(B^2 + 4nC)}\}/2\sqrt{n(n-3)}; \tag{5.11.6}$$

$v_{[i]}$, $u_{[i]}$, $\bar{v}_{[.]}$, $\bar{u}_{[.]}$, G, H, D and K have exactly the same expressions as those in (5.8.6)-(5.8.8) with

$$B = (2p/k)\sum_{i=1}^{n}\alpha_i \{v_{[i]} - \bar{v}_{[.]} - G(u_{[i]} - \bar{u}_{[.]})\} \text{ and}$$

$$C = (2p/k)\sum_{i=1}^{n}\beta_i \{v_{[i]} - \bar{v}_{[.]} - G(u_{[i]} - \bar{u}_{[.]})\}^2 \tag{5.11.7}$$

$$= (2p/k)\left\{\sum_{i=1}^{n}\beta_i (v_{[i]} - \bar{v}_{[.]})^2 - G\sum_{i=1}^{n}\beta_i (u_{[i]} - \bar{u}_{[.]})v_{[i]}\right\}.$$

The expressions (5.11.5)-(5.11.7) are refined versions of those given in Tiku et al. (1999, pp. 320-321).

Remark: The estimators (5.11.5)-(5.11.6) are computed in three iterations, exactly along the same lines as the MML estimators (5.8.4)-(5.8.5). If for a sample, C in (5.11.7) is negative, the MML estimators are computed with α_i and β_i replaced by $\alpha_i^* = 0$ and $\beta_i^* = 1/\{1 + 1/k)t_{(i)}^2\}$, respectively.

Least squares: The LS estimators are obtained from (5.11.5)-(5.11.7) by equating α_i and β_i to 0 and 1, respectively, and equating 2p/k to 1.

Asymptotic relative eficiency: To evaluate the asymptotic efficiencies of LS estimators relative to the MML estimators for a given ϕ, we have the following results. These results also apply to the MML estimators $\hat{\mu}$, $\hat{\delta}$ and $\hat{\sigma}$ for the STS distributions with $\text{Cov}_\phi(\hat{\mu}, \hat{\delta}, \hat{\sigma})$ evaluated from the corresponding Fisher information matrix.

Theorem 5.1: For a given ϕ, the MML estimators $\hat{\mu}(\phi)$, $\hat{\delta}(\phi)$ and $\hat{\sigma}(\phi)$ are asymptotically unbiased with variance-covariance matrix $\text{Cov}_\phi(\hat{\mu}, \hat{\delta}, \hat{\sigma})$

$$= \frac{(p+1)\sigma^2}{n(p-1/2)}\begin{bmatrix} \frac{(p-3/2)}{p}\left(1 + \frac{\bar{u}^2}{s_u^2}\right) & -\frac{(p-3/2)}{p}\frac{\bar{u}}{s_u^2} & 0 \\ -\frac{(p-3/2)}{p}\frac{\bar{u}}{s_u^2} & \frac{(p-3/2)}{p}\frac{1}{s_u^2} & 0 \\ 0 & 0 & 1/2 \end{bmatrix}; \tag{5.11.8}$$

$$u_i = x_i - \phi x_{i-1}, \quad \bar{u} = \sum_i u_i/n \quad \text{and} \quad s_u^2 = \sum_{i=1}^{n}(u_i - \bar{u})^2/n.$$

Proof: The asymptotic unbiasedness follows from Taylor series expansions. Consider, for example, $\partial \ln L^*/\partial\mu$. For large n (Kendall and Stuart, 1979, p. 52)

$$E(\hat{\mu}) \cong \mu - E\left(\frac{\partial \ln L^*}{\partial \mu}\right) \Big/ E\left(\frac{\partial^2 \ln L^*}{\partial \mu^2}\right).$$

But from the results given in Appendix 2A, $E(\partial \ln L^*/\partial\mu)$ is asymptotically equal to $E(\partial \ln L/\partial\mu)$ and the latter is zero. Therefore, $E(\hat{\mu}) \cong \mu$ for large n. Similarly, $E(\hat{\delta}) \cong \delta$ and $E(\hat{\sigma}) \cong \sigma$ for large n.

The information matrix $I_\phi(\hat{\mu}, \hat{\delta}, \hat{\sigma})$ is given by the expected values of the second derivatives. Realizing that (equation 2A.4)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \beta_i (u_{[i]} - \bar{u}_{[.]})^2 = \frac{(p - 1/2)}{p + 1} s_u^2,$$

(5.11.8) is obtained by taking the inverse of the information matrix.

Lemma 5.3: For a given ϕ, the LS estimators $\tilde{\mu}, \tilde{\delta}$ and $\tilde{\sigma}$ are asymptotically unbiased with variance-covariance matrix

$$\mathrm{Cov}_\phi(\tilde{\mu}, \tilde{\delta}, \tilde{\sigma}) = \frac{\sigma^2}{n}\begin{bmatrix} 1+\frac{\bar{u}^2}{s_u^2} & -\frac{\bar{u}}{s_u^2} & 0 \\ -\frac{\bar{u}}{s_u^2} & \frac{1}{s_u^2} & 0 \\ 0 & 0 & \frac{1}{2}\left(1+\frac{1}{2}\lambda_4\right) \end{bmatrix}; \quad (5.11.9)$$

$\lambda_4 = (\mu_4/\mu_2^2) - 3$ and μ_4/μ_2^2 is the kurtosis of the distribution (5.11.1).

Comment: For a given ϕ, the asymptotic relative efficiencies of the LS estimators $\tilde{\mu}$ and $\tilde{\delta}$ are

$$RE(\tilde{\mu}) = RE(\tilde{\delta}) = 100(p + 1)(p - 1/2)/p(p - 1/2) \quad (5.11.10)$$

which assumes values between 50 and 100 percent as p increases from 2 to ∞. Realizing that for the family (5.11.1), the kurtosis is

$$\mu_4/\mu_2^2 = 3(p - 3/2)/(p - 5/2), \quad (5.11.11)$$

the asymptotic relative efficiency of the LS estimator $\tilde{\sigma}$ is

$$RE(\tilde{\sigma}) = 100(p + 1)/(p - 1/2)\left[1 + \frac{3}{2}\left\{\frac{(p - 3/2)}{(p - 5/2)} - 1\right\}\right] \quad (5.11.12)$$

which assumes values between zero and 100 percent as p increases from 2 to ∞. Clearly, the LS estimators have very low efficiencies unless $p = \infty$ in which case (5.11.1) reduces to normal $N(0, \sigma^2)$ and the MML estimators reduce to the LS estimators.

Small sample efficiency: Tiku et al. (1999) reported the results of an extensive simulation study carried out to determine the relative efficiencies of the LS estimators. They considered three designs: (a) x_i generated from a uniform distribution (– 1, 1), (b) x_i generated from a standard normal distribution (Tan and Lin, 1993), and (c) x_i generated from a Cauchy distribution. They found the MML estimators enormously more efficient than the LS estimators. For illustration, we give the values for designs (a) and (b) in Table 5.5. It can be seen that even for small sample sizes, the bias in both the MML as well as the LS estimators is negligible. It is also interesting to see that the relative efficiencies of the LS estimators are almost the same for the two designs (a) and (b). The values given in Table 5.5 are slightly different than those of Tiku et al. (1999). This is due to the fact that we generate (y_t, x_t) here somewhat differently as said earlier.

Table 5.5: Means and mean square errors of the estimators;
$\mu = 0$, $\delta = 1$, $\phi = 0.5$ and $\sigma = 1$ and $p = 3.5$; $n = 30$.

		Design (a)			Design (b)		
	Estimator	Mean	$(\text{Bias})^2$	MSE	Mean	$(\text{Bias})^2$	MSE
μ	LS	– 0.006	0.000	0.056	0.003	0.000	0.057
	MML	– 0.005	0.000	0.042	0.003	0.000	0.043
δ	LS	0.999	0.000	0.097	0.999	0.000	0.035
	MML	1.000	0.000	0.071	0.999	0.000	0.025
ϕ	LS	0.400	0.010	0.039	0.400	0.010	0.038
	MML	0.429	0.005	0.034	0.427	0.005	0.034
σ	LS	0.919	0.007	0.039	0.919	0.006	0.039
	MML	0.977	0.001	0.028	0.977	0.000	0.028

It may be noted that the design effect is pronounced only in estimating the regression coefficient δ, e.g., the variances of $\hat{\delta}$ and $\tilde{\delta}$ change substantially with the design. The variances of other estimators are unaffected (almost).

5.12 HYPOTHESIS TESTING FOR LTS DISTRIBUTIONS

To test the null hypothesis H_0: $\delta = 0$, we have the following result.

Lemma 5.4: For a given ϕ, the MML estimator is asymptotically the MVB estimator with variance

$$(p + 1)(p - 3/2)\sigma^2/np(p - 1/2)s_u^{\,2},$$

$$s_u^{\,2} = \sum_{i=1}^{n}(u_i - \bar{u})^2/n \, (u_i = x_i - \phi x_{i-1}), \tag{5.12.1}$$

and is normally distributed.

Proof: The modified likelihood equation $\partial \ln L^*/\partial\delta = 0$ assumes the form $(k = 2p - 3,\ p \geq 2)$

$$\frac{\partial \ln L}{\partial \delta} \cong \frac{\partial \ln L^*}{\partial \delta} = \frac{2np}{k\sigma^2}\left[\frac{1}{n}\sum_{i=1}^{n}\beta_i(u_{[i]} - \bar{u}_{[.]})^2\right](\hat{\delta}(\phi) - \delta) = 0, \tag{5.12.2}$$

and the result follows.

The asymptotic value of the constant on the right hand side of (5.12.2) is the reciprocal of that in (5.12.1); see also (5.10.3). Therefore, to test the null hypothesis H_0: $\delta = 0$, the proposed statistic is (Tiku et al., 1999)

$$T_3 = \sqrt{\left\{\frac{np(p - 1/2)\hat{s}_u^{\,2}}{(p + 1)(p - 3/2)}\right\}}\left(\frac{\hat{\delta}}{\hat{\sigma}}\right), \tag{5.12.3}$$

where $\hat{s}_u^{\,2}$ is $s_u^{\,2}$ with u_i replaced by $\hat{u}_i = x_i - \hat{\phi}x_{i-1}$ $(1 \leq i \leq n)$. Large values of T_3 lead to the rejection of H_0 in favour of H_1: $\delta > 0$. Since $\hat{\phi}$ and $\hat{\sigma}$ converge to ϕ and σ, respectively, as n becomes large, the asymptotic null distribution of T_3 is normal N(0, 1). Its power function P(T $\geq z_\alpha | H_1$) is

$$\text{Power} = \text{Prob}\,(Z \geq z_\alpha - |\lambda_3|); \tag{5.12.4}$$

$$\lambda_3^2 = \frac{np(p-1/2)s_u^2}{(p+1)(p-3/2)}\left(\frac{\delta}{\sigma}\right)^2 \tag{5.12.5}$$

is the non-centrality parameter.

The corresponding statistic based on the LS estimators is

$$G_3 = \sqrt{(ns_u^2)}(\tilde{\delta}/\tilde{\sigma}). \tag{5.12.6}$$

The asymptotic null distribution of G_3 is normal N(0, 1). The power function of the G_3 test is

$$\text{Power} = P(Z \geq z_\alpha - |\lambda_0|) \tag{5.12.7}$$

where $\lambda_0^2 = ns_u^2(\delta/\sigma)^2$. The ratio $\lambda_3^2/\lambda_0^2 = p(p-1/2)/(p+1)(p-3/2)$ is greater than 1. The T_3 test is, therefore, asymptotically more powerful than the G_3 test, as for other families of non-normal distributions.

Comment: Simulations reveal that the normal approximations are adequate for the null distributions of T_3 and G_3 for $n \geq 30$. The T_3 test is, however, considerably more powerful than the G_3 test (Tiku et al., 1999). We show in Chapter 8 that the T_3 test is remarkably robust to LTS distributions and to outliers in a sample.

Türker (2002) has extended the estimation and hypothesis testing procedures above to k autoregressive models

$$y_{i,t} - \phi_i y_{i,t-1} = \mu_i + \delta_i(x_{i,t} - \phi_{i,} x_{i\,t-1}) + a_{it} \quad (1 \leq i \leq k,\ 1 \leq t \leq n_i). \tag{5.12.8}$$

For example, she develops a procedure for testing $\delta_1 = \delta_2 = \ldots\ldots = \delta_k$, a_{it} being non-normal iid innovations.

Remark: For $\phi = \delta = 0$, (5.2.2) reduces to a location-scale model considered in Chapter 2. For $\phi = 0$, it reduces to a linear regression model considered in Chapter 3. For $\delta = 0$, (5.2.2) reduces to a time series model. We now briefly discuss such models when the innovations $a_t\,(1 \leq t \leq n)$ have non-normal distributions.

5.13 TIME SERIES MODEL

In the first place, consider a time series AR(1) model

$$y_t = \phi y_{t-1} + a_t \quad (1 \leq t \leq n); \tag{5.13.1}$$

$|\phi| < 1$, and $E(a_t) = 0$ and $V(a_t) = \sigma^2$. We take $y_0 = a_0/\sqrt{(1-\phi^2)}$. Assume that the innovations $a_0, a_1, \ldots\ldots, a_n$ are iid and have the LTS distribution (5.11.1). Realize that $E(y_0) = 0$ and $V(y_t) = \sigma^2/(1-\phi^2)$, $1 \leq t \leq n$. Writing

$$z_t = (y_t - \phi y_{t-1})/\sigma,$$

the likelihood function conditional to y_0 is

$$L \propto \left(\frac{1}{\sigma}\right)^n \prod_{t=1}^{n}\left(1 + \frac{a_t^2}{k\sigma^2}\right)^{-p}. \tag{5.13.2}$$

The likelihood equations are

$$\frac{\partial \ln L}{\partial \phi} = \frac{2p}{k\sigma}\sum_{t=1}^{n} y_{t-1}\, g(z_t) = 0 \text{ and}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma}\sum_{t=1}^{n} z_t\, g(z_t) = 0; \tag{5.13.3}$$

$g(z) = z/(1 + z^2/k)$. It may be noted that y_{t-1} and z_t are independent of each other. Solving (5.13.3) by iteration is problematic. Let $z_{(i)} = a_{(i)}/\sigma$ be the ordered variates, where $a_{(i)} = y_{[i]} - \phi y_{[i]-1}$ (for a given ϕ) are the order statistics of a_i ($1 \le i \le n$). The equations (5.13.3) can be written as

$$\frac{\partial \ln L}{\partial \phi} = \frac{2p}{k\sigma}\sum_{i=1}^{n} y_{[i]-1}\, g(z_{(i)}) = 0 \text{ and}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{2p}{k\sigma}\sum_{i=1}^{n} z_{(i)}\, g(z_{(i)}) = 0. \tag{5.13.4}$$

Replacing $g(z_{(i)})$ by the linear function (2.3.13), we obtain the modified likelihood equations. The solutions of these equations are the MML estimators:

$$\hat{\phi} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \{B + \sqrt{(B^2 + 4nC)}\}/2n; \tag{5.13.5}$$

$$K = \sum_{i=1}^{n}\beta_i y_{[i]} y_{[i]-1} \Big/ \sum_{i=1}^{n}\beta_i y^2{}_{[i]-1}, \quad D = \sum_{i=1}^{n}\alpha_i y_{[i]-1} \Big/ \sum_{i=1}^{n}\beta_i y^2{}_{[i]-1},$$

$$B = \frac{2p}{k}\sum_{i=1}^{n}\alpha_i (y_{[i]} - K y_{[i]-1}) \text{ and} \tag{5.13.6}$$

$$C = \frac{2p}{k}\sum_{i=1}^{n}\beta_i (y_{[i]} - K y_{[i]-1})^2 = \frac{2p}{k}\left(\sum_{i=1}^{n}\beta_i (y_{[i]}{}^2 - K\sum_{i=1}^{n}\beta_i y_{[i]} y_{[i]-1}\right).$$

The divisor n in (5.13.5) may be replaced by $\sqrt{n(n-1)}$ as a bias correction. If for a sample, C in (5.13.6) is negative, the MMLE are computed by replacing α_i and β_i by α_i*and β_i*, respectively, as said earlier. That ensures that $\hat{\sigma}$ is always real and positive. Realize that $\hat{\phi}$ is scale invariant and $\hat{\sigma}$ is location invariant, as they should be.

Computations: The MML estimators (5.13.5) are computed in two iterations. In the first iteration, $\hat{\phi}$ is computed from the order statistics of $a_i = y_i - \tilde{\phi} y_{i-1}$ ($1 \le i \le n$); $\tilde{\phi} = \sum_{i=1}^{n} y_i y_{i-1} \Big/ \sum_{i=1}^{n} y_{i-1}^2$ is the LS estimator. In the second iteration, $\tilde{\phi}$ is replaced by $\hat{\phi}$ and a revised value of $\hat{\phi}$ obtained. The MML estimator $\hat{\sigma}$ is then computed. Not more than two iterations are needed for the estimates to stabilize sufficiently enough (Tiku et al., 2000).

5.14 ASYMPTOTIC PROPERTIES

Realizing that the modified likelihood equations (and the expected values of the first two derivatives of ln L*) are asymptotically equivalent to the corresponding likelihood equations (and the first two derivatives of ln L), the following result is true.

Thorem 5.2: The MML estimators $\hat{\phi}$ and $\hat{\sigma}$ are asymptotically unbiased with variance-covariance matrix ($p \ge 2$)

$$\text{Cov}\,(\hat{\phi}, \hat{\sigma}) \cong \begin{bmatrix} \dfrac{(p+1)(p-3/2)}{p(p-1/2)}\dfrac{(1-\phi^2)}{n} & 0 \\ 0 & \dfrac{p+1}{(p-1/2)}\dfrac{\sigma^2}{2n} \end{bmatrix}. \tag{5.14.1}$$

Proof: Unbiasedness follows from Taylor series expansions as in Theorem 5.1. Now, the expected values of the second derivatives of ln L* are exactly the same asymptotically as those of ln L (Appendix 2A). Using the integral (2.3.20), the values of the Fisher information $I(\phi, \sigma)$ are obtained; I^{-1} gives the covariance matrix (5.14.1).

Least squares: The LS estimator of ϕ is as usual obtained by minimizing $\sum_{i=1}^{n} a_i^2$. That gives $\tilde{\phi} = \sum_{i=1}^{n} y_i y_{i-1} / \sum_{i=1}^{n} y_{i-1}^2$. (5.14.2)

The LS estimator of σ is

$$\tilde{\sigma} = \sqrt{\sum_{i=1}^{n} (y_i - \tilde{\phi}_i\, y_{i-1})^2 / n}\,. \tag{5.14.3}$$

Since in time series analysis n is usually large (say, $n \ge 50$), it is immaterial whether the divisor in (5.14.3) is taken to be n or n – 1. The estimators $\tilde{\phi}$ and $\tilde{\sigma}$ are asymptotically unbiased with covariance matrix (asymptotic)

$$\text{Cov}\,(\tilde{\phi}, \tilde{\sigma}) \cong \begin{bmatrix} \dfrac{(1-\phi^2)}{n} & 0 \\ 0 & \left(1 + \dfrac{1}{2}\lambda_4\right)\dfrac{\sigma^2}{2n} \end{bmatrix}, \tag{5.14.4}$$

$\lambda_4 = (\mu_4/\mu_2^2) - 3$; the expression for the kurtosis μ_4/μ_2^2 is given in (5.11.12). The efficiencies of the LS estimators $\tilde{\phi}$ and $\tilde{\sigma}$ relative to the MML estimators $\hat{\phi}$ and $\hat{\sigma}$ are, therefore, the same as those in (5.11.11) and (5.11.13), respectively. The efficiencies are quite low for small p, as stated earlier.

The covariance matrix (5.14.1) is valid for $p \ge 2$. Tiku et al. (2000) give the simulated values of the means and variances for $p < 2$. We report some of their values for p = 1.5 as follows; $\phi = 0.5$, $\sigma = 1$. Note that for $1 < p \le 1.5$, the mean of the distribution exists but not its variance, and σ is simply a scale parameter. This is called Mandelbrot's phenomenon. It can be seen that the LS estimators have very undesirable features: the relative efficiency of $\tilde{\phi}$ is very low and decreases with increasing n, and the bias in $\tilde{\sigma}$ is very high even for sample sizes as large as n = 300. The bias in the MML estimator $\hat{\sigma}$ is almost nonexistent for large n, and $\hat{\phi}$ and $\hat{\sigma}$ are highly efficient. We show in Chapter 8 that the MML estimators are robust due to the umbrella ordering of the coefficients β_i $(1 \le i \le n)$.

		n=100			n=300		
		Mean	(Bias)^2	MSE	Mean	(Bias)^2	MSE
				p = 1.5			
ϕ	LS	0.491	0.000	0.0062	0.496	0.000	0.002
	MML	0.502	0.000	0.0030	0.501	0.000	0.0007
σ	LS	1.945	0.892	6.602	2.100	1.210	5.244
	MML	1.200	0.040	0.655	1.090	0.008	0.134

Small sample efficiency: Since the MMLE are asymptotically the MVB estimators, the LSE are bound to have low relative efficiencies for large n. We give in Table 5.6 the simulated values of the means and variances for smaller sample sizes; (1) = $(1/\sigma)$Mean, (2) = (n/σ^2)Variance, and RE is the relative efficiency of the LS estimator:

Table 5.6: Values of the means and variances: $\phi = 0.5$

Parameter		(1)	(2)	RE	(1)	(2)	RE
				p = 2			
			n = 50			n = 100	
ϕ	LS	0.482	0.730	78	0.491	0.730	70
	MML	0.497	0.570		0.500	0.510	
σ	LS	0.934	5.560	75	0.955	8.000	46
	MML	1.144	4.170		1.093	3.690	
				p = 3.5			
			n = 50			n = 100	
ϕ	LS	0.480	0.785	94	0.489	0.770	92
	MML	0.487	0.735		0.497	0.710	
σ	LS	0.980	1.075	97	0.989	1.110	86
	MML	1.051	1.040		1.035	0.960	

It can be seen that the bias in both the MML as well as the LS estimators is negligible. The MMLE are, however, considerably more efficient than the LSE particularly for smaller values of p. A disconcerting feature of the LSE is that their efficiencies decrease as the sample size n increases.

Hypothesis testing: Testing $\phi = 0$ (random walk) and $\phi = 1$ (unit root problem) are of enormous interest in time series: $\phi = 0$ implies that y_t $(1 \le t \le n)$ are random, and $\phi = 1$ implies that the differences $y_t - y_{t-1}$ $(1 \le t \le n)$ are random. We first discuss the problem of testing $H_0 : \phi = 0$ in the AR(1) model.

Under H_0, the asymptotic variance of the MML estimator $\hat{\phi}$ is $(p + 1)(p - 3/2)/p(p - 1/2)$ $(p \ge 2)$ which is free of σ^2. Therefore, we define the statistic (Tiku et al., 2000)

$$T^* = \sqrt{\left\{\frac{np(p - 1/2)}{(p + 1)(p - 3/2)}\right\}}\ \hat{\phi}\ . \qquad (5.14.5)$$

Large values of T^* lead to the rejection of H_0 in favour of H_1: $\phi > 0$.

Theorem 5.3: The null distribution of T^* is asymptotically normal N(0, 1), for $p \ge 2$.

Proof. Using Taylor's theorem, we have (Kendall and Stuart, 1979, pp. 46-7)

$$\frac{1}{n}\left(\frac{\partial \ln L^*}{\partial \phi}\right) \cong (\hat{\phi} - \phi)\left\{\frac{1}{n}\left(-\frac{\partial^2 \ln L^*}{\partial \phi^2}\right)\right\}_{\phi=\phi^*}$$

where ϕ^* is some value between $\hat{\phi}$ and ϕ (true value). The last term on the right hand side converges to its expected value as n tends to infinity, and $\hat{\phi}$ converges to ϕ and so does ϕ^*. Moreover,

$$\frac{1}{n}\left(\frac{\partial \ln L}{\partial \phi} - \frac{\partial \ln L^*}{\partial \phi}\right) \cong 0 \quad \text{and}$$

$$\frac{1}{n} E\left(-\frac{\partial^2 \ln L}{\partial \phi^2}\right) \cong \frac{1}{n} E\left(-\frac{\partial^2 \ln L^*}{\partial \phi^2}\right) = R(\phi);$$

$R(\phi) = p(p- 1/2)/(p + 1)(p - 3/2)$ under H_0. The result then follows from the fact that $V(\partial \ln L/\partial\phi) = - E(\partial^2 \ln L/\partial\phi^2)$ and $(1/n)(\partial \ln L/\partial\phi)$ is under H_0 the mean of n iid random variables and is, therefore, normally distributed (asymptotically) by the well known Central Limit Theorem.

Power function: For testing H_0: $\phi = 0$ against H_1: $\phi > 0$, the asymptotic power function of the T^* test is

$$\text{Power} = P(Z \geq z_\alpha - |\lambda_1|), \tag{5.14.6}$$

$\lambda_1^2 = np(p - 1/2)\phi^2/(p + 1)(p - 3/2)(1 - \phi^2)$ being the noncentrality parameter; Z is a standard normal variate and z_α is its $100(1 - \alpha)$% point.

The test based on the LS estimator is (Davis and Rensick, 1986; Martin and Yohai, 1985)

$$G^* = \sqrt{n}\tilde{\phi}. \tag{5.14.7}$$

The null distribution of G^* is asymptotically normal N(0, 1). Its asymptotic power function is

$$\text{Power} = P(Z \geq z_\alpha - |\lambda_0|), \tag{5.14.8}$$

$\lambda_0^2 = n\phi^2/(1 - \phi^2)$ being the noncentrality parameter. Since

$$\lambda_1^2/\lambda_0^2 = p(p - 1/2)/(p + 1)(p - 3/2) \tag{5.14.9}$$

is greater than 1 for all values of p (≥ 2), the T* test is asymptotically more powerful than the G* test. Tiku et al. (2000) give the simulated values of the type I error and power of the T* and G* tests. They show that for $n \geq 50$, the normal N(0,1) distribution provides accurate approximations to their percentage points and the T* test has higher power than the G* test. This is due to the fact that the MML estimator $\hat{\phi}$ is more efficient than the LS estimator $\tilde{\phi}$.

For small n (< 50), Tiku et al. (2000) recommend the use of the following statistics,

$$T_1^* = \sqrt{\left\{\left(\frac{2p}{k}\right)\sum_{i=1}^{n} \beta_i y_{[i]-1}^{\,2}\right\}}\left(\frac{\hat{\phi}}{\hat{\sigma}}\right) \quad (p \geq 2) \tag{5.14.10}$$

based on the MML estimators, and (Dickey and Fuller, 1979)

$$G_1^* = \sqrt{\left(\sum_{i=1}^{n} y_{y-i}^{\,2}\right)}\left(\frac{\tilde{\phi}}{\tilde{\sigma}}\right) \tag{5.14.11}$$

based on the LS estimators. The null distributions of T_1^* and G_1^* are referred to the Student t with n – 1 degrees of freedom. The T_1^* test, however, has somewhat higher power than the G_1^* test particularly for small p. Moreover, T_1^* test is robust but not the G_1^* (Chapter 8).

5.15 UNIT ROOT PROBLEM

Testing H_0: $\phi = 1$ is of enormous interest in Econometrics and a vast literature exists on the subject. Most of the work reported is based on the normality assumption; see, for example, Dickey and Fuller (1979), Philips and Perron (1988), Abadir (1993, 1995), and Vinod and Shenton

(1996). The problem is difficult because the variance $V(y_t) = \sigma^2/(1 - \phi^2)$ does not exist if $\phi = 1$. The authors above give a galaxy of mathematical tecniques, in particular applications of Brownian motion, to derive the asymptotic distribution of $n(\tilde{\phi} - 1)$ under H_0: $\phi = 1$; $\tilde{\phi}$ is the LS estimator. Of course, they assume the innovations a_t in (5.13.1) to be normal $N(0, \sigma^2)$.

Tiku and Wong (1998) have a different approach and give solutions in terms of 3-moment chi-square and 4-moment F distributions, assuming that the innovations a_t have one of the distributions in the family (5.11.1). Although their solutions are approximate in nature but they are remarkably accurate for all values of n small or large.

Testing for unit root: To test H_0: $\phi = 1$ against H_1: $\phi < 1$, define the statistic

$$R_1 = \sqrt{n}(2 - \hat{\phi}) \qquad (5.15.1)$$

where $\hat{\phi}$ is the MML estimator (5.13.5). Also

$$R_0 = \sqrt{n}(2 - \tilde{\phi}). \qquad (5.15.2)$$

Large values of R_1 (and R_0) lead to the rejection of H_0 in favour of H_1. Under H_0, $E(R_1) \cong \sqrt{n}$ and $E(R_0) \cong \sqrt{n}$. It may be noted that R_0 and the more traditional statistic $n(\tilde{\phi} - 1)$ are both linear functions of $\tilde{\phi}$ and, therefore, the tests based on them have exactly the same power.

Vinod and Shenton (1996) give the first four moments of $\tilde{\phi}$ when $\phi = 1$ assuming normality of the innovations. Tiku and Wong (1998) give the simulated values of the first four moments of R (and R_0) when $\phi = 1$ and the innovatons have the long-tailed symmetric distribution (5.11.1). As in Dickey and Fuller (1979, p.427), they work with Vinod and Shenton Model A (y_0 is zero) but the results are essentially the same under their Model B. A very interesting finding is that the distributions of R_1 and R_0 are both positively skew and the Pearson coefficients of skewness and kurtosis

$$\beta_1^* = \mu_3^2/\mu_2^3 \quad \text{and} \quad \beta_2^* = \mu_4/\mu_2^2 \qquad (5.15.3)$$

are very close to the Type III line or are located in the F-region (Pearson and Tiku, 1970, Fig. 1). The values of the mean and variance, and β_1^* and β_2^*, of both R_0 and R_1 are given in Tiku and Wong (1999, Table III). For $p = \infty$ (normal innovations) and $n \geq 100$, these values are in close agreement with the asymptotic equations of Vinod and Shenton (1996, eqs 3.5).

Three-moment chi-square: Let μ_1' and μ_2 be the mean and variance of a random variable Y and β_1^* and β_2^* its skewness and kurtosis, $\mu_3 > 0$. If the coefficients β_1^* and β_2^* satisfy the condition (Pearson, 1959; Tiku, 1963, 1966b)

$$|\beta_2^* - (3 + 1.5\beta_1^*)| \leq 0.5, \qquad (5.15.4)$$

then the distribution of

$$\chi^2 = (Y + a)/f \qquad (5.15.5)$$

is effectively a central chi-square with ν degrees of freedom. The values of a, f, and ν are obtained by equating the first three moments on both sides of (5.15.5). That gives

$$\nu = 8/\beta_1^*, \quad f = \sqrt{\mu_2/2\nu} \quad \text{and} \quad a = f\nu - \mu_1'. \qquad (5.15.6)$$

Realize that for a chi-square distribution $\beta_2^* = 3 + 1.5\beta_1^*$ which is called the Type III line (Pearson and Tiku, 1970, Fig. 1).

Four-moment F: Let (Tiku and Yip, 1978)

$$F = (Y + g)/h \qquad (5.15.7)$$

where F has a central F distribution with (ν_1, ν_2) degrees of freedom. Equating the first four moments on both sides of (5.15.7), the values of ν_2, ν_1, h and g are obtained:

$$v_2 = 2\left[3 + \frac{\beta_2^* + 3}{\beta_2^* - (3 + 1.5\beta_1^*)}\right]$$

$$v_1 = \frac{1}{2}(v_2 - 2)\left(-1 + \sqrt{1 + \frac{32(v_2-4)/(v_2-6)^2}{\beta_1^* - 32(v_2-4)/(v_2-6)^2}}\right) \qquad (5.15.8)$$

$$h = \sqrt{\left(\frac{v_1(v_2-2)^2(v_2-4)}{2v_2^2(v_1+v_2-2)}\mu_2\right)} \quad \text{and} \quad g = \frac{v_2}{(v_2-2)}\, h - \mu_1'.$$

For the F distribution $v_1 > 0$ and $v_2 > 0$ and, therefore, for (5.15.7) to be valid the (β_1*, β_2*) values of Y should satisfy

$$\beta_1^* > 32(v_2-4)/(v_2-6)^2 \qquad \text{and} \qquad \beta_2^* > 3 + 1.5\beta_1^*. \qquad (5.15.9)$$

It is interesting to note that the inequalities (5.15.9) determine the F-region in the Pearson plane bounded by the χ^2-line and the reciprocal χ^2-line (Pearson and Tiku, 1970). Whenever the (β_1*, β_2*) points of Y lie within this F-region, the 4-moment approximation (5.15.7) provides remarkably accurate values for the probability integral and the percentage points of Y (Tiku and Yip, 1978). Thus, the 100(1 – α)% point of Y is approximately $hF_\alpha(v_1,v_2) - g$, where $F_\alpha(v_1, v_2)$ is the 100(1 – α)% point of the F distribution with (v_1, v_2) degrees of freedom. An IMSL subroutine is available to evaluate the percentage points of F for both integer as well as non-integer values of v_1 and v_2.

It is interesting to note that the (β_1*, β_2*) values of R_1 and R_0 satisfy (5.15.9) for all values of p and n, even for p =1 in which case (5.11.1) is a Cauchy distribution; for p < 2, k is equated to 1. For n ≥ 100, however, the (β_1*, β_2*) values generally satisfy the condition (5.15.4). Therefore, the chi-square and F approximations above are applicable.

To illustrate the accuracy of these approximations, we give in Table 5.7 (reproduced from Tiku and Wong, 1998) the values of the probabilities

$$P(R_1 \ge d_1 \mid \phi = 1) \quad \text{and} \quad P(R_0 \ge d_0 \mid \phi = 1) \qquad (5.15.10)$$

where d_1 and d_0 are the 95% points obtained from (5.15.7). If, however, the chi-square approximation is applicable we obtain them from (5.15.5). Realize that the LS estimator $\tilde{\phi}$ is not defined if p =1 (Cauchy distribution).

It can be seen that the chi-square and F approximations are remarkably accurate for all n.

Tiku and Wong (1998) show that the R_1 test is considerably more powerful than the R_0 test.

It may be noted that the β_1* and β_2* values of R_0 (and R_1) are never close to zero and 3, respectively, even under the assumption of normality. The normal approximations suggested in the literature (see, for example, Abadir 1995) should, therefore, be used with great deal of caution. They can give erroneous results.

5.16 UNKNOWN LOCATION

In numerous situations one has the AR(1) model

$$y_t = \mu + \phi y_{t-1} + a_t \ (1 \le t \le n). \qquad (5.16.1)$$

Assume that a_t are iid and have one of the LTS distributions in the family (5.11.1). Here,

Table 5.7: Values of (a) 95 % points of R_1 and R_0, and (b) the probabilities $P(R_1 \geq d_1|\phi = 1)$ and $P(R_0 \geq d_0|\phi = 1)$.

		(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
n		p = 1		p = 1.5		p = 2		p = 2.5	
30	d_1	5.906	0.037	6.200	0.044	6.397	0.046	6.557	0.051
	d_0	—	—	6.756	0.049	6.765	0.053	6.797	0.050
50	d_1	7.355	0.038	7.593	0.046	7.752	0.049	7.874	0.050
	d_0	—	—	8.087	0.050	8.104	0.052	8.132	0.050
100	d_1	10.171	0.041	10.327	0.045	10.460	0.048	10.532	0.049
	d_0	—	—	10.761	0.047	10.778	0.052	10.764	0.051
		p = 3		p = 3.5		p = 5		p = 10	
30	d_1	6.591	0.049	6.766	0.050	6.828	0.051	6.806	0.049
	d_0	6.794	0.050	6.832	0.049	6.858	0.050	6.811	0.049
50	d_1	7.948	0.051	8.022	0.051	8.146	0.047	8.135	0.051
	d_0	8.134	0.049	8.162	0.050	8.172	0.046	8.141	0.051
100	d_1	10.588	0.053	10.684	0.051	10.763	0.051	10.768	0.051
	d_0	10.749	0.053	10.814	0.049	10.793	0.051	10.776	0.052

$$E(y_t) = \mu/(1-\phi) \quad \text{and} \quad V(y_t) = \sigma^2/(1-\phi^2).$$

The modified likelihood equation for estimating μ is

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{2p}{k\sigma}\sum_{i=1}^{n}(\alpha_i - \beta_i z_{(i)}) = 0; \qquad (5.16.2)$$

$z_{(i)} = \{y_{[i]} - \mu - \phi y_{[i]-1}\}/\sigma$. Since $\sum_i \alpha_i = 0$, this gives

$$\mu = \sum_{i=1}^{n}\beta_i\{y_{[i]} - \phi y_{[i]-1}\}/m, \quad m = \sum_{i=1}^{n}\beta_i. \qquad (5.16.3)$$

Incorporating this in (5.13.3) and solving the resulting equations, we get the MML estimators $\hat{\phi}$ and $\hat{\sigma}$. They are exactly the same as (5.13.5) with $y_{[i]}$ and $y_{[i]-1}$ replaced by $w_{[i]}$ and $w_{[i-1]}$, respectively;

$$w_{[i]} = y_{[i]} - (1/m)\sum_{i=1}^{n}\beta_i y_{[i]} \quad \text{and} \quad w_{[i-1]} = y_{[i]-1} - (1/m)\sum_{i=1}^{n}\beta_i y_{[i]-1}. \qquad (5.16.4)$$

It may be noted that

$$\sum_{i=1}^{n}\beta_i w_{[i]} = 0 \quad \text{and} \quad \sum_{i=1}^{n}\beta_i w_{[i-1]} = 0.$$

The MML estimator of μ is

$$\hat{\mu} = \sum_{i=1}^{n}\beta_i\{y_{[i]} - \hat{\phi} y_{[i]-1}\}/m. \qquad (5.16.5)$$

The asymptotic variances and covariances are given by $I^{-1}(\mu, \phi, \sigma)$, I being the Fisher information matrix consisting of the elements $-E(\partial^2 \ln L/\partial\mu^2)$, $-E(\partial^2 \ln L/\partial\mu\partial\phi)$, etc. They can be obtained exactly along the same lines as in Appendix 5B. In particular, $\hat{\mu}$ and $\hat{\phi}$ are asymptotically uncorrelated with $\hat{\sigma}$.

The MML estimators of μ, θ and σ can readily be obtained if a_t has a skew distribution. For a gamma distribution, for example, Tiku et al. (1999, Section 6) give the MML estimators. The asymptotic variance-covariance matrix they give needs correcting for the fact that $(1/n)\sum_i \alpha_i$ is equal to $2/(k-1)$ for large n and not $1/(k-1)$; see Akkaya and Tiku (2001b, p.2230). The variances and covariances of the estimators can be obtained from the appropriate expressions in Appendix 5B simply by equating δ to zero. Tiku et al. (1999, Section 8) also give results for the Generalized Logistic GL(b, σ). The expression for $-E(\partial^2 \ln L^*/\partial\phi\partial\sigma)$ in their equation (34) should be multiplied by $\psi(b+1)-\psi(2)$.

5.17 GENERALIZATION TO AR(q) MODELS

The method of modified likelihood immediately generalizes to the AR(q) model

$$y_t = \sum_{j=1}^{q} \phi_j y_{y-j} + a_t \quad (1 \le t \le n). \tag{5.17.1}$$

Assume that a_t has one of the LTS distributions in the family (5.11.1). The MML estimators of ϕ_j $(1 \le j \le q)$ and σ are the solutions of the q +1 equations

$$C_{j1}\phi_1 + C_{j2}\phi_2 + \ldots + C_{jq}\phi_q = K_j + D_j\sigma \quad (1 \le j \le q) \quad \text{and} \tag{5.17.2}$$

$$n\sigma^2 - \frac{2p}{k}\left(\sum_{i=1}^{n} \alpha_i y_{[i]}\right)\sigma - \frac{2p}{k}\sum_{i=1}^{n} \beta_i\, y_{[i]}\{y_{[i]} - \phi_1 y_{[i]-1} \ldots - \phi_q y_{[i]-q}\} = 0 \tag{5.17.3}$$

where $(y_{[i]}, y_{[i]-1}, \ldots, y_{[i]-q}]$ is that observation $(y_i, y_{i-1}, \ldots, y_{i-q})$ which determines $z_{(i)}$ $(1 \le i \le n)$, as in the AR(1) model (5.13.1). Here (j, l = 1, 2,…, q)

$$C_{jl} = \sum_{i=1}^{n} \beta_i y_{[i]} y_{[i]-l}, \quad C_{jl} = C_{lj} \quad j \ne l, \quad \text{and}$$

$$K_j = \sum_{i=1}^{n} \beta_i y_{[i]} y_{[i]-j} \quad \text{and} \quad D_j = \sum_{i=1}^{n} \alpha_i y_{[i]-j}, \quad 1 \le j \le q. \tag{5.17.4}$$

The MML estimators are

$$\hat{\varphi} = C^{-1}(K + D\hat{\sigma}) \quad \text{and} \quad \hat{\sigma} = \{B + \sqrt{(B^2 + 4nH)}\,/\,2n,$$

$$B = \frac{2p}{k}\left(\sum_{i=1}^{n} \alpha_i y_{[i]} - K'C^{-1}D\right) \quad \text{and} \quad H = \frac{2p}{k}\left(\sum_{i=1}^{n} \beta_i y_{[i]}{}^2 - K'C^{-1}K\right); \tag{5.17.5}$$

$\hat{\varphi} = (\phi_1, \phi_2, \ldots, \phi_q)$, $K = (K_1, K_2, \ldots, K_q)$ and $C = (C_{ij})$. For q = 1, they reduce to (5.13.5). If for a sample $H < 0$, we calculate the estimators with α_i and β_i replaced by zero and $\beta_1{}^*$, respectively, as said earlier. The estimators $\hat{\varphi}$ and $\hat{\sigma}$ have the same efficiency properties as for q = 1.

Cryer (1986) gives numerous real life time series data sets mostly having non-normal error distributions. For these data sets, the MMLE are considerably more efficient than the corresponding LSE.

SUMMARY

In linear regression models $y_i = \theta_0 + \theta_1 x_i + e_i$ $(1 \le i \le n)$ considered in Chapter 3, the errors e_i $(1 \le i \le n)$ are assumed to be independently distributed. In numerous applications, however, e_i are correlated in a special way which gives rise to autoregressive models $y_t - \phi y_{t-1} = \mu + \delta(x_t - \phi x_{t-1}) + a_t$; a_t $(1 \le t \le n)$ are independent innovations. In this chapter, we consider the estimation of the parameters μ, δ, ϕ and σ(the scale parameter in the distribution of a_t) for both normal as well as non-normal innovations. This has been a very difficult problem analytically and computationally. The MLE being intractable, we obtain the MMLE and show that the estimators are highly efficient. We compare the efficiencies of the MMLE and the LSE and show that the latter are considerably less efficient for all the three families of distributions considered. For normal innovations, the MMLE are identical to the LSE. We develop procedures for testing H_0: $\delta = 0$. For $\delta = 0$, the model reduces to a time series AR(1) model $y_t = \mu + \phi y_{t-1} + a_t$ $(1 \le t \le n)$: We derive the MMLE of μ, ϕ and σ when a_t have a normal or a non-normal distribution. The estimators are asymptotically fully efficient and are highly efficient for small sample sizes. The relative efficiencies of the LSE are low and decrease with increasing n. We develop procedures for testing the two important hypotheses $\phi = 0$ and $\phi = 1$ (the important unit root problem). We generalize some of the results to time series AR(q) models.

APPENDIX 5A

EXPECTED VALUE AND VARIANCE OF y_t

Suppose that $E(y_0) = 0$ and $V(y_0) = \sigma^2/(1 - \phi^2)$. Let $E(a_t) = c\sigma$ and $V(a_t) = \sigma^2$. Since y_{t-1} is independent of a_t,

$$V(y_t) = \left(1 + \frac{\phi^2}{1-\phi^2}\right)\sigma^2 = \frac{\sigma^2}{(1-\phi^2)}, \quad 1 \le t \le n\,; \tag{5A.1}$$

$E(y_t)$ is different from $E(y_0)$ and is hard to evaluate. The sum of the expected values of y_t can, however, be evaluated for large n as follows:

From the model (5.2.2),

$$\begin{aligned} E(y_1) &= \mu^* + \delta(x_1 - \phi x_0) \\ E(y_2) &= \phi[\mu^* + \delta(x_1 - \phi x_0)] + \mu^* + \delta(x_2 - \phi x_1) \\ E(y_3) &= \phi^2(\mu^* + \delta(x_1 - \phi x_0)] + \phi[\mu^* + \delta(x_2 - \phi x_1)] + \mu^* + \delta(x_3 - \phi x_2) \end{aligned} \tag{5A.2}$$

and so on.

For sufficiently large n, therefore, we have

$$A_n = \frac{1}{n}\sum_{t=1}^{n} E(y_t) \cong \frac{1}{1-\phi}\left[\mu^* + \frac{\delta}{n}\{((x_1 - \phi x_0) + (x_2 - \phi x_1) + \ldots\}\right]; \tag{5A.3}$$

$(n + 1) A_{n+1} - nA_n$ suggests the solution

$$E(y_t) \cong \frac{1}{1-\phi}[\mu^* + \delta(x_t - \phi x_{t-1})] \quad (\mu^* = \mu + c\sigma), \tag{5A.4}$$

for $|\phi| < |$. We have not, however, used it to evaluate the asymptotic variances of the MMLE and the LSE of δ, the parameter of main interest in (5.2.2).

The modified likelihood methodology extends to k design variables on the right hand side of (5.2.2). The details are given in Turker and Akkaya (2004). The methodology readily extends to the model

$$y_t - \phi y_{t-1} = \mu + \delta_1 x_{1,t} + \ldots + \delta_k x_{k,t} + a_t \quad (1 \le t \le n). \tag{5A.5}$$

Remark: For the Gamma (k, σ) the bias in the MML and the LS estimators $\hat{\mu}$ and $\tilde{\mu}$ is considerable for small n. However, the bias tends to zero as n becomes large. For symmetric distributions, the bias in $\hat{\mu}$ and $\tilde{\mu}$ is negligible even for small n. See also Chapter 8. We recommend, however, that the first observation be ignored. This may reduce the bias by a margin.

CHAPTER 6

Analysis of Variance in Experimental Design

6.1 INTRODUCTION

Experimental design constitutes a very important area from both theoretical as well as applications point of view. Treatments are compared and their interactions evaluated. This is done by using the technique of analysis-of-variance (ANOVA) and defining variance ratio F statistics. Traditionally, the assumption of normality has been invoked in the development of the technique. In practice, however, non-normal distributions are more prevalent as said earlier. It has, therefore, been of great interest to study the effect of non-normality on the F statistics used for testing main effects and interaction (Geary, 1947; Gayen, 1950; Srivastava, 1959; Tiku, 1964; Donaldson, 1968; Tiku, 1971b; Spjøtvoll and Aastveit, 1980; Rasch, 1980; Tan and Tiku, 1999). It is known that for numerous non-normal distributions, the type I error of the variance ratio tests is not much different than that for a normal distribution (essentially due to the Central Limit Theorem) but the power is considerably lower (due to inefficiency of the sample mean). This has already been illustrated in Chapter 1 (Examples 1.1-1.4). One way of handling non-normal data is to invoke Box and Cox (1964) lambda-transformation so that the transformed data is normal, at any rate close to it. Bickel and Doksum (1981) make some interesting comments on this method. It should, however, be clear that all non-normal data can not be amenable to this transformation or for that matter to any other normalizing transformation. Moreover, it is often difficult to interpret transformations. It is, therefore, desirable to work with the original data without subjecting it to any so called normalizing transformation. There is, therefore, a need to develop analysis-of-variance procedures for non-normal data. That is exactly what we try to accomplish in this chapter. We will consider estimation of main effects and interaction and develop appropriate variance ratio F statistics. Our approach is applicable to any location-scale distribution but we will focus on the Generalized Logistic family GL(b; σ) for its flexibility. Moreover, GL(b; σ) provides a good model for the Box and Cox(1964) biometrical data which has received a great deal of attention in recent years. We also give procedures for testing linear contrasts. Extensions to other non-normal distributions, and to nonidentical error distributions, are given.

6.2 ONE-WAY CLASSIFICATION

Consider the one-way-classification fixed-effects model

$$y_{ij} = \mu + \gamma_i + e_{ij} \quad (i = 1, 2, ..., k;\ j = 1, 2, ..., n), \tag{6.2.1}$$

having k blocks with n observations in each block; μ is a constant and γ_i is the effect of the i^{th} block. Without loss of generality we assume that $\sum_{i=1}^{k} \gamma_i = 0$. The errors e_{ij} are assumed to be iid.

Normal errors: As said earlier, e_{ij} have traditionally been assumed to be iid normal $N(0, \sigma^2)$ which gives the likelihood function

$$L \propto \left(\frac{1}{\sigma}\right)^N \prod_{i=1}^{k} \prod_{j=1}^{n} e^{-(y_{ij} - \mu - \gamma_i)^2/2\sigma^2} \quad (N = nk). \tag{6.2.2}$$

The ML estimators are the solutions of the likelihood equations $\partial \ln L/\partial\mu = 0$, $\partial \ln L/\partial\gamma_i = 0$ and $\partial \ln L/\partial\sigma = 0$. They are

$$\hat{\hat{\mu}} = \bar{y}_{..} = (1/N) \sum_{i=1}^{k} \sum_{j=1}^{n} y_{ij},\ \hat{\hat{\gamma}}_i = \bar{y}_i - \bar{y}_{..},\ \bar{y}_i = (1/n) \sum_{j=1}^{k} y_{ij}, \tag{6.2.3}$$

and

$$\hat{\hat{\sigma}}^2 = s_e^{\,2} = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2/(N - k). \tag{6.2.4}$$

Here, we have the following forms of the derivatives of ln L:

$$\frac{\partial \ln L}{\partial \mu} = \frac{N}{\sigma^2}(\bar{y}_{..} - \mu), \quad \frac{\partial \ln L}{\partial \gamma_i} = \frac{n}{\sigma^2}(\bar{y}_{i.} - \mu_i)\ (\mu_i = \mu + \gamma_i), \tag{6.2.5}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = \frac{N}{\sigma^3}(S^2 - \sigma^2),\ S^2 = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \mu - \gamma_i)^2.$$

We also have the Cochran identity

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2 \equiv n \sum_{i=1}^{k} (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 \tag{6.2.6}$$

with expected values

$$(N - 1)\sigma^2 \equiv (k - 1)\sigma^2 + (N - k)\sigma^2 \quad (\gamma_i = 0 \text{ for all } i).$$

The equations (6.2.5)-(6.2.6) are exactly similar to those in Section 2.7 which gives the result that if $\gamma_i = 0$ $(1 \le i \le k)$, then the two sums of squares on the right hand side of (6.2.6) are independently distributed as multiples of chi-square. In fact,

$$(k - 1)s_b^{\,2}/\sigma^2 = n \sum_{j=1}^{n} (\bar{y}_{i.} - \bar{y}_{..})^2/\sigma^2$$

and

$$(k - 1)s_e^{\,2}/\sigma^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2/\sigma^2 \tag{6.2.7}$$

are independently distributed as chi-squares with $\nu_1 = k - 1$ and $\nu_2 = N - k$ degrees of freedom. Consequently, the distribution of

$$F = s_b^{\,2}/s_e^{\,2} \tag{6.2.8}$$

under $H_0 : \gamma_i = 0$ $(1 \le i \le k)$ is central F with $\nu_1 = k - 1$ and $\nu_2 = N - k = k(n - 1)$ degrees of freedom. Under $H_1 : \gamma_1 \ne 0$ (for some i) the two sums of squares on the right hand side of (6.2.6) when divided by σ^2 are independently distributed as noncentral chi-square and central chi-square distributions, respectively. Thus, the distribution of F is noncentral F with (ν_1, ν_2) degrees of freedom and noncentrality parameter

$$\lambda^2 = n \sum_{i=1}^{k} (\gamma_i/\sigma)^2. \tag{6.2.9}$$

Large values of F lead to the rejection of H_0 in favour of H_1. Under the normality assumption, the F statistic provides the most powerful test of H_0. Under non-normality, its type I error is generally not much different than that under normality but its power is adversely affected. This has been illustrated in Chapter 1 (Example 1.4).

For unbalanced designs, i.e. n_i not necessarily equal, we assume without loss of generality that $\sum_{i=1}^{k} n_i \gamma_i = 0$. The F statistic is the same as in (6.2.8) but

$$(k-1)s_b^2 = \sum_{i=1}^{k} n_i(\bar{y}_{i.} - \bar{y}_{..})^2 \quad \text{and} \quad (N-k)s_e^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} (\bar{y}_{ij} - \bar{y}_{i.})^2; \tag{6.2.10}$$

$$N = \sum_{i=1}^{k} n_i,\ \bar{y}_{i.} = \sum_{j=1}^{n} y_{ij}/n_i \quad \text{and} \quad \bar{y}_{..} = \sum_{j=1}^{k} n_i \bar{y}_{i.}/N = \sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}/N.$$

The distribution of $F = s_b^2/s_e^2$ is the same as for balanced designs but the noncentrality parameter is

$$\lambda^2 = \sum_{i=1}^{k} n_i(\gamma_i/\sigma)^2. \tag{6.2.11}$$

By adopting the methodology of modified likelihood, we extend the technique to non-normal distributions as follows:

6.3 GENERALIZED LOGISTIC

Suppose that the errors e_{ij} are iid and have one of the distributions in the family of Generalized Logistic (b > 0)

$$GL(b;\sigma) : f(e) = \frac{b}{\sigma}\frac{\exp(-e/\sigma)}{\{1+\exp(-e/\sigma)\}^{b+1}} \quad (-\infty < e < \infty). \tag{6.3.1}$$

Its mean and variance and the Pearson coefficients of skewness and kurtosis are given in Appendix 2D (Chapter 2).

For the model (6.2.1), the likelihood function L is

$$L \propto \left(\frac{1}{\sigma}\right)^N \prod_{i=1}^{k}\prod_{j=1}^{n} [\exp(-z_{ij})/\{1+\exp(-z_{ij})\}^{b+1}]; \tag{6.3.2}$$

$$z_{ij} = (y_{ij} - \mu - \gamma_i)/\sigma. \tag{6.3.3}$$

The likelihood equations for estimating μ, γ_i $(1 \le i \le k)$ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{N}{\sigma} - \frac{(b+1)}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n} g(z_{ij}) = 0 \tag{6.3.4}$$

$$\frac{\partial \ln L}{\partial \gamma_i} = \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\sum_{j=1}^{n} g(z_{ij}) = 0 \tag{6.3.5}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij} - \frac{(b+1)}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}\, g(z_{ij}) = 0; \tag{6.3.6}$$

the function g(z) is given by

$$g(z) = e^{-z}/(1+e^{-z}). \tag{6.3.7}$$

The only way to solve the equations (6.3.4)-(6.3.6) is by iteration but that is difficult and time consuming indeed, since there are k + 1 equations to iterate simultaneously.

6.4 MODIFIED LIKELIHOOD

Let $y_{i(1)} \le y_{i(2)} \le \ldots\ldots \le y_{i(n)} \quad (1 \le i \le k)$ (6.4.1)

be the order statistics of the n observations y_{ij} $(1 \le j \le n)$ in the i^{th} block. Then

$$z_{i(j)} = \{y_{i(j)} - \mu - \gamma_i\}/\sigma \quad (1 \le i \le k) \qquad (6.4.2)$$

are the ordered z_{ij} $(1 \le i \le n)$ variates. Alternative expressions of the likelihood equations are obtained by replacing z_{ij} by $z_{i(j)}$.

As in (2.5.4), we have the linear approximations $(1 \le i \le k)$

$$g(z_{i(j)}) \cong \alpha_j - \beta_j z_{i(j)} \quad (1 \le j \le n) \qquad (6.4.3)$$

where $\alpha_j = (1 + e^t + te^t)/(1 + e^t)^2$ and $\beta_j = e^t/(1 + e^t)^2$, $t = t_{(j)} = E\{z_{i(j)}\}$. (6.4.4)

For $n \ge 10$, we use the approximate values of $t_{(j)}$ given in equations (2.5.4)-(2.5.7), i.e.,

$$t_{(j)} = -\ln(q_j^{-1/b} - 1), \quad q_j = j/(n + 1). \qquad (6.4.5)$$

As said earlier, the use of these approximate values in place of the true values does not affect the efficiencies of the resulting estimators too adversely. The true values are available in Balakrishnan and Leung (1988) for $n \le 15$.

Incorporating (6.4.3) in the alternative expressions obtained by replacing z_{ij} in (6.3.4)-(6.3.6) by $z_{i(j)}$, we get the modified likelihood equations $\partial \ln L^*/\partial\mu = 0$, $\partial \ln L^*/\partial\gamma_i = 0$ and $\partial \ln L^*/\partial\sigma = 0$. These equations have explicit solutions, the MML estimators (Şenoğlu and Tiku, 2001):

$$\hat{\mu} = \hat{\mu}_{..} - (\Delta / m)\hat{\sigma} \qquad (6.4.6)$$

$$\hat{\gamma}_i = \hat{\mu}_{i.} - \hat{\mu}_{..} \text{ and } \hat{\sigma} = \{-B + \sqrt{B^2 + 4knC}\}/2\sqrt{k(n-1)}; \qquad (6.4.7)$$

$$\Delta = \sum_{j=1}^{n} \Delta_j, \; \Delta_j = \alpha_j - (b+1)^{-1}; \; m = \sum_{j=1}^{n} \beta_j; \; B = \sum_{i=1}^{k} B_i, \; C = \sum_{j=1}^{k} C_i;$$

$$B_i = (b+1)\sum_{j=1}^{n} \Delta_j\,(y_{i(j)} - \hat{\mu}_{i.}), \qquad (6.4.8)$$

$$C_i = (b+1)\sum_{j=1}^{n} \beta_j\,(y_{i(j)} - \hat{\mu}_{i.})^2 = (b+1)\left(\sum_{j=1}^{n} \beta_j\, y_{i(j)}{}^2 - m\hat{\mu}_i{}^2\right);$$

$$\hat{\mu}_{..} = (1/km)\sum_{i=1}^{k}\sum_{j=1}^{n} \beta_j y_{i(j)} = (1/k)\sum_{i=1}^{k} \hat{\mu}_{i.} \text{ and } \hat{\mu}_{i.} = (1/m)\sum_{j=1}^{n} \beta_j y_{i(j)}. \qquad (6.4.9)$$

The estimators are explicit functions of sample observations. No iterations, whatsoever, are required to compute the MML estimators (6.4.6)-(6.4.9). Also, $\hat{\sigma}$ is real and positive since $\beta_j > 0$ for all j.

The MML estimators above have the following interesting properties.

Lemma 6.1: Asymptotically (n tends to infinity), the estimator $\hat{\mu}_i(\sigma) = \hat{\mu}_{i.} - (\Delta/m)\sigma$ is conditionally (σ known) the MVB estimator of $\mu_i = \mu + \gamma_i$ $(1 \le i \le k)$ and is normally distributed with variance

$$V\{\hat{\mu}_i(\sigma)\} \cong \sigma^2/m(b + 1). \qquad (6.4.10)$$

Proof: The result follows from the fact that $\partial \ln L^*/\partial\mu_i$ is asymptotically equivalent to $\partial \ln L/\partial\mu_i$ and assumes the form

$$\frac{\partial \ln L^*}{\partial \mu_i} = \frac{m(b+1)}{\sigma^2}\{\hat{\mu}_i(\sigma) - \mu_i\} \qquad (6.4.11)$$

and (equation 2A.4)

$$\lim_{n \to \infty} \frac{m}{n} = E\{e^{-z}/(1 + e^{-z})^2\} = b/(b + 1)(b + 2). \quad (6.4.12)$$

The normality follows from the fact that $E(\partial^r \ln L^*/\partial\mu_i^r) = 0$ for all $r \geq 3$.

Lemma 6.2: The estimator $\hat{\gamma}_i = \hat{\mu}_{i.} - \hat{\mu}_{..}$ is an unbiased estimator of γ_i for all n, and is asymptotically normally distributed with variance $\sigma^2/m(b + 1)$.

Proof: This is true exactly for the same reasons as in Lemma 6.1, and the fact that in view of $\partial \ln L^*/\partial\mu = 0$, $\partial \ln L^*/\partial\gamma_i$ when re-organized assumes the form

$$\frac{\partial \ln L^*}{\partial \gamma_i} = \frac{m(b+1)}{\sigma^2}(\hat{\gamma}_i - \gamma_i) \quad (6.4.13)$$

with $E(\partial^r \ln L^*/\partial\gamma_i^r) = 0$ for $r \geq 3$. Since for any n, $E(\hat{\mu}_i - \mu - \gamma_i)$ are equal for all i = 1, 2, …, k and $\sum_i \gamma_i = 0$, $\hat{\gamma}_i$ is an unbiased estimator of γ_i. Realize that $m(b + 1) \cong nb/(b + 2)$ from equation (6.4.12).

Corollary: Since $\hat{\mu}_{..} = (1/k)\sum_{i=1}^{k} \hat{\mu}_{i.}$ and $\hat{\mu}_{i.}$ $(1 \leq i \leq k)$ are independent of one another, a sharper result is that

$$V(\hat{\gamma}_i) \cong (k - 1)\sigma^2/mk(b + 1), \ 1 \leq i \leq k. \quad (6.4.14)$$

Remark: Asymptotically, the estimators $\hat{\gamma}_i$ $(1 \leq i \leq k)$ are independent of $\hat{\sigma}$. This follows from the fact that all mixed derivatives $E(\partial^{r+s} \ln L^*/\partial\gamma_i^r\, \partial\sigma^s) = 0$ for all $r \geq 1$ and $s \geq 1$ as explained in Chapter 2. The estimators $\hat{\gamma}_i$ and $\hat{\sigma}$ are, however, uncorrelated for all n.

Lemma 6.3: Asymptotically, $N\hat{\sigma}^2(\mu)/\sigma^2$ is conditionally ($\mu_i = \mu + \gamma_i$ known) distributed as chi-square with N = nk degrees of freedom.

Proof: Write $B_0 = \sum_{i=1}^{k} B_{i0}$ and $C_0 = \sum_{i=1}^{k} C_{i0}$, where

$$B_{i0} = (b + 1)\sum_{j=1}^{n} \Delta_j(y_{i(j)} - \mu_i) \text{ and } C_{i0} = (b + 1)\sum_{j=1}^{n} \beta_j(y_{i(j)} - \mu_i)^2.$$

Realizing that $B_0/\sqrt{(nC_0)} \cong 0$ for large n, it is easy to show that

$$\frac{\partial \ln L^*}{\partial \sigma} \cong \frac{N}{\sigma^3}\left(\frac{C_0}{N} - \sigma^2\right). \quad (6.4.15)$$

The result then follows from the values of $E(\partial^r \ln L^*/\partial\sigma^r)$ as in Chapter 2. Note also that like normal-theory sample means and sums of squares,

$$\sum_{i=1}^{k}\sum_{j=1}^{n} \beta_j(y_{i(j)} - \hat{\mu}_{..})^2 \equiv m\sum_{i=1}^{k}(\hat{\mu}_{i.} - \hat{\mu}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n} \beta_j(y_{i(j)} - \hat{\mu}_{i.})^2. \quad (6.4.16)$$

The identity is instrumental in defining the variance ratio statistic for testing the equality of block effects.

Exact variances: From the results (6.4.13)-(6.4.14), it follows that $\hat{\mu}_{i.}$ is asymptotically the minimum variance bound estimator with variance $(b + 2)\sigma^2/nb$. To see how efficient it is for small sample sizes, we give below the exact values of the variance $V(\hat{\mu}_{i.})$ calculated from the expected values and the variances and covariances of order statistics. Also given are the exact relative efficiencies $E_1 = 100\{MVB/V(\hat{\mu}_{i.})\}$ and $E_2 = 100\{MVB/V(\bar{y}_{i.})\}$. It can be seen that the MML estimator $\hat{\mu}_{i.}$ is highly efficient even for sample sizes as small as $n = 6$. The LS estimator (sample mean) has low relative efficiency $E = 100\{V(\hat{\mu}_{i.})/V(\bar{y}_{i.})\}$ which decreases with increasing n. That is not a good prospect from a theoretical as well as a practical point of view. The same is true (Senoğlu and Tiku,2001) about the relative efficiency of the LS estimator $s_e = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2/(N-k)}$ as compared to the ML estimator $\hat{\sigma}$.

Minimum variance bounds and the exact variances for $GL(b;\sigma)$.

n =	6	10	15	n = 6	10	15
		b = 2			b = 6	
$(1/\sigma^2)$MVB	0.333	0.200	0.133	0.208	0.125	0.0833
$(1/\sigma^2)\ V(\hat{\mu}_{i.})$	0.344	0.204	0.135	0.221	0.129	0.0850
E_1	96.8	98.0	98.5	94.1	96.9	98.0
E_2	87.3	87.3	87.3	68.4	68.4	68.4

6.5 TESTING BLOCK EFFECTS

Testing that the block effects $\mu_i = \mu + \gamma_i$ are all equal is equivalent to testing the null hypothesis $H_0 : \gamma_1 = 0$ for all i = 1, 2, 3, ..., k. The normal-theory variance ratio statistic is

$$F = n\sum_{i=1}^{n}(\bar{y}_{i.} - \bar{y}_{..})^2/(k-1)\ s_e^2,\quad s_e^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i.})^2/(N-k). \tag{6.5.1}$$

The variance ratio statistic based on the estimators (6.4.7)-(6.4.9) is

$$F^* = m(b+1)\sum_{i=1}^{k}\hat{\gamma}_i^{\,2}/(k-1)\,\hat{\sigma}^2. \tag{6.5.2}$$

Like the statistic F, large values of F* lead to the rejection of H_0 in favour of $H_1 : \gamma_i \neq 0$ (for some i).

In view of the results (6.4.13)-(6.4.15), the null distribution of F* for large n is central F with $\nu_1 = k - 1$ and $\nu_2 = N - k = k(n-1)$ degrees of freedom. The distribution of F* under H_1 is for large n noncentral F with (ν_1, ν_2) degrees of freedom and noncentrality parameter

$$\lambda_1^2 = m(b+1)\sum_{i=1}^{k}(\gamma_i/\sigma)^2. \tag{6.5.3}$$

The noncentrality parameter of the non-null distribution of F is

$$\lambda_2^2 = n\sum_{i=1}^{k}(\gamma_i/\sigma)^2. \tag{6.5.4}$$

Using the result (6.4.12), $\lambda_1^2/\lambda_2^2 = (b+2)/b$ which is always greater than 1. The F* test is, therefore, more powerful for large n than the traditional F test.

Small sample results: The distributional results above are applicable to small samples as well. Given in Table 6.1 are the simulated values of the type I error and the power of the F* and F tests, reproduced from Şenoğlu and Tiku (2001). For d = 0, the values are Type I errors.

Table 6.1: Values of the power of F* and F tests for GL(b; σ); k = 4, n = 10 .

b	d =	0.0	0.2	0.4	0.6	0.8	1.0
0.5	F*	0.048	0.10	0.34	0.68	0.91	0.99
	F	0.043	0.09	0.28	0.58	0.83	0.95
2	F*	0.043	0.10	0.30	0.61	0.87	0.97
	F	0.043	0.09	0.27	0.57	0.83	0.96
4	F*	0.041	0.11	0.33	0.66	0.90	0.98
	F	0.043	0.09	0.28	0.57	0.83	0.96
6	F*	0.042	0.11	0.34	0.68	0.92	0.99
	F	0.043	0.09	0.28	0.57	0.83	0.96

It can be seen that the F* test has higher power. We will show in Chapter 8 that F* has both criterion robustness as well as efficiency robustness. Incidentally, the values given in Table 6.1 are obtained by adding and subtracting a constant d to the observations in the first and the second blocks, respectively.

It is not only for the GL(b, σ) family that F* provides a more powerful test but it does so for other families of non-normal distributions. Consider, for example, the family of Weibull distributions which has support on the real line IR: (0, ∞).

6.6 THE WEIBULL FAMILY

Suppose now that the errors in (6.2.1) are iid and have one of the distributions in the Weibull family W(p, σ), p > 1 is assumed known:

$$W(p, \sigma) : f(e) = \frac{p}{\sigma^p} e^{p-1} \exp\left\{-\left(\frac{e}{\sigma}\right)^p\right\}, \; 0 < e < \infty. \tag{6.6.1}$$

If p is unknown, its value can be determined by using one of the techniques given in Chapters 9 and 11. Incidentally, W(p, σ) has many applications in the areas of engineering sciences (Meeker and Escobar, 1988), biomedical sciences (Johnson and Johnson, 1979), seismology (Akkaya, 1995), air quality determination (Murani et al., 1986), etc. Like GL(b, σ), the Weibull W(p, σ) represents unimodal distributions.

The ML estimators of the parameters in the model (6.2.1) are intractable. However, the MML estimators can be obtained exactly along the same lines as in Section 6.4. They are the same as in (6.4.6)-(6.4.9) with m(b + 1) replaced by $m = \sum_{j=1}^{n} \delta_j$;

$$\delta_j = (p-1)\beta_{j0} + p\beta_j, \; \beta_j = (p-1)t_{(j)}^{\,p-2}, \; \beta_{j0} = t_{(j)}^{\,-2} \; (p > 1), \tag{6.6.2}$$

$$t_{(j)} = \left[-\ln\left(1 - \frac{j}{n+1}\right)\right]^{1/p}, \; 1 \le j \le n, \tag{6.6.3}$$

and the expressions for $\hat{\mu}_i$, B_i and C_i are

$$\hat{\mu}_{i.} = (1/m)\sum_{j=1}^{n} \delta_j y_{i(j)},\ B_i = \sum_{j=1}^{n} \Delta_j (y_{i(j)} - \hat{\mu}_{i.}),$$

$$C_i = \sum_{j=1}^{n} \delta_j (y_{i(j)} - \hat{\mu}_{i.})^2 = \sum_{j=1}^{n} \delta_j y_{i(j)}{}^2 - m\hat{\mu}_{i.}{}^2; \tag{6.6.4}$$

$$\Delta_j = (p-1)\alpha_{j0} - p\alpha_j,\ \alpha_j = (2-p)t_{(j)}{}^{p-1},\ \alpha_{j0} = 2t_{(j)}{}^{-1}.$$

Given in Table 6.2 are the type I errors of the tests based on F* and the normal-theory F. It can be seen that the F* test has higher power. Another important feature of the F* test is that for $1 \le p \le 2$ in the Weibull $W(p, \sigma)$, it has considerably smaller type I error than the F test and has higher power unless d is small. Small values of d, however, are not of much practical interest since the power is only marginally bigger than the type I error. We will show in Chapter 8 that this interesting property (smaller type I error and higher power) is true in several other situations, e.g., outlier and contamination models.

Table 6.2: Values of the power for the Weibull $W(p, \sigma)$; $k = 4$, $n = 10$.

p	d =	0.0	0.2	0.4	0.6	0.8	1.0
1.5	F*	0.019	0.06	0.28	0.71	0.87	1.00
	F	0.043	0.09	0.27	0.57	0.71	0.96
2	F*	0.032	0.08	0.27	0.60	0.88	0.98
	F	0.045	0.09	0.27	0.56	0.83	0.96
4	F*	0.046	0.09	0.27	0.57	0.84	0.97
	F	0.044	0.09	0.26	0.56	0.83	0.97
6	F*	0.044	0.09	0.27	0.57	0.84	0.97
	F	0.047	0.09	0.26	0.55	0.82	0.96

The values of the power above are obtained by adding and subtracting a constant d from the observations in the first and the second blocks, respectively. We show in Chapter 8 that F* has excellent robustness properties due to the umbrella or half-umbrella ordering of the β_j (or δ_j) cofficients which are the weights given to the order statistics $y_{i(j)}$ $(l \le j \le n)$ in each block.

6.7 TWO-WAY CLASSIFICATION AND INTERACTION

Consider the two-way classification fixed-effects model

$$y_{ijl} = \mu + \gamma_i + \delta_j + \tau_{ij} + e_{ijl} \quad (1 \le i \le k,\ 1 \le j \le c,\ 1 \le l \le n) \tag{6.7.1}$$

with usual interpretation of the parameters μ, γ_i, δ_j and τ_{ij} (the interaction between the i^{th} block and the j^{th} column) ; e_{ijl} are iid random errors. Without loss of generality we assume that

$$\Sigma_i\gamma_i = \Sigma_j\delta_j = \Sigma_i\tau_{ij} = \Sigma_j\tau_{ij} = 0. \tag{6.7.2}$$

It is of great practical interest to test the null hypotheses

$$H_{01} : \gamma_i = 0 \text{ (for all i)},\ H_{02} : \delta_j = 0 \text{ (for all j) and} \tag{6.7.3}$$
$$H_{03} : \tau_{ij} = 0 \text{ (for all i and j)}.$$

Assuming that the random errors e_{ijl} are iid normal, the ML estimators of γ_i, δ_j, τ_{ij} and σ are, respectively,

$$\bar{y}_{i..} - \bar{y}_{...},\ \bar{y}_{.j.} - \bar{y}_{...},\ \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \quad \text{and}$$

$$s_e = \sqrt{\{\Sigma_i\Sigma_j\Sigma_l(y_{ijl} - \bar{y}_{ij.})^2/(N - kc)\}} \quad (N = nkc); \tag{6.7.4}$$

$$\bar{y}_{i..} = \Sigma_j\Sigma_l y_{ijl}/nc,\ \bar{y}_{.j.} = \Sigma_i\Sigma_l y_{ijl}/nc \text{ and } \bar{y}_{ij.} = \Sigma_l y_{ijl}/n$$

are the means of the observations in the i^{th} block, the j^{th} column and the $(i, j)^{th}$ cell, respectively.

The F statistics for testing H_{01}, H_{02} and H_{03} are, respectively,

$$F_1 = nc\sum_{i=1}^{k}(\bar{y}_{i..} - \bar{y}_{...})^2 / (k-1)s_e^2 \tag{6.7.5}$$

$$F_2 = nk\sum_{j=1}^{c}(\bar{y}_{.j.} - \bar{y}_{...})^2/(c-1)s_e^2 \tag{6.7.6}$$

and

$$F_3 = n\sum_{i=1}^{k}\sum_{j=1}^{c}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2/(k-1)(c-1)s_e^2. \tag{6.7.7}$$

The null distributions of F_1, F_2 and F_3 are central F with degrees of freedom (ν_1, ν_4), (ν_2, ν_4) and (ν_3, ν_4), respectively:

$$\nu_1 = k-1, \quad \nu_2 = c-1, \quad \nu_3 = (k-1)(c-1) \quad \text{and} \quad \nu_4 = kc(n-1) = N - kc. \tag{6.7.8}$$

Large values of F_1, F_2 and F_3 lead to the rejection of H_{01}, H_{02} and H_{03}, respectively.

If the null hypotheses in (6.7.3) are not true, then the distributions of F_1, F_2 and F_3 are noncentral F with degrees of freedom given in (6.7.8) and noncentrality parameters

$$nc\sum_{i=1}^{k}(\gamma_i/\sigma)^2,\ nk\sum_{j=1}^{k}(\delta_j/\sigma)^2 \quad \text{and} \quad n\Sigma_i\Sigma_j(\tau_{ij}/\sigma)^2, \tag{6.7.9}$$

respectively.

6.8 EFFECTS UNDER NON-NORMALITY

Assume that the random errors e_{ijl} in (6.7.1) have the Generalized Logistic distribution GL(b, σ). Here, the likelihood function is (N = nkc)

$$L \propto \left(\frac{1}{\sigma}\right)^N \prod_{i=1}^{k}\prod_{j=1}^{c}\prod_{l=1}^{n}[\exp(-z_{ijl})/\{1+\exp(-z_{ijl})\}^{b+1}]; \tag{6.8.1}$$

$$z_{ijl} = (y_{ijl} - \mu - \gamma_i - \delta_j - \tau_{ij})/\sigma.$$

Let

$$y_{ij(1)} \le y_{ij(2)} \le \ldots \le y_{ij(n)} \quad (1 \le i \le k,\ 1 \le j \le c) \tag{6.8.2}$$

be the order statistics of the n observations in the (i, j)th cell. Writing

$$z_{ij(l)} = (y_{ij(l)} - \mu - \gamma_i - \delta_j - \tau_{ij})/\sigma \quad \text{and} \quad g(z) = e^{-z}/(1+e^{-z}), \tag{6.8.3}$$

the likelihood equations are

$$\begin{aligned}
\frac{\partial \ln L}{\partial \mu} &= \frac{N}{\sigma} - \frac{(b+1)}{\sigma}\Sigma_i\Sigma_j\Sigma_l\, g(z_{ij(l)}) = 0\\
\frac{\partial \ln L}{\partial \gamma_i} &= \frac{nc}{\sigma} - \frac{(b+1)}{\sigma}\Sigma_j\Sigma_l\, g(z_{ij(l)}) = 0\\
\frac{\partial \ln L}{\partial \delta_j} &= \frac{nk}{\sigma} - \frac{(b+1)}{\sigma}\Sigma_i\Sigma_l\, g(z_{ij(l)}) = 0\\
\frac{\partial \ln L}{\partial \tau_{ij}} &= \frac{n}{\sigma} - \frac{(b+1)}{\sigma}\Sigma_l\, g(z_{ij(l)}) = 0 \quad \text{and}
\end{aligned} \tag{6.8.4}$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma} \Sigma_i\Sigma_j\Sigma_l z_{ij(l)} - \frac{(b+1)}{\sigma} \Sigma_i\Sigma_j\Sigma_l z_{ij(l)} g(z_{ij(l)}) = 0.$$

Clearly, there are too many equations to solve by iteration which is almost an impossible task. The ML estimators are, therefore, not available. The LS estimators are inefficient as we have shown in Section 6, and their relative efficiencies decrease with increasing n.

To obtain the MML estimators which are asymptotically fully efficient, and are highly efficient for small sample sizes, we linearize the equations (6.8.4) by replacing $g(z_{ij(l)})$ by

$$g(z_{ij(l)}) \cong \alpha_l - \beta_l z_{ij(l)}, \quad 1 \le l \le n, \tag{6.8.5}$$

and obtain the modified likelihood equations by incorporating (6.8.5) in (6.8.4). The coefficients α_l and β_l are given in (6.4.4) with $t_{(j)}$ replaced by $t_{(l)} = E\{z_{ij(l)}\}$. The solutions of the modified likelihood equations are the following MML estimators:

$$\hat{\mu} = \hat{\mu}_{...} - (\Delta/m)\hat{\sigma} \tag{6.8.6}$$

$$\hat{\gamma}_i = \hat{\mu}_{i..} - \hat{\mu}_{...}, \quad \hat{\delta}_j = \hat{\mu}_{.j.} - \hat{\mu}_{...}, \quad \hat{\tau}_{ij} = \hat{\mu}_{ij.} - \hat{\mu}_{i..} - \hat{\mu}_{.j.} + \hat{\mu}_{...},$$

and

$$\hat{\sigma} = \{-B + \sqrt{B^2 + 4NC}\}/2\sqrt{N(N-kc)}; \tag{6.8.7}$$

$$\Delta = \sum_{l=1}^{n} \Delta_l, \Delta_l = \alpha_l - (b+1)^{-1}, \quad m = \sum_{l=1}^{n} \beta_l, B = \sum_{i=1}^{k}\sum_{j=1}^{c} B_{ij},$$

$$C = \sum_{i=1}^{k}\sum_{j=1}^{c} C_{ij};\ B_{ij} = (b+1) \sum_{l=1}^{n} \Delta_l (y_{ij(l)} - \hat{\mu}_{ij.}),$$

$$C_{ij} = (b+1) \sum_{l=1}^{n} \beta_l (y_{ij(l)} - \hat{\mu}_{ij.})^2 = (b+1)\left(\sum_{l=1}^{n} \beta_l y_{ij(l)}^2 - m\hat{\mu}_{ij.}^{\ 2}\right);$$

$$\hat{\mu}_{...} = (1/kcm)\, \Sigma_i\Sigma_j\Sigma_l\, \beta_l y_{ij(l)},\ \hat{\mu}_{i..} = (1/cm)\Sigma_j\Sigma_l\beta_l y_{ij(l)},$$

$$\hat{\mu}_{.j.} = (1/km)\Sigma_i\Sigma_l\, \beta_l y_{ij(l)} \quad \text{and} \quad \hat{\mu}_{ij.} = (1/m)\Sigma_l\, \beta_l y_{ij(l)}. \tag{6.8.8}$$

Note that $\hat{\sigma}$ is always real and positive since $\beta_l > 0$ $(1 \le l \le n)$.

The pivotal estimator is $\hat{\mu}_{ij.}$ which has the same efficiency and distributional properties as $\hat{\mu}_{i.}$ in (6.4.9). In particular, we have the following properties of the MML estimators.

Efficiency: Since the estimators $\hat{\gamma}_i$, $\hat{\delta}_j$ and $\hat{\tau}_{ij}$ are linear contrasts of $\hat{\mu}_{ij.}$, they are unbiased and uncorrelated (asymptotically independent) with $\hat{\sigma}^2$. In fact, they are the BAN (best asymptotically normal) estimators. The estimator $\hat{\sigma}^2$ is asymptotically fully efficient and is distributed as a multiple of chi-square (Chapter 2).

Minimum variance bounds: The minimum variance bounds for estimating γ_i, δ_j and τ_{ij} are $(k-1)\sigma^2/(b+1)M$, $(c-1)\sigma^2/(b+1)M$ and $(k-1)(c-1)\sigma^2/(b+1)M$, respectively; $M = mkc$. The true variances of the estimators $\hat{\gamma}_i$, $\hat{\delta}_j$ and $\hat{\tau}_{ij}$ are very close to these bounds even for small sample sizes as in Section 6.4. This clearly establishes these estimators as highly efficient. Realize that the MVB estimators do not exist for the GL(b, σ) family and, therefore, all estimators will have their variances bigger than the minimum variance bounds.

6.9 VARIANCE RATIO STATISTICS

To test the null hypothesis H_{0i} (i = 1, 2, 3) in (6.7.3), we define the variance ratio statistics

$$F_1^* = \frac{mc(b+1)\sum_{i=1}^{k} \hat{\gamma}_i^2}{(k-1)\hat{\sigma}^2}, \quad F_2^* = \frac{mk(b+1)\sum_{j=1}^{c} \hat{\delta}_j^2}{(c-1)\hat{\sigma}^2} \tag{6.9.1}$$

and
$$F_3^* = \frac{m(b+1)\sum_{i=1}^{k}\sum_{j=1}^{c} \hat{\tau}_{ij}^2}{(k-1)(c-1)\hat{\sigma}^2}. \tag{6.9.2}$$

Large values of F_i^* lead to the rejection of H_{0i} (i = 1, 2, 3), respectively. For large n, the null distributions of F_1^*, F_2^* and F_3^* are central F with degrees of freedom (ν_1, ν_4),(ν_2, ν_4) and (ν_3, ν_4), respectively; ν_i are given in (6.7.8).

The central F distributions provide remarkably accurate approximations to the percentage points even for small n. For example, we have the simulated values of the probabilities

$$\text{Prob}\{F_i \geq F_{0.05}(\nu_i, \nu_4) \mid H_{0i}\} \quad (i = 1, 2, 3) \tag{6.9.3}$$

given in Table 6.3, reproduced from Şenoğlu and Tiku (2001):

Table 6.3: Values of the probabilities ; k = 3, c = 4

		b = 0.5	1	2	3	4	6
n = 4	F_1^*	0.049	0.045	0.045	0.043	0.041	0.040
	F_2^*	0.050	0.046	0.047	0.046	0.044	0.042
	F_3^*	0.041	0.041	0.045	0.044	0.042	0.041
n = 5	F_1^*	0.052	0.047	0.047	0.045	0.044	0.042
	F_2^*	0.052	0.050	0.050	0.050	0.047	0.045
	F_3^*	0.048	0.045	0.048	0.047	0.047	0.046
n = 6	F_1^*	0.054	0.050	0.050	0.048	0.047	0.044
	F_2^*	0.054	0.054	0.054	0.051	0.050	0.048
	F_3^*	0.053	0.049	0.052	0.051	0.050	0.049

Power properties: For large n, the non-null distributions of F_1^*, F_2^* and F_3^* are noncentral F with degrees of freedom given in (6.7.8) and noncentrality parameters

$$\lambda_1^2 = mc(b+1)\sum_{i=1}^{k}(\gamma_i/\sigma)^2, \quad \lambda_2^2 = mk(b+1)\sum_{j=1}^{k}(\delta_j/\sigma)^2$$

and
$$\lambda_3^2 = m(b+1)\sum_{i=1}^{k}\sum_{j=1}^{c}(\tau_{ij}/\sigma)^2, \tag{6.9.4}$$

respectively. The corresponding noncentrality parameters of the non-null distributions of the statistics F_1, F_2 and F_3 are for large n the same as in (6.9.4) with m(b + 1) replaced by n. Because of (6.4.12), the ratios of the noncentrality parameters of F_i^* to F_i(i = 1, 2, 3) are all equal to (b + 2)/b which is greater than 1 for all values of b. The F_i^* tests are, therefore, more powerful than the F_i tests. For small n, the F_i^* tests are more powerful than the F_i tests by the same margins as those in Table 6.1 (Şenoğlu and Tiku, 2001).

6.10 BOX-COX DATA

Consider the Box-Cox (1964, p.220) data which give the survival times x (in units of 10 hours) of 48 animals exposed to three different poisons and four different treatments. The experiment was set in a 3 × 4 factorial design with four observations x_{ijl} $(1 \leq l < 4)$ in each cell:

Poison	Treatment A	B	C	D
I	0.31	0.82	0.43	0.45
	0.45	1.10	0.45	0.71
	0.46	0.88	0.63	0.66
	0.43	0.72	0.76	0.62
II	0.36	0.92	0.44	0.56
	0.29	0.61	0.35	1.02
	0.40	0.49	0.31	0.71
	0.23	1.24	0.40	0.38
III	0.22	0.30	0.23	0.30
	0.21	0.37	0.25	0.36
	0.18	0.38	0.24	0.31
	0.23	0.29	0.22	0.33

The objective is to test for block (I, II and III) effects, column (A, B, C and D) effects and the interaction effects (block × column). As the tradition has it, assume that e_{ijl} in the model (6.7.1) are normal $N(0, \sigma^2)$. The values of the F_1, F_2 and F_3 statistics in (6.7.5)-(6.7.7) reveal that the block and columns effects are significant at 1% significance level but the interaction effects are not significant even at 5% significance level. The latter is erroneous since the data are known to have interaction (Schrader and McKean, 1977, p.889). This is apparently due to the wrongful assumption of normality.

Several authors have discussed transformations (Box and Cox, 1964) to normalize the data which also reduces the interaction effects in this case; see for example, Brown (1975). The recommendation is that an inverse transformation be applied and Y = 1/X be regarded as having a near-normal distribution. Usually, it is difficult to interpret transformations as said earlier but here Y = 1/X has a natural appeal and represents the rate of dying. Calculations with the F_1, F_2 and F_3 statistics for the transformed data reveal that the block and column effects are significant at 1% significance level but the interaction effects are not significant at 5% significance level (Brown, 1975; Schrader and McKean, 1977).

To have an idea about the underlying distribution, we constructed a Q-Q plot of the residuals

$$\tilde{e}_{ijl} = y_{ijl} - \bar{y}_{ij}, \quad \bar{y}_{ij.} = (1/n)\sum_{l=1}^{n} y_{ijl}\,; \tag{6.10.1}$$

Q-Q plots and goodness-of-fit tests are discussed in Chapters 9 and 11. It was found that the Generalized Logistic GL(b, σ) with b = 0.5 provides "close to a straight line" pattern. Also, the multi-sample goodness-of-fit test does not reject GL(0.5, σ) at 10% significance level (Chapter 9). We conclude, therefore, that GL(0.5, σ) is overall a plausible distribution for the errors. The resulting values of the F_i^*(i = 1, 2, 3) statistics are given in Table 6.4 for both the original as well as the transformed data.

Table 6.4: ANOVA of Box-Cox data based on GL(0.5, σ).

		Original data			Transformed data		
Source	df	SS	MSS	F*	SS	MSS	F*
Poison	2	0.322	0.116	30.68**	8.052	4.026	63.90**
Treatment	3	0.280	0.093	17.80**	5.091	1.697	26.93**
Interaction	6	0.089	0.015	2.83*	0.336	0.056	0.88
Error	36	0.180	0.005		2.268	0.063	
	Scale estimate = $\sqrt{0.005}$ = 0.072				Scale estimate = 0.251		

The block and column effects in the original as well as the transformed data are significant at 1% significance level. The interaction effects in the original data are significant at 5% significance level but not in the transformed data. This agrees with the findings of Schrader and McKean (1977) who used the Huber (1981) method and the nonparametric R (relative ranks) method. These methods are very computer intensive and the Huber method is suited particularly to long-tailed symmetric distributions and not to skew or short-tailed symmetric distributions (Chapters 2, 7, 8). It may also be noted that the error SS (on 36 degrees of freedom) obtained by using the R method are 1.656 and 8.28 for the original and the transformed data, respectively. The corresponding normal-theory values are 0.801 and 8.64, respectively. These values are considerably bigger than those given in Table 6.4, i.e., 0.180 and 2.268. It can, therefore, be concluded that the method based on the MML estimators with Generalized Logistic GL(0.5, σ) provides a more precise analysis of the data besides being computationally quite straightforward. Moreover, the method extends to other families of distributions and maintains high power (Tables 6.1-6.2). We will show in Chapter 8 that the variance ratio F statistics based on the MML estimators have also criterion robustness as well as efficiency robustness. Also, the method readily extends to unbalanced designs, i.e. $n_i \neq n$ in (6.2.1).

6.11 LINEAR CONTRASTS

It has to be recognized that the variance ratio F statistics provide an overall assessment of the block (column) differences. The F or F* tests given in Section 6.5, for example, give an overall assessment whether block differences exist or not. If F or F* statistics are not significantly large, that does not necessarily imply that no block (column) differences exist. It is, therefore, always advisable to construct linear contrasts to assess the block (column) effects. It suffices to consider the one-way classification model (6.2.1).

Consider the linear contrast

$$\eta = \sum_{i=1}^{k} l_i \gamma_i = \sum_{i=1}^{k} l_i \mu_i \ \ (\mu_i = \mu + \gamma_i), \quad \sum_{i=1}^{k} l_i = 0 \tag{6.11.1}$$

Without loss of generality we assume that $\sum_{i=1}^{k} l_i^2 = 1$ in which case η is called a standardized linear contrast; $l_i (1 \le i \le k)$ are constant coefficients. Two contrasts

$$\eta_1 = \sum_{i=1}^{k} l_{1i}\mu_i \quad \text{and} \quad \eta_2 = \sum_{i=1}^{k} l_{2i}\mu_i \tag{6.11.2}$$

are called orthogonal if $\sum_{i=1}^{k} l_{1i} l_{2i} = 0$. A convenient way of constructing standardized orthogonal linear contrasts is through Helmert transformation:

$$\begin{aligned} \eta_1 &= (\mu_1 - \mu_2)/\sqrt{2} \\ \eta_2 &= (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6} \\ &\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots \\ \eta_{k-1} &= (\mu_1 + \ldots\ldots + \mu_{k-1} - (k-1)\mu_k)/\sqrt{k(k-1)}; \end{aligned} \tag{6.11.3}$$

these are all the possible orthogonal contrasts, k – 1 in number. Each one of them is orthogonal to the mean vector

$$(\mu_1 + \mu_2 + \ldots\ldots + \mu_k)/k. \tag{6.11.4}$$

Replacing μ_i by $\hat{\mu}_i (1 \le i \le k)$ gives the estimator $\hat{\eta}_i$ of η_i.

If $\hat{\mu}_i$ are unbiased or have an equal amount of bias, i.e. $E(\hat{\mu}_i)$ is equal to μ_i or is equal to $\mu_i + \delta$ $(1 \le i \le k)$, then $\hat{\eta}_i$ is clearly an unbiased estimator of η_i. If $\hat{\mu}_i (1 \le i \le k)$ are independent of each other (or uncorrelated with each other), then

$$\text{var}(\hat{\eta}) = \left(\sum_{i=1}^{k} l_i^2 V_i\right)\sigma^2, \quad V(\hat{\mu}_i) = V_i\sigma^2. \tag{6.11.5}$$

Further, if $\hat{\mu}_i$ are asymptotically normal and σ^2 converges to σ as n becomes large, then the distribution of the statistic

$$T = \left(\sum_{i=1}^{k} l_i \hat{\mu}_i\right) \Big/ \hat{\sigma}\sqrt{\sum_{i=1}^{k} l_i^2 V_i} \tag{6.11.6}$$

is asymptotically normal N(0, 1). If, however, $\hat{\mu}_i$ are normally distributed and $(N - k)\hat{\sigma}^2/\sigma^2$ is distributed as chi-square and is independent of $\hat{\mu}_i (1 \le i \le k)$, then T is distributed as the Student t.

Efficiency and power: Consider the model (6.2.1) with, for example, k = 3 (three blocks). We consider the two orthogonal linear contrasts

$$\eta_1 = (\mu_1 - \mu_2)/\sqrt{2} \quad \text{and} \quad \eta_2 = (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6}.$$

If the errors e_{ij} are iid normal $N(0, \sigma^2)$, then the maximum likelihood estimators are

$$\hat{\hat{\eta}}_1 = (\bar{y}_1 - \bar{y}_2)/\sqrt{2} \quad \text{and} \quad \hat{\hat{\eta}}_2 = (\bar{y}_1 + \bar{y}_2 - 2\bar{y}_3)/\sqrt{6}, \tag{6.11.7}$$

each having variance σ^2/n which is estimated by s^2/n;

$$s^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2/k(n-1) \ (k = 3)$$

is the pooled sample variance. Under the null hypotheses

$$H_{01} : \eta_1 = 0 \quad \text{and} \quad H_{02} : \eta_2 = 0, \tag{6.11.8}$$

the null distributions of

$$t_1 = \sqrt{n}\,(\hat{\hat{\eta}}_1/s) \quad \text{and} \quad t_2 = \sqrt{n}\,(\hat{\hat{\eta}}_2/s) \tag{6.11.9}$$

are the Student t with ν degrees of freedom and the two are independently distributed; ν = 3(n – 1). Their non-null distributions are noncentral t with ν df and noncentrality parameters λ_1 and λ_2, respectively;

$$\lambda_1^2 = n\eta_1^2/\sigma^2 \quad \text{and} \quad \lambda_2^2 = n\eta_2^2/\sigma^2. \tag{6.11.10}$$

6.12 NON-NORMAL DISTRIBUTIONS

Suppose that the errors in (6.2.1) have the Generalized Logistic distribution GL(b, σ). The MML estimators of η_1 and η_2 are

$$\hat{\eta}_1 = (\hat{\mu}_{1.} - \hat{\mu}_{2.})/\sqrt{2} \quad \text{and} \quad \hat{\eta}_2 = (\hat{\mu}_{1.} + \hat{\mu}_{2.} - 2\hat{\mu}_{3}.)/\sqrt{6}; \qquad (6.12.1)$$

$\hat{\mu}_{1.}$ are given in (6.4.9); $\hat{\eta}_1$ and $\hat{\eta}_2$ are unbiased and their variances for large n are $\sigma^2/m(b + 1) \cong b\sigma^2/(b + 2)$ (Lemma 6.2). To test H_{01} and H_{02} in (6.11.8), we define the statistics

$$T_1 = \sqrt{\{m(b+1)\}}\,(\hat{\eta}_1/\hat{\sigma}) \quad \text{and} \quad T_2 = \sqrt{\{m(b+1)\}}\,(\hat{\eta}_2/\hat{\sigma}); \qquad (6.12.2)$$

$\hat{\sigma}$ is the MML estimator of σ given in (6.4.7). For large n, the null distributions of T_1 and T_2 are normal N(0, 1). Large values of $|T_1|$ and $|T_2|$ lead to the rejection of H_{01} and H_{02}, respectively. The non-null distributions are for large n normal with means

$$\sqrt{\{(b+2)/b\}\lambda_1^2} \quad \text{and} \quad \sqrt{\{(b+2)/b\}\lambda_2^2}; \qquad (6.12.3)$$

respectively, and variances both equal to 1. Since (b + 2)/b is always greater than 1, the T_i tests are more powerful than the $t_i (i = 1, 2)$ tests. For small n, the null and non-null distributions of T_1 and T_2 are referred to central and noncentral t with ν df and noncentrality parameters given in (6.12.13).

Given in Table 6.5 are the values of the type I error and power of the T_2 and t_2 tests; the differences in the values for T_1 and t_1 tests are essentially the same. It can be seen that the T test is considerably more powerful than the t test. Not only is the type I error of T smaller than that of t, but its power is higher. This is an interesting phenomenon. See also Chapter 8 for similar results.

Table 6.5: Type I error and power for testing linear contrasts; k = 3 and $\tau = \tau_1 = -\tau_2/2 = \tau_3$.

$b_1 = b_2 = b_3 = 0.5$	n = 10		15		20	
τ	T	t	T	t	T	t
0.0	0.041	0.054	0.041	0.046	0.045	0.049
0.2	0.09	0.10	0.12	0.12	0.16	0.14
0.4	0.24	0.22	0.35	0.31	0.46	0.40
0.6	0.46	0.42	0.68	0.59	0.80	0.71
0.8	0.71	0.65	0.89	0.83	0.96	0.92
1.0	0.89	0.83	0.98	0.95	1.00	0.99
$b_1 = b_2 = b_3 = 4.0$	n = 10		15		20	
τ	T	t	T	t	T	t
0.0	0.039	0.051	0.037	0.047	0.034	0.051
0.2	0.18	0.20	0.28	0.28	0.37	0.36
0.4	0.61	0.59	0.81	0.76	0.91	0.87
0.6	0.93	0.90	0.99	0.97	1.00	0.99
0.8	1.00	0.98	1.00	1.00	1.00	1.00
1.0	1.00	1.00	1.00	1.00	1.00	1.00

6.13 NON-IDENTICAL ERROR DISTRIBUTIONS

In the model (6.2.1), and in (6.7.1), we assumed that the random errors e_{ij} have the same distribution. This may not always be true in practice. Let us revisit the Box-Cox data considered in Section 6.10. Consider the three blocks each consisting of 16 observations, receiving Poison I, II and III. Denote these observations by e_{il} (i = 1, 2, 3 ; l = 1, 2,......, 16). Let $\bar{y}_{i.}$(i = 1, 2, 3) be the means of the three blocks. The residuals $\tilde{e}_{il} = y_{il} - \bar{y}_{i.}$ $(1 \le l \le 16)$ for the three blocks when arranged in ascending order of magnitude are

$\tilde{e}_{1(1)}$:	– 0.546	– 0.312	– 0.281	– 0.275	– 0.264	– 0.254	– 0.174
	– 0.161	– 0.076	– 0.027	0.056	0.225	0.359	0.462
	0.532	0.738					
$\tilde{e}_{2(1)}$:	– 0.768	– 0.721	– 0.586	– 0.490	– 0.441	– 0.306	– 0.293
	– 0.213	0.084	0.143	0.179	0.245	0.511	0.647
	0.930	1.074					
$\tilde{e}_{3(1)}$:	– 0.454	– 0.397	– 0.326	– 0.314	– 0.264	– 0.257	– 0.098
	– 0.061	– 0.04	0.082	0.134	0.241	0.280	0.304
	0.419	0.752					

The Q-Q plots, i.e., the ordered residuals plotted against the population quantiles of the $GL(b, \sigma)$ family, namely,

$$t_{i(1)} = t_l = -\ln (q_l^{-1/b_i} - 1, \quad q_l = l/17, \quad 1 \le l \le 16, \tag{6.13.1}$$

are discussed in Chapter 11. It can be seen that 'close to a straight line' patterns are achieved by taking $b_1 = 2.0$, $b_2 = 1.0$ and $b_3 = 6.0$ in the three blocks. This makes it necessary to be able to find efficient estimators of linear contrasts $\sum_{i=1}^{k} l_i \mu_i$ when the error distributions from block to block are not necessarily identical. We also need efficient and robust hypothesis testing procedures to test assumed values of these contrasts. It may be noted that the normal-theory F statistics are non-robust if the distributions in some of the blocks are positively skew and are negatively skew in others. This has been illustrated in Chapter 1; see particularly equations (1.7.8)-(1.7.9). See also Şenoğlu ad Tiku (2002).

In the model (6.2.1), suppose the errors e_{ij} $(1 \le j \le k)$ have the Generalized Logistic distribution $GL(b_i, \sigma)$. Since $b_i \ne b$ $(1 \le i \le k)$, the variances are also unequal in the blocks. Let

$$y_{i(1)} \le y_{i(2)} \le \ldots\ldots \le y_{i(n)} \quad (1 \le i \le k) \tag{6.13.2}$$

be the order statistics of the n random observations in the ith block. The likelihood equations when written in terms of the order statistics are

$$\frac{\partial \ln L}{\partial \mu} = \frac{N}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n}(b_i + 1)\, g(z_{i(j)}) = 0 \tag{6.13.3}$$

$$\frac{\partial \ln L}{\partial \gamma_i} = \frac{n}{\sigma} - \frac{(b_i + 1)}{\sigma}\sum_{j=1}^{n} g(z_{i(j)}) = 0 \tag{6.13.4}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n} z_{i(j)} - \frac{1}{\sigma}\sum_{i=1}^{k}\sum_{j=1}^{n}(b_i + 1) z_{i(j)} g(z_{i(j)}) = 0; \tag{6.13.5}$$

$$z_{i(j)} = (y_{i(j)} - \mu - \gamma_i)/\sigma \qquad \text{and} \qquad g(z) = e^{-z}/(1 + e^{-z}).$$

Solving these equations is enormously more difficult than those with $b_i (1 \le i \le k)$ all equal. Consequently, the ML estimators are even more elusive than before.

Modified likelihood: To obtain the MML estimators, we first formulate the modified likelihood equations by replacing $g(z_{i(j)})$ by the linear functional

$$g(z_{i(j)}) \cong \alpha_{ij} - \beta_{ij} z_{i(j)}, \qquad 1 \le j \le n; \tag{6.13.6}$$

the values of α_{ij} and β_{ij} $(1 \le j \le n)$ are given in (6.4.3) with $t_j = t_{i(j)}$. If $b_i = b$, then $\alpha_{ij} = \alpha_i$ and $\beta_{ij} = \beta_{ij}$ as in (6.4.4).

The modified likelihood equations are asymptotically equivalent to the likelihood equations and their solutions are the following MML estimators, assuming $\sum_{i=1}^{k} (b_i + 1) m_i \gamma_i = 0$ without any loss of generality (Şenoğlu and Tiku, 2002):

$$\hat{\mu} = \hat{\mu}_{..} - \Delta_1 \hat{\sigma}, \quad m_i = \sum_{j=1}^{n} \beta_{ij} \tag{6.13.7}$$

$$\hat{\gamma}_i = \hat{\mu}_{i.} - \hat{\mu}_{..} + (\Delta_1 - (\Delta_{i.}/m_i))\hat{\sigma} \quad (1 \le i \le k), \tag{6.13.8}$$

and

$$\hat{\sigma} = \{-B + \sqrt{B^2 + 4NC}\}/2\sqrt{N(N-k)}, \quad N = nk; \tag{6.13.9}$$

$$\begin{aligned} \hat{\mu}_{..} &= (1/M)\, \Sigma_i (b_i + 1) m_i \hat{\mu}_i, \; M = \Sigma_i (b_i + 1) m_i, \\ \Delta_1 &= (1/M)\, \{\Sigma_i \Sigma_j (b_i + 1\} \alpha_{ij} - N\}; \\ \hat{\mu}_{i.} &= (1/m_i) \Sigma_j \beta_{ij} y_{i(j)}, \quad \Delta_{i.} = \Sigma_j \Delta_{ij}, \quad \Delta_{ij} = \alpha_{ij} - (b_i + 1)^{-1}, \\ B &= \Sigma_i \Sigma_j (b_i + 1) \Delta_{ij} (y_{i(j)} - \hat{\mu}_{i.}) \end{aligned} \tag{6.13.10}$$

and

$$C = \Sigma_i \Sigma_j (b_i + 1) \beta_{ij} (y_{i(j)} - \hat{\mu}_{i.})^2 = \Sigma_i (b_i + 1)(\Sigma_j \beta_{ij} y_{i(j)}{}^2 - m_i \hat{\mu}_{i.}{}^2). \tag{6.13.11}$$

Since the coefficients β_{ij} are all positive, $\hat{\sigma}$ is real and positive. If $b_i (1 \le i \le k)$ are all equal and equal to b, the estimators (6.13.7)-(6.13.11) reduce to (6.8.6)-(6.8.8). The estimators above are asymptotically fully efficient since the MML estimators are equivalent (asymptotically) to the ML estimators (Chapter 2).

6.14 LINEAR CONTRAST WITH NON-IDENTICAL DISTRIBUTIONS

Consider the estimation of the linear contrast

$$\theta = \sum_{i=1}^{k} l_i \gamma_i = \sum_{i=1}^{k} l_i \mu_i, \quad \mu_i = \mu + \gamma_i (1 \le i \le k); \quad \sum_{i=1}^{k} l_i = 0. \tag{6.14.1}$$

The MML estimator of μ_i is

$$\hat{\mu}_i = \hat{\mu} + \hat{\gamma}_i = \hat{\mu}_{i.} - (\Delta_{i.}/m_i)\hat{\sigma} \quad (1 \le i \le k), \tag{6.14.2}$$

and the MML estimator of θ is

$$\hat{\theta} = \sum_{i=1}^{k} l_i \hat{\mu}_i = \sum_{i=1}^{k} l_i (\hat{\mu}_{i.} - (\Delta_{i.}/m_i)\hat{\sigma}). \tag{6.14.3}$$

Realize that $\hat{\mu}_{i.}$ $(1 \le i \le k)$, are independent of each other since they are calculated from independent samples. To derive the expression for the variance of $\hat{\theta}$, we have the following result.

Asymptotic covariance matrix: The asymptotic covariance matrix of $\hat{\mu}_i$ and $\hat{\sigma}$ is given by $I^{-1}(\mu_i, \sigma)$, where I is the Fisher Information matrix consisting of the elements $-E(\partial^2 \ln L/\partial\mu_i^2)$, $-E(\partial^2 \ln L/\partial\mu_i\partial\sigma)$ and $-E(\partial^2 \ln L/\partial\sigma^2)$ and is given by $I(\mu_i, \sigma)$

$$= \frac{n}{\sigma^2}\begin{bmatrix} \dfrac{b_i}{(b_i+2)} & \dfrac{b_i}{(b_i+2)}(\psi(b_i+1)-\psi(2)) \\ \dfrac{b_i}{(b_i+2)}(\psi(b_i+1)-\psi(2)) & k+\sum_{i=1}^{k}\dfrac{b_i}{(b_i+2)}([\psi'(b_i+1)+\psi'(2)]+(\psi(b_i+1)-\psi(2)]^2) \end{bmatrix}$$

(6.14.4)

The values of the functions $\psi(b)$ and $\psi'(b)$ are given in Appendix 2D (Chapter 2). It is easy to invert I and we will write

$$\mathrm{Cov}(\hat{\mu}_i, \hat{\sigma}) = I^{-1} = \frac{\sigma^2}{n}\begin{bmatrix} v_{i1} & v_{i2} \\ v_{2i} & v_{22} \end{bmatrix} \quad (1 \le i \le k). \tag{6.14.5}$$

Realize that v_{22} is the same constant for all i = 1, 2,......, k. The covariance $\mathrm{Cov}(\hat{\mu}_i, \hat{\sigma})$ and the variance $V(\hat{\sigma})$ are of order $O(N^{-1})$. For large n, $\hat{\theta}$ is an unbiased estimator of θ (follows from Lemma 6.1). Its variance to order $O(N^{-1})$ is

$$V(\hat{\theta}) \cong \sum_{i=1}^{k} l_i^2\, V(\hat{\mu}_{i.}) = (\sigma^2/n)\sum_{i=1}^{k} l_i^2 v_{i1}\,. \tag{6.14.6}$$

Given below are the values of (a) the bias $(E(\hat{\theta}) - \theta)/\sigma$, (b) the variance $V(\hat{\theta})/\sigma^2$, and (c) the (asymptotic variance)/σ^2 calculated from (6.14.6); k = 3, $l_1 = 1/\sqrt{6}$, $l_2 = -2/\sqrt{6}$ and $l_3 = 1/\sqrt{6}$ for illustration:

		$(b_1 = 0.5, b_2 = 1.0, b_3 = 4.0)$	$(b_1 = 0.2, b_2 = 1.0, b_3 = 6.0)$
n = 10	(a)	– 0.0044	– 0.059
	(b)	0.326	0.449
	(c)	0.313	0.415
n = 20	(a)	0.008	– 0.014
	(b)	0.158	0.214
	(c)	0.156	0.207

The bias is negligible and the closeness of the values (b) and (c) implies that $\hat{\theta}$ is highly efficient, as expected.

Hypothesis testing: To test the null hypothesis H_0: $\sum_{i=1}^{k} l_i\gamma_i = \sum_{i=1}^{k} l_i\mu_i = 0$, we define the statistic

$$T = \sqrt{n}\left(\sum_{i=1}^{k} l_i\hat{\mu}_i\right)\Bigg/\hat{\sigma}\sqrt{\sum_{i=1}^{k} l_i^2 v_i}\,, \quad v_i = v_{i1}\,, \tag{6.14.7}$$

v_{i1} are from (6.14.5). The statistic T is location and scale invariant. Large values of $|T|$ lead to the rejection of H_0. Since $\hat{\sigma}$ converges to σ as n becomes large, the null distribution of T is asymptotically normal N(0, 1). This follows from Lemma 6.1. The asymptotic power function of the $|T|$ test (with type I error α) is

$$\text{Power} \cong \text{Prob}\,(|Z| \ge z_{\alpha/2} - |\lambda|) \tag{6.14.8}$$

where Z is a standard normal variate and

$$\lambda^2 = n\left(\sum_{i=1}^{k} l_i\mu_i\right)^2 \Bigg/ \left(\sigma^2 \sum_{i=1}^{k} l_i^{\,2}\, \nu_i\right) \qquad (6.14.9)$$

is the noncentrality parameter. The normal distribution gives accurate approximations even for small n. For example, we have the following simulated values of the probability Prob ($| T | \ge z_{\alpha/2} \mid H_0$ |), presumed type I error being 0.050; k = 3 and $l_1 = 1$, $l_2 = -2$ and $l_3 = 1$:

($b_1 = 0.5$, $b_2 = 1.0$, $b_3 = 4.0$)			($b_1 = 0.2$, $b_2 = 1.0$, $b_3 = 6.0$)		
n = 10	15	20	10	15	20
0.055	0.053	0.051	0.057	0.053	0.054

Comment: The normal theory statistic

$$G = \sqrt{n}\left(\sum_{i=1}^{k} l_i \bar{y}_i\right) \Bigg/ s_e\sqrt{\sum_{i=1}^{k} l_i^{\,2}} \qquad (6.14.10)$$

should not be used in this situation. It has unacceptably high type I error (Chapter 1), usually much larger than 0.065 (Şenoğlu and Tiku, 2002).

6.15 NORMAL THEORY TEST WITH NON-IDENTICAL BLOCKS

Assume that in the model (6.2.1), e_{ij} ($1 \le j \le n$) are iid normal with mean and variance exactly equal to the mean and variance of the Generalized Logistic GL(b_i, σ). The ML (maximum likelihood) estimators (equivalently, the weighted LS estimators) of $\mu_i = \mu + \gamma_i$ ($1 \le j \le n$) and σ^2 corrected for bias are

$$\tilde{\mu}_i = \bar{y}_i - \{\psi(b_i) - \psi(1)\}s \quad \text{and} \quad \tilde{\sigma}^2 = s^2 = \sum_{i=1}^{k} w_i s_i^{\,2}/k\,; \qquad (6.15.1)$$

$$\bar{y}_i = \sum_{j=1}^{n} y_{ij}/n\,, \quad s_i^{\,2} = \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2/(n-1), \quad w_i = 1/V_i\ (1 \le i \le k),$$

$$V_i = \psi'(b_i) + \psi'(l)$$

Since $\bar{y}_i$ and s^2 are independently distributed, and var(s) = $E(s^2) - [E(s)]^2 \cong 0$ for large N,

$$\text{var}(\tilde{\mu}_i) \cong V_i(\sigma^2/n), \quad 1 \le i \le k.$$

To test H_0, therefore, the normal-theory statistic is

$$t = \frac{\sqrt{n}\left\{\sum_{i=1}^{k} l_i\,\bar{y}_i - \left(\sum_{i=1}^{k} l_i\psi(b_i)\right)s\right\}}{s\sqrt{\sum_{i=1}^{k} l_i^{\,2} V_i}}\,, \quad \sum_{i=1}^{k} l_i = 0. \qquad (6.15.2)$$

The null distribution of t is asymptotically normal N(0, 1). Large values of | t | lead to the rejection of H_0. The asymptotic power function of the | t | test is (with type I error α)

$$\text{Power} \cong \text{Prob}\,(| Z | \ge z_{\alpha/2} - | \delta |). \qquad (6.15.3)$$

Here, the noncentrality parameter is

$$\delta^2 = n\left(\sum_{i=1}^{k} l_i\mu_i\right)^2 \Bigg/ \left(\sigma^2 \sum_{i=1}^{k} l_i^{\,2} V_i\right). \tag{6.15.4}$$

Calculations show that $V_i \geq v_i$ $(1 \leq i \leq c)$. Hence, $\delta^2 \leq \lambda^2$ and the | t | test is asymptotically less powerful than the | T | test.

Şenoğlu and Tiku (2002) simulated the type I errors and power of the two tests. They showed that the normal distributions provide accurate approximations to the null distributions of T and t, but the | T | test has considerably higher power than the | t | test. We reproduce their values in Table 6.6. It can be seen that the | T | test is clearly superior to the | t | test. We will show in Chapter 8 that the | T | test is also robust but not the | t | test.

The percentage points for T and t are, respectively, 2.033 and 1.911 for n = 10, 2.014 and 1.905 for n = 15, 2.004 and 1.891 for n = 20. They were obtained by simulation in order that both the tests have the same type I error 0.050. These percentage points are, of course, very close to those obtained from normal approximations above. For an accurate comparison of the power, it is advisable to have the same type I error for the two tests. That was the motivation for using the simulated percentage points.

Table 6.6: Power of | T | and | t |, for testing a linear contrast; $k = 5$ and $\mu_1 = \mu_2 = -\mu_3 = \mu_4 = -\mu_5/2 = \tau$.

$(b_1 = 0.2, b_2 = 0.5, b_3 = 1.0, b_4 = 4.0, b_5 = 6.0)$						
	n =	10		15		20
τ	T	t	T	T	T	T
0.0	0.050	0.050	0.050	0.050	0.050	0.050
0.2	0.10	0.10	0.17	0.13	0.22	0.17
0.4	0.38	0.31	0.57	0.35	0.72	0.56
0.6	0.74	0.62	0.92	0.80	0.97	0.90
0.8	0.94	0.86	0.99	0.96	1.00	0.99
1.0	0.99	0.97	1.00	1.00	1.00	1.00

Numerical example: Consider the Box-Cox data given in Section 6.10 representing the survival times. As mentioned earlier, the GL(b, σ) family with b = 2, 1 and 6 provide plausible models for the reciprocal of the observations in the three blocks, respectively. We consider the two orthogonal linear contrasts of interest, namely,

$$\theta_1 = (\mu_1 - \mu_2)/\sqrt{2} \quad \text{and} \quad \theta_2 = (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6}. \tag{6.15.5}$$

The three estimators of θ_1 and θ_2, based on the MML estimators (6.13.7)-(6.13.10), the weighted LS estimators (6.15.1), and the classical (sample mean and sample variance) estimators of location and scale, and the standard errors of the estimates are given below. Also given are the values of the corresponding statistics T, t and G:

Estimator	Estimate	Standard error	Statistic
$\hat{\theta}_1 = (\hat{\mu}_1 - \hat{\mu}_2)/\sqrt{2}$	– 0.017	± 0.050	| T | = 0.340
$\tilde{\theta}_1 = (\tilde{\mu}_1 - \tilde{\mu}_2)/\sqrt{2}$	– 0.037	± 0.052	| t | = 0.711
$\bar{\theta}_1 = (\bar{y}_1 - \bar{y}_2)/\sqrt{2}$	0.052	± 0.063	| G | = 0.825

Similarly, we have the following for θ_2:

$\hat{\theta}_2$	0.361	± 0.045	$\lvert T \rvert = 8.022$
$\tilde{\theta}_2$	0.430	± 0.046	$\lvert t \rvert = 9.348$
$\bar{\theta}_2$	0.249	± 0.052	$\lvert G \rvert = 4.788$

The three tests are in agreement in not rejecting $\theta_1 = 0$ but rejecting $\theta_2 = 0$. The conclusion, therefore, is that the first two block effects are not different from one another but are different from the third block.

To explore whether the column-effects are different from one another, we first constructed Q-Q plots of the twelve residuals $e_{ijl} = y_{ijl} - \bar{y}_{ij.}$ $(i = 1, 2, 3; l = 1, 2, 3, 4)$ in each of the four columns. The plausible models for the four columns turned out to be GL(b, σ) with b = 1, 6, 6 and 1, respectively. We then considered the three orthogonal linear contrasts

$$\eta_1 = (\mu_1 - \mu_2)/\sqrt{2}, \quad \eta_2 = (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{6}$$

and

$$\eta_3 = (\mu_1 + \mu_2 + \mu_3 - 3\mu_4)/\sqrt{12}. \tag{6.15.6}$$

Only the estimate of η_1 turned out to be significantly different from zero at 5% significance level, with $\lvert T \rvert = 5.810$, $\lvert t \rvert = 6.204$ and $\lvert G \rvert = 2.736$, the estimates of η_1 being

$$\hat{\eta}_1 = (\hat{\mu}_1 - \hat{\mu}_2)/\sqrt{2} = 2.068, \quad \tilde{\eta}_1 = (\tilde{\mu}_1 - \tilde{\mu}_2)/\sqrt{2} = 2.098,$$

In fact, the $\lvert T \rvert$ and $\lvert t \rvert$ tests provide overwhelming evidence for $H_0 : \eta_1 = 0$ not being true.

We will show in Chapter 8 that the T test is robust and has both criterion as well as efficiency robustness. The t and G tests are not robust. We reiterate, since deviations from an assumed distribution are very common, one cannot feel comfortable with assuming a particular distribution and believing it to be exactly correct. For example, somewhat different values of b_i will also provide 'close to a straight line' pattern for the Box-Cox data above. It is, therefore, advantageous to use robust procedures to safeguard against plausible deviations from an assumed distribution and to outliers and inliers in a sample. This will be discussed in detail in Chapter 8.

SUMMARY

In this Chapter, we consider one-way and two-way classification models in experimental design. We consider the realistic situations when the errors have non-normal distributions. We obtain the MMLE for block, column and interaction effects. We show that the estimators are explicit functions of sample observations and are highly efficient. We define statistics analogous to the normal-theory F statistics to test the block, column and interaction effects. We show that the new tests are more powerful than the traditional normal-theory tests. We give MMLE of linear contrasts and use them for testing that a linear contrast is zero. We extend the results to situations when the distributions are nonidentical from block to block. We revisit the Box-Cox data and give appropriate statistical analyses.

CHAPTER 7

Censored Samples from Normal and Non-Normal Distributions

7.1 INTRODUCTION

In numerous situations a few observations in a sample are not available due to experimental constraints. Such samples are called censored samples. Consider a life-test experiment with n units when the experimenter decides to terminate the experiment with a number of units still functioning, or the experimenter decides to terminate the experiment at a pre-determined time with a number of units still functioning. The available observations constitute a censored sample. To distinguish between the two situations above, Type I and Type II censoring are used. Type I censoring occurs when a sample of size n is drawn and the observations which have values below a pre-determined limit or above a pre-determined limit are censored; the limits are known but the number of censored observations is random. Type II censoring occurs if a pre-determined number (or proportion) of smallest or a pre-determined number (or proportion) of largest observations are censored. For Type II censoring, the number (or proportion) of censored observations is known while the censoring limits can be considered random. In this chapter, we primarily deal with Type II censored samples and refer to them simply as censored samples. The reason is that censored samples are used in constructing robust procedures, both estimation and hypothesis testing. They are also used for detecting outliers in a sample and in formulating goodness-of-fit tests (Chapter 9). The Tukey and Tiku estimators mentioned in Chapters 1 and 2 are based on symmetrically censored samples and are used as robust estimators of a location parameter. In a symmetric censored sample, an equal number (or proportion) of smallest and largest observations are censored. In life-test experiments, however, only a few largest observations are usually censored. Therefore, we consider the very general situation when a number $r_1 \ge 0$ of smallest observations and a number $r_2 \ge 0$ of largest observations in a random sample of size n are censored. The remaining observations arranged in ascending order of magnitude constitute a censored sample of size $n - r_1 - r_2$.

7.2 ESTIMATION OF PARAMETERS

We consider the estimation of location and scale parameters from a censored sample of size $n - r_1 - r_2$. In certain situations, the ML estimators are explicit functions of the order statistics (censored sample of size $n - r_1 - r_2$)

$$y_{(r_1+1)} \le y_{(r_1+2)} \le \dots \le y_{(n-r_2)} \tag{7.2.1}$$

and their distributions and efficiencies are easy to determine. In most situations, however, the ML estimators are intractable. We illustrate this by considering the exponential, normal, Rayleigh, and the LTS family of distributions (2.2.9). We derive the MML estimators and show that they are explicit functions of sample observations (7.2.1) and are remarkably efficient.

Likelihood function: Let the distribution of the random variable Y be denoted by $(1/\sigma)f((y-\mu)/\sigma)$. The likelihood function L is the joint pdf of the order statistics (7.2.1),

$$L = \frac{n!}{r_1!\,r_2!}\left(\frac{1}{\sigma}\right)^{n-r_1-r_2}(F\{z_{(r_1+1)}\})^{r_1}(1-F\{z_{(n-r_2)}\})^{r_2}\prod_{i=r_1+1}^{n-r_2} f(z_{(i)}) \tag{7.2.2}$$

where $F(z)=\int_{-\infty}^{z} f(z)\,dz$ is the cdf and $z_{(i)}=(y_{(i)}-\mu)/\sigma$, $r_1+1 \le i \le n-r_2$. Since Y is assumed to have a location-scale distribution, f(z) is free of μ and σ.

Exponential: Let the distribution of Y be the exponential

$$E(\theta, \sigma): (1/\sigma)\exp\{-(y-\theta)/\sigma\}, \quad \theta < y < \infty. \tag{7.2.3}$$

Writing $z=(y-\theta)/\sigma$,

$$f(z) = e^{-z} \quad \text{and} \quad F(z)=\int_0^z f(z)\,dz = 1-e^{-z}. \tag{7.2.4}$$

The likelihood function is

$$L \propto \sigma^{-(n-r_1-r_2)}\{1-\exp(-z_{(r_1+1)})\}^{r_1}\exp\left\{-\left(\sum_{i=r_1+1}^{n-r_2} z_{(i)} + r_2 z_{(n-r_2)}\right)\right\}; \tag{7.2.5}$$

$$z_{(i)}=\{y_{(i)}-\theta\}/\sigma.$$

To estimate θ and σ, we have the likelihood equations (Epstein and Sobel, 1953; Tiku, 1967b; Kambo, 1978)

$$\frac{\partial \ln L}{\partial \theta} = \frac{1}{\sigma}[(n-r_1) - r_1 g(z_{(r_1+1)})] = 0 \tag{7.2.6}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = \frac{1}{\sigma}\left[-(n-r_1-r_2) + \sum_{i=r_1+1}^{n-r_2} z_{(i)} - r_1 z_{(r_1+1)} g(z_{(r_1+1)}) + r_2 z_{(n-r_2)}\right] = 0 \tag{7.2.7}$$

where $g(z)=f(z)/F(z)$; f(z) and F(z) are given in (7.2.4).

The solutions of (7.2.6)-(7.2.7) give the ML estimators which when corrected for bias are (Kambo, 1978)

$$\hat{\theta}_c = y_{(r_1+1)} - a\hat{\sigma}, \; a = \sum_{i=1}^{r_1+1} 1/(n-i+1), \tag{7.2.8}$$

and

$$\hat{\sigma}_c = \left[\sum_{i=r_1+1}^{n-r_2} y_{(i)} + r_2 y_{(n-r_2)} - (n-r_1) y_{(r_1+1)}\right] / (n-r_1-r_2-1) \tag{7.2.9}$$

To evaluate the variances and the covariance of the estimators, we express $\hat{\sigma}_c$ in terms of sample spacings.

Spacings: Consider the order statistics

$$y_{(1)} \le y_{(2)} \le \ldots\ldots \le y_{(n)} \tag{7.2.10}$$

of a random sample of size n from the exponential $E(\theta, \sigma)$. The exponential sample spacings are defined as

$$D_i = (n-i+1)\{y_{(i)} - y_{(i-1)}\} \quad (y_{(0)}=\theta), \quad 1 \le i \le n. \tag{7.2.11}$$

The spacings D_i are iid exponential $E(0, \sigma)$; see Appendix 7A. Therefore, $2D_i/\sigma$ $(1 \le i \le n)$ are independently distributed as chi-squares, each with 2 degrees of freedom.

Lemma 7.1: The estimator $\hat{\sigma}_c$ is unbiased and its variance is

$$V(\hat{\sigma}_c) = \sigma^2/(n - r_1 - r_2 - 1).$$

Proof: It is easy to show that $(n - r_1 - r_2 - 1)\,\hat{\sigma}_c$ is the sum of $n - r_1 - r_2 - 1$ spacings D_i, i.e.,

$$(n - r_1 - r_2 - 1)\,\hat{\sigma}_c = \sum_{i=r_1+2}^{n-r_2} D_i\,. \tag{7.2.12}$$

Since D_i $(1 \le i \le n)$ are iid with $E(D_i) = \sigma$ and $V(D_i) = \sigma^2$, the result follows.

Corollary: The distribution of $2(n - r_1 - r_2 - 1)\hat{\sigma}_c{}^2/\sigma^2$ is chi-square with $2(n - r_1 - r_2 - 1)$ degrees of freedom.

Lemma 7.2: The statistics $y_{(r_1+1)}$ and $\hat{\sigma}_c$ are independently distributed.

Proof: This follows from the fact that the likelihood function (7.2.5) factorizes (Tiku, 1981b), that is,

$$L \propto \exp\{-(n - r_1)(y_{(r_1+1)} - \theta)/\sigma\}\{1 - \exp\{-(y_{(r_1+1)} - \theta)/\sigma\}\}^{r_1} \times e^{-(n - r_1 - r_2 - 1)\hat{\sigma}_c/\sigma}. \tag{7.2.13}$$

The result is important for hypothesis testing.

Lemma 7.3: The estimator $\hat{\theta}_c$ is unbiased with variance

$$V(\hat{\theta}_c) = \left[\sum_{i=1}^{r_1+1} \frac{1}{(n-i+1)^2} + \frac{a^2}{n - r_1 - r_2 - 1}\right]\sigma^2.$$

Proof: Follows immediately from the results given in (1A.8) and Lemmas 7.1-7.2.

Corollary: The covariance between $\hat{\theta}_c$ and $\hat{\sigma}_c$ is

$$\mathrm{Cov}\,(\hat{\theta}_c, \hat{\sigma}_c) = -\,a\sigma^2/(n - r_1 - r_2 - 1). \tag{7.2.14}$$

One-sided censoring: In most life-test experiments assuming exponentiality, $r_1 = 0$ (Epstein and Sobel, 1953; Weissman, 1978). Writing $r_2 = r$, the following results follow as particular cases since $a = 1/n$ for $r_1 = 0$:

(a) The ML estimators $\hat{\theta}_c = y_{(1)} - \hat{\sigma}_c/n$ and $\hat{\sigma}_c = \sum_{i=1}^{n-r} y_{(i)}/(n-r-1)$ are unbiased for θ and σ, respectively, and

$$V(\hat{\theta}_c) = \left[1 + \frac{1}{n - r - 1}\right]\frac{\sigma^2}{n^2},\quad V(\hat{\sigma}_c) = \frac{\sigma^2}{n - r - 1} \tag{7.2.15}$$

and $$\mathrm{Cov}(\hat{\theta}_c, \hat{\sigma}_c) = \frac{\sigma^2}{n(n - r - 1)}.$$

(b) The random variables $2n(y_{(1)} - \theta)/\sigma$ and $2(n - r - 1)\hat{\sigma}_c/\sigma$ are independently distributed as chi-squares with 2 and $2(n - r - 1)$ degrees of freedom.

The result (b) is important for hypothesis testing.

To test H_0: $\theta = 0$ against H_0: $\theta > 0$, the statistic

$$F = n(y_{(1)} - \theta_0)/\hat{\sigma}_c \tag{7.2.16}$$

is used. Large values of F lead to the rejection of H_0 in favour of H_1. The null distribution of F is central F with $\nu_1 = 2$ and $\nu_2 = 2(n - r - 1)$ degrees of freedom.

Example 7.1: Ten electronic components of each of the two types A and B were put to test. Due to some technical difficulty, recording of the failure times could not begin until two components in each of the two types had already failed. The failure times (in hours) of the remaining components were recorded as follows (Tiku, 1981b):

A: —, —, 409, 455, 509, 519, 541, 543, 550, 583

B: —, —, 429, 463, 526, 527, 542, 560, 567, 598.

Assuming exponential distributions $E(\theta_i, \sigma)$ (i = 1, 2) for the life times, one wants to test the null hypothesis $H_0 = \theta_1 = \theta_2$ against $H_1 = \theta_1 \neq \theta_2$; θ_1 and θ_2 are the minimum life times of the two types A and B, respectively.

Consider the general situation when

$$y_{1,(r_1+1)} \quad \text{and} \quad y_{2,(r_2+1)}$$

are the smallest order statistics in the two censored samples

$$y_{i,(r_i+1)} \le y_{i,(r_i+2)} \le \ldots\ldots \le y_{i,(n_i-s_i)} \quad (i = 1, 2) \tag{7.2.17}$$

from the two exponentials $E(\theta_i, \sigma)$ (i = 1, 2); $s_i \ge 0$ are the number of largest observations censored in the two samples, respectively. To test H_0 against H_1, Tiku (1981b) defines the statistic

$$U = |y_{1,(r_1+1)} - y_{2,(r_2+1)}|/\hat{\sigma}_c \tag{7.2.18}$$

where $\hat{\sigma}_c$ is the pooled estimator, namely,

$$\hat{\sigma}_c = \sum_{i=1}^{2} (n_i - r_i - s_i - 1)\hat{\sigma}_i / \sum_{i=1}^{2} (n_i - r_i - s_i - 1); \tag{7.2.19}$$

$$\hat{\sigma}_i = \left\{ \sum_{j=r_i+1}^{n_i - s_i} y_{i,(j)} + s_i y_{i,(n_i - s_i)} - (n_i - r_i) y_{i,(r_i+1)} \right\} \Big/ (n_i - r_i - s_i - 1) \qquad (i = 1, 2)$$

are the MLE of σ calculated from the two samples. Large values of U lead to the rejection of H_0. For $r_1 = r_2 = 0$, U reduces to the Kumar and Patel (1971) statistic.

Tiku (1981b) derives the null distribution of U using the independence of $y_{i,(r_i+1)}$ and $\hat{\sigma}_c$. He gives the exact expression for the probability

$$P(U \le x \mid H_0) = \int_0^x f(u)\,du$$

$$= \frac{1}{\beta(n_1 - r_1, r_1 + 1)\beta(n_2 - r_2, r_2 + 1)} \left[\sum_{l=0}^{r_2} (-1)^l \binom{r_2}{l} \beta(f + l, r_1 + 1) \right. \tag{7.2.20}$$

$$\times (1/h_2 d)\{1 - (1 + h_2 x)^{-d}\} + \sum_{l=0}^{r_2} (-1)^l \binom{r_1}{l} \beta(f + l, r_2 + 1)$$

$$\left. \times (1/h_1 d)\{1 - (1 + h_1 x)^{-d}\} \right]$$

where $f = n_1 + n_2 - r_1 - r_2$, $d = f - s_1 - s_2 - 2$, $h_1 = (n_1 - r_1 + l)/d$, $h_2 = (n_2 - r_2 + l)/d$ and $\beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$; see Appendix 7B.

For the data above, we have $n_1 = n_2 = n = 10$, $r_1 = r_2 = r = 2$, $s_1 = s_2 = 0$, $d = \sum_{i=1}^{2} (n_i - r_i - s_i - 1) = 14$, $y_{1,(r+1)} = 409$, $y_{2,(r+1)} = 429$ and $\hat{\sigma} = 115.5$.

The statistic U = 20/115.5 = 0.1732 and the probability $P(U \le 0.1732/H_0) = 0.49$ calculated from (7.2.20). The null hypothesis H_0 is not rejected.

Kambo and Awad (1985) generalize the U statistic to k > 2 exponential populations $E(\theta_i, \sigma)$. Tiku and Vaughan (1991) give a modification of U which results in better power properties. They also generalize the modified U statistic to k > 2 exponential populations (Vaughan and Tiku, 1993). See also Shetty and Joshi (1989) who have similar statistics.

Comment: The efficiency of $\hat{\theta}_c$ in (7.2.8) sharply decreases as r_1 increases. For n = 10 and $r_2 = 2$, for example, the values of $(1/\sigma^2)\,V(\hat{\theta}_c)$ for r_1 = 0, 1 and 2 are 0.0243, 0.0575 and 0.1051, respectively. Censoring of smallest observations should, therefore, be avoided as much as possible. Censoring of largest observations in a sample from $E(\theta, \sigma)$ is less damaging; it has the effect of reducing the sample size from n to $n - r_2$.

7.3 CENSORED SAMPLES FROM NORMAL DISTRIBUTION

Consider now the situation when the censored sample (7.2.1) comes from a normal distribution $N(\mu, \sigma^2)$. The likelihood function L is given by (7.2.2) with

$$F(z) = \int_{-\infty}^{z} f(z)\,dz \quad \text{and} \quad f(z) = (2\pi)^{-1/2}\,e^{-z^2/2};\; z = (y - \mu)/\sigma. \tag{7.3.1}$$

The likelihood equations are

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\left[\sum_{i=r_1+1}^{n-r_2} z_{(i)} - r_1 g_1(z_{(r_1+1)}) + r_2 g_2(z_{(n-r_2)})\right] = 0 \tag{7.3.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = \frac{1}{\sigma}\left[-(n - r_1 - r_2) + \sum_{i=r_1+1}^{n-r_2} z_{(i)}^2 - r_1 z_{(r_1+1)}\, g_1(z_{(r_1+1)}) + r_2 z_{(n-r_2)}\, g_2(z_{(n-r_2)})\right] = 0; \tag{7.3.3}$$

$z_{(i)} = (y_{(i)} - \mu)/\sigma$, $g_1(z) = f(z)/F(z)$ and $g_2(z) = f(z)/(1 - F(z))$. The equations have no explicit solutions and have to be solved by iteration. In fact, the Newton-Raphson method can be applied to solve these equations (Schneider, 1986, p. 73). Nonetheless, the fact remains that the ML estimators are implicit and it is difficult to make any analytical study of them.

To derive the MMLE which are asymptotically fully efficient, we note that both the functions $g_1(z)$ and $g_2(z)$ are almost linear in the vicinity of z (Tiku, 1967, p. 155). We, therefore, use the linear approximations

$$g_1(z_{(r_1+1)}) \cong \alpha_1 - \beta_1 z_{(r_1+1)} \quad \text{and} \quad g_2(z_{(n-r_2)}) \cong \alpha_2 + \beta_2 z_{(n-r_2)}. \tag{7.3.4}$$

The coefficients (α_1, β_1) and (α_2, β_2) are obtained from the first two terms of Taylor series expansions around $t_1 = E(z_{(r_1+1)})$ and $t_2 = E(z_{(n-r_2)})$. This gives

$$\beta_1 = \frac{f(t_1)}{q_1}\left[t_1 + \frac{f(t_1)}{q_1}\right] \quad \text{and} \quad \alpha_1 = \frac{f(t_1)}{q_1} + \beta_1 t_1 \tag{7.3.5}$$

and

$$\beta_2 = -\frac{f(t_2)}{q_2}\left[t_2 - \frac{f(t_2)}{q_2}\right] \quad \text{and} \quad \alpha_2 = \frac{f(t_2)}{q_2} - \beta_2 t_2, \tag{7.3.6}$$

$f(t) = (2\pi)^{-1/2} e^{-t^2/2}$; $q_1 = r_1/n$ and $q_2 = r_2/n$. It suffices to use the approximate values of t_1 and t_2 obtained from the equations

$$F(t_1) = r_1/n \quad \text{and} \quad F(t_2) = 1 - r_2/n; \quad F(z) = \int_{-\infty}^{z} f(t)\,dt. \tag{7.3.7}$$

In (7.3.7), r_1/n and r_2/n may as well be equated to $r_1/(n + 1)$ and $r_2(n + 1)$, respectively. That does not make much difference to the results. It may be noted that β_1 and β_2 are both positive fractions ($0 \le \beta_i < 1$) and so are α_1 and α_2. Their values are given in Appendix 7C for easy accessibility.

Incorporating (7.3.5) – (7.3.6) in (7.3.2) – (7.3.3) gives the modified likelihood equations:

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\left[\sum_{i=r_1+1}^{n-r_2} z_{(i)} - r_1(\alpha_1 - \beta_1 z_{(r_1+1)}) + r_2(\alpha_2 + \beta_2 z_{(n-r_2)})\right]$$

$$= \frac{m}{\sigma^2}(K + D\sigma - \mu) = 0 \qquad (7.3.8)$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{1}{\sigma}\left[-(n - r_1 - r_2) + \sum_{i=r_1+1}^{n-r_2} z^2{}_{(i)} - r_1 z_{(r_1+1)}(\alpha_1 - \beta_1 z_{(r_1+1)}) + r_2 z_{(n-r_2)}(\alpha_2 + \beta_2 z_{(n-r_2)})\right]$$

$$= -\frac{1}{\sigma^3}[(A\sigma^2 - B\sigma - C) - m(K - \mu)(K + D\sigma - \mu)] = 0 \qquad (7.3.9)$$

where $A = n - r_1 - r_2$, $m = n - r_1 - r_2 + r_1\beta_1 + r_2\beta_2$, $D = (r_2\alpha_2 - r_1\alpha_1)/m$,

$$K = \left[\sum_{i=r_1+1}^{n-r_2} y_{(i)} + r_1\beta_1 y_{(r_1+1)} + r_2\beta_2 y_{(n-r_2)}\right]/m$$

$$B = r_2\alpha_2(y_{(n-r_2)} - K) + r_1\alpha_1(y_{(r_1+1)} - K) \qquad (7.3.10)$$

and

$$C = \sum_{i=r_1+1}^{n-r_2} (y_{(i)} - K)^2 + r_1\beta_1(y_{(r_1+1)} - K)^2 + r_2\beta_2(y_{(n-r_2)} - K)^2$$

$$= \sum_{i=r_1+1}^{n-r_2} y_{(i)}{}^2 + r_1\beta_1 y^2_{(r_1+1)} + r_2\beta_2 y^2_{(n-r_2)} - mK^2 .$$

The solutions of these equations are the MML estimators:

$$\hat{\mu}_c = K + D\hat{\sigma}_c \quad \text{and} \quad \hat{\sigma}_c = \{B + \sqrt{(B^2 + 4AC)}\}/2\sqrt{A(A-1)}; \qquad (7.3.11)$$

$D = 0$ if $r_1 = r_2$ (symmetric censoring).

It is interesting to note that (7.3.8) – (7.3.11) are exactly of the same form as (2.4.2) – (2.4.3) based on complete samples. Realize also that $\hat{\sigma}_c$ is always real and positive.

The MMLE $\hat{\mu}_c$ and $\hat{\sigma}_c$ are asymptotically equivalent to the corresponding MLE. This is due to the asymptotic equivalence of the modified likelihood and the likelihood equations. The MMLE $\hat{\mu}_c$ and $\hat{\sigma}_c$ are, therefore, asymptotically fully efficient. In fact, Bhattacharyya (1985) proved rigorously that the MMLE obtained by using the linear approximations such as (7.3.4) are asymptotically fully efficient not only for the normal $N(\mu, \sigma^2)$ but also for any other location-scale distribution. This also follows from the results given in Appendix 2A.

Incidentally, by solving the differential equations (7.3.8) – (7.3.9) as in Appendix 2C, we get the modified likelihood function (Tan, 1985; Tiku et al., 1986, p. 38)

$$L^* = \sigma^{-A} \exp\left[-\frac{B}{\sigma} - \frac{C}{2\sigma^2} - \frac{m}{2\sigma^2}(K + D\sigma - \mu)^2\right] H(y) \qquad (7.3.12)$$

where H(y) is an analytic function free of μ and σ; L* has been called robust likelihood function since for $r_1 > 0$ and $r_2 > 0$ it is devoid of the smallest and largest order statistics in a sample and is, therefore, not affected much by outliers or long tails. It also follows from (7.3.12) that K, B and C are asymptotically a minimum set of sufficient statistics for (μ, σ).

Asymptotic covariance matrix: The asymptotic variances and the covariance of $\hat{\mu}_c$ and $\hat{\sigma}_c$ are obtained from the expected values of the second derivatives of ln L or ln L*, the latter being easier to work with in this case. Asymptotically

$$g_1(z_{(r_1+1)}) \cong g_1(t_1) = f(t_1)/q_1 = \alpha_1 - \beta_1 t_1$$

and

$$g_2(z_{(n-r_2)}) \cong g_2(t_2) = f(t_2)/q_2 = \alpha_2 + \beta_2 t_2,$$

and (Harter and Moore, 1966)

$$\lim_{n \to \infty} E\left(\frac{1}{n}\sum_{i=r_1+1}^{n-r_2} z_{(i)}\right) = \frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2} ze^{-z^2/2}\,dz = -[f(t_2) - f(t_1)]. \tag{7.3.13}$$

Similarly,

$$\lim_{n \to \infty} E\left(\frac{1}{n}\sum_{i=r_1+1}^{n-r_2} z_{(i)}^2\right) = (1 - q_1 - q_2) - [t_2 f(t_2) - t_1 f(t_1)]. \tag{7.3.14}$$

Using these results, the elements of the Fisher information matrix I(μ, σ) are obtained:

$$I_{11} = m/\sigma^2 \quad (m = n - r_1 - r_2 + r_1\beta_1 + r_2\beta_2), \quad I_{12} = -(r_2\alpha_2 - r_1\alpha_1)/\sigma^2$$

and

$$I_{22} = [2(n - r_1 - r_2) - (r_2\alpha_2 t_2 - r_1\alpha_1 t_1)]/\sigma^2. \tag{7.3.15}$$

The results (7.3.15) also follow simply by differentiating (7.3.8) with respect to μ and σ and taking expectations as in 2A.10 (Chapter 2). Thus, we have the following results. Realize that $q_1 = r_1/n$ and $q_2 = r_2/n$ are usually less than or equal to 0.2.

Theorem 7.1: Asymptotically (n tend to infinity, r_1 and r_2 or q_1 and q_2 fixed), the MMLE $\hat{\mu}_c$ and $\hat{\sigma}_c$ are unbiased and

$$V(\hat{\mu}_c) \cong \frac{\sigma^2}{m}\left[1 - \frac{(r_2\alpha_2 - r_1\alpha_1)^2}{m\Delta}\right]^{-1}, \ \Delta = 2(n - r_1 - r_2) - (r_2\alpha_2 t_2 - r_1\alpha_1 t_1),$$

$$\mathrm{Cov}(\hat{\mu}_c, \hat{\sigma}_c) \cong \frac{\sigma^2}{m}\frac{(r_2\alpha_2 - r_1\alpha_1)}{\Delta}\left[1 - \frac{(r_2\alpha_2 - r_1\alpha_1)^2}{m\Delta}\right]^{-1} \tag{7.3.16}$$

and

$$V(\hat{\sigma}_c) \cong \frac{\sigma^2}{2A}\left[1 - \frac{(r_2\alpha_2 t_2 - r_1\alpha_1 t_1)}{2A} - \frac{(r_2\alpha_2 - r_1\alpha_1)^2}{2mA}\right]^{-1}.$$

Comment: Since $(r_2\alpha_2 - r_1\alpha_1)^2/m\Delta$ and $(r_2\alpha_2 - r_1\alpha_1)^2/mA$ are very small, they may be ignored. Thus, the equations (7.3.16) simplify further. In particular, the covariance is zero (almost) since $(r_2\alpha_2 - r_1\alpha_1)^2/m\Delta$ is also small.

To test the null hypothesis H_0: μ = 0, we have the following result.

Lemma 7.4: Asymptotically, the conditional (σ known) distribution of $\hat{\mu}_c(\sigma) = K + D\sigma$ is normal with mean μ and variance σ^2/m.

Proof: The result follows from the fact that ∂ ln L*/ ∂μ assumes the form (7.3.8) and

$$E(\partial^r \ln L^*/\partial\mu^r) = 0 \quad (r \geq 3) \tag{7.3.17}$$

Hypothesis testing: Since from the first equation in (7.3.16), $V(\hat{\mu}_c) \cong \sigma^2/m$ to test H_0: $\mu = 0$ we define the statistic

$$T = \sqrt{m}(\hat{\mu}_c/\hat{\sigma}_c) \tag{7.3.18}$$

Large values of $|T|$ lead to the ejection of H_0 in favour of H_1: $\mu \neq 0$.

Since $\hat{\sigma}_c$ converges to σ as $A = n - r_1 - r_2$ becomes large, the null distribution of T is closely approximated by normal N(0, 1) for large $A \geq 20$. For smaller A and $q_1 + q_2 < 0.4$, the probability $P(|t| \geq t_\alpha | H_0)$ is closely approximated by $P(|t| \geq t_\alpha)$, t being the Student t statistic with A – 1 degrees of freedom. The distribution of $(A-1)\hat{\sigma}_c{}^2/h\sigma^2$ is closely approximated by χ^2 with A-1 degrees of freedom; h = V + 1 and

$$V = \text{Var}(\hat{\sigma}_c/\sigma) = (\sigma^2/A)\,[1 - (r_2\alpha_2 t_2 - r_1\alpha_1 t_1)/2A]^{-1}.$$

This result is obtained as in (2.11.19). For q_1+q_2<0.4, however, V is close to$1/2(A-1)$. For example, we have the following values:

n	q_1	q_2	$(1/\sigma^2)V(\hat{\sigma}_c)$	$1/2(A-1)$
10	0.0	0.1	0.058	0.062
	0.1	0.1	0.070	0.071
	0.0	0.2	0.068	0.071
	0.2	0.2	0.098	0.100

The constant h is, therefore, close to 1 and that gives the result that $(A-1)\hat{\sigma}^2/\sigma^2$ is closely approximated by χ^2 with A – 1 degrees of freedom. The result can also be established as follows.

Lemma 7.5: For large A, the distribution of $(A-1)\hat{\sigma}^2/\sigma^2$ is chi-square with A – 1 degrees of freedom.

Proof: This follows from the fact that $\partial \ln L^*/\partial\sigma$ assumes exactly the same form as (2.9.4) with n replaced by A and

$$B_0 = r_2\alpha_2(y_{(n-r_2)} - \mu) - r_1\alpha_1(y_{(r_1+1)} - \mu)$$

and

$$C_0 = \sum_{i=r_1+1}^{n-r_2}(y_{(i)} - \mu)^2 + r_1\beta_1(y_{(r_1+1)} - \mu)^2 + r_2\beta_2(y_{(n-r_2)} - \mu)^2. \tag{7.3.19}$$

Moreover, $B_0/\sqrt{(AC_0)} \cong 0$ and $C_0 = m(K-\mu)^2 + C$;

K and C are given in (7.3.10) – (7.3.11). Thus, the distribution of $(A-1)\hat{\sigma}^2/\sigma^2$ is chi-square with A – 1 degrees of freedom since K, a linear function of order statistics, is asymptotically normally distributed.

7.4 SYMMETRIC CENSORING

If in (7.2.1) $r_1 = r_2 = r$, then the resulting sample

$$y_{(r+1)} \leq y_{(r+2)} \leq \dots \leq y_{(n-r)} \tag{7.4.1}$$

is a symmetric censored sample of size $A = n - 2r = n(1-2q)$, $q = r/n$. All the results given in the previous section greatly simplify now. For example, we have

$$V(\hat{\mu}_c) \cong \sigma^2/m, \text{ Cov }(\hat{\mu}_c, \hat{\sigma}_c) = 0 \tag{7.4.2}$$

and $$V(\hat{\sigma}_c) \cong (\sigma^2/2A)[1 - r\alpha t/A]^{-1};$$

t is given by $F(t) = 1 - q$, $m = n - 2r + 2r\beta$, $\alpha_1 = \alpha_2 = \alpha$ and $\beta_1 = \beta_2 = \beta$ in (7.3.4), and α and β are both positive fractions between 0 and 1. Also, we have the following beautiful result.

Lemma 7.6: The estimator $\hat{\mu}_c$ is BAN and is independent of $\hat{\sigma}_c$.

Proof: Follows from the fact that $\partial \ln L^* / \partial\mu$ assumes the form

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{m}{\sigma^2}(\hat{\mu}_c - \mu) \tag{7.4.3}$$

and $E(\partial^{r+s} \ln L^*/\partial\mu^r \partial\sigma^s) = 0$ for all $r \geq 1$ and $s \geq 1$.

Remark: The estimator $\hat{\mu}_c$ is unbiased for all n. This follows from symmetry. Also, the minimum variance bound for estimating μ is σ^2/m (for large A).

To have an idea about how efficient $\hat{\mu}_c$ is for small n, we give below in Table 7.1 the values of $MVB(\mu) = \sigma^2/m$ and the exact variance of $\hat{\mu}_c$ calculated from

$$V(\hat{\mu}_c) = (\beta' \, \Omega \, \beta)\sigma^2/m^2, \tag{7.4.4}$$

$\beta' = (1 + r\beta, 1, 1, \ldots, 1, 1 + r\beta)$ is a vector with $n - 2r$ elements and Ω is the variance-covariance matrix of the standardized nor mal variates $z_{(i)} = (y_{(i)} - \mu)/\sigma$, $r + 1 \leq i \leq n - r$. The elements of Ω are given in Pearson and Hartley (1972, Table 10) for $n \leq 20$. Also given in Table 7.1 are the values of the variance $V(\mu_c^*)$; μ_c^* is the BLUE obtained exactly along the same lines as in Section 2.7 (Chapter 2). Among linear unbiased estimators, μ_c^* has the minimum possible variance as said earlier.

Table 7.1: The exact values of (a) $(1/\sigma^2)V(\mu_c^*)$, (b) $(1/\sigma^2)V(\hat{\mu}_c)$ and (c)$(1/\sigma^2)MVB(\mu)$, for symmetric censored normal samples.

	n = 5			n = 10			n = 20		
Q	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)
0.0	0.200	0.200	0.200	0.100	0.100	0.100	0.0500	0.0500	0.0500
0.1				0.104	0.104	0.103	0.0520	0.0520	0.0517
0.2	0.226	0.226	0.217	0.111	0.111	0.109	0.0552	0.0552	0.0547
0.3				0.122	0.122	0.119	0.0602	0.0602	0.0593
0.4				0.138	0.138	0.133	0.0679	0.0679	0.0665

The MMLE $\hat{\mu}_c$ is so highly efficient even for small n that there is perhaps no room for any other estimator including the MLE to be more efficient. Moreover, the computation of $\hat{\mu}_c$ and $\hat{\sigma}_c$ is so straightforward.

Exactly along the same lines as Lemma 7.5, it follows that for a symmetric censored sample $(A - 1)\hat{\sigma}_c^2/\sigma^2$ is a multiple 1/h ($h \cong 1$ for large A) of chi-square with $A - 1$ degrees of freedom. Moreover, $\hat{\mu}_c$ and $\hat{\sigma}_c$ are independently distributed (asymptotically).

To have an idea about how efficient $\hat{\sigma}_c$ is, we give below the simulated values of the mean and variance of $\hat{\sigma}_c$. Also given are the variances of the BLUE σ_c^* and the minimum variance bound $MVB(\sigma) \cong \sigma^2/2(A - 1)$; $\sigma = 1$ without loss of generality:

q	$E(\hat{\sigma}_c)$	$V(\hat{\sigma}_c)$	1/ 2(A – 1)	$V(\sigma_c^*)$
		n = 10		
0.1	0.96	0.075	0.071	0.082
0.2	0.93	0.104	0.100	0.129
		n = 20		
0.1	0.99	0.034	0.033	0.038
0.2	0.99	0.049	0.045	0.058

It can be seen that $\hat{\sigma}_c$ is highly efficient and is more efficient than the BLUE σ_c^* as expected. This is due to the fact that $\hat{\sigma}_c$ is nonlinear and is the MVB estimator for large n (r_1 and r_2 or q_1 and q_2 fixed); $\hat{\sigma}_c$ has the minimum variance only among linear unbiased estimators. Moreover, $\hat{\sigma}_c$ is so easy to compute but the computation of σ_c^* is much involved.

Example 7.2: In an accelerated life-test experiment involving specimens of electrical insulation (Lawless, 1982, p.226), 10 specimens were put on test and the test was terminated at the time of the eight failure. The eight observed log-failure times are

6.00 6.43 6.77 7.07 7.40 7.66 8.10 8.40 — —

which constitute a censored sample with n = 10, $r_1 = 0$ and $r_2 = 2$ in (7.2.1). Assuming normality $N(\mu, \sigma^2)$, the ML and MML estimates are

		μ	σ
Cohen	MLE	7.59	1.069 (corrected for bias)
Tiku	MMLE	7.5895	1.0696 (equation 7.3.10)

The MMLE are numerically the same as the MLE.

Example 7.3: The data below (Gupta, 1952, p.271) show the days on which the first 7 of a sample of 10 tested mice died after being inoculated with a uniform culture of human tuberculosis:

Days after inoculation	41	44	46	54	55	58	60	—	—	—
Log days after inoculation	1.613	1.644	1.663	1.732	1.740	1.763	1.778	—	—	—

The distribution of Y = log (to the base e) days is assumed to be normal $N(\mu, \sigma^2)$. Here, the estimates are

		μ	σ
Cohen	MLE	1.7422	0.0792 (Schneider, 1986, p.104)
Tiku	MMLE	1.7423	0.0794 (Schneider, 1986, p.104)
Lloyd	BLUE	1.746	0.101 (Sarhan and Greenberg, 1962, pp.233-256)

The ML and MML estimates are numerically very close to one another. They also have essentially the same standard errors. Since $V(\hat{\mu}_c) \cong \sigma^2/m$ and $V(\hat{\sigma}_c) \cong \sigma^2/2(A-1)$, $m = n - r_1 - r_2 + r_1\beta_1 + r_2\beta_2$, n = 10, $r_1 = 0$, $r_2 = 3$, $\beta_2 = 0.7355$, the standard errors are

$$S.E.(\hat{\mu}_c) = \pm\ 0.079\sqrt{0.1086} = \pm\ 0.026$$

and
$$S.E.(\hat{\sigma}_c) = \pm\ 0.079\sqrt{0.0833} = \pm\ 0.023.$$

The standard errors of the BLUE are

$$S.E.(\mu_c^*) = \pm\ 0.034 \quad \text{and} \quad S.E.\ (\sigma_c^*) = \pm\ 0.032.$$

As expected, BLUE have somewhat larger standard errors. Moreover, their computation is much involved.

Example 7.4: In a time-mortality experiment (Cohen, 1957), m_1 specimens die before observation begins. The experiment is subsequently terminated with m_2 specimens remaining alive. Actual survival times are recorded for the specimens which die during the period of full observation. For a specific sample of this type in which Y, the log survival time in days, is assumed to be normally distributed,

$$m_1 = 2,\ m = 40,\ m_2 = 5,\ \bar{y} = \sum_{i=3}^{42} y_i/40 = 1.62011,\ y_{(3)} = 1.301030,$$

$$y_{(42)} = 1.903090 \quad \text{and} \quad \sum_{i=3}^{42} y_i^2/40 = 2.646499.$$

Here n = 47, $q_1 = 0.04255$, $q_2 = 0.10638$, and (Tiku, 1967, p.164) $\alpha_1 = 0.62689$, $\beta_1 = 0.87523$, $\alpha_2 = 0.69364$, $\beta_2 = 0.83065$.

Substituting these values in equations (7.3.10)-(7.3.11) gives the following MMLE and their variances and the covariance (Tiku et al., 1986, p.164):

$$\hat{\mu}_c = 1.64336, \quad \hat{\sigma}_c = 0.20287$$

$$(1/\sigma^2)V(\hat{\mu}_c) = 0.8978 \times 10^{-3}, \quad (1/\sigma^2)V(\hat{\sigma}_c) = 0.5598 \times 10^{-3},$$

$$(1/\sigma^2)\ \mathrm{Cov}(\hat{\mu}_c, \hat{\sigma}_c) = 0.0268 \times 10^{-3}.$$

These may be compared with Cohen's (1957, p.235) MLE. Incidentally, there is a mistake in Cohen's inverse matrix giving the variances and the covariance of the ML estimators. This has been corrected and the following obtained:

$$\hat{\hat{\mu}}_c = 1.64330, \quad \hat{\hat{\sigma}}_c = 0.20278$$

$$(1/\sigma^2)V(\hat{\hat{\mu}}_c) = 0.8962 \times 10^{-3}, \quad (1/\sigma^2)V(\hat{\hat{\sigma}}_c) = 0.5591 \times 10^{-3},$$

$$(1/\sigma^2)\ \mathrm{Cov}(\hat{\hat{\mu}}_c, \hat{\hat{\sigma}}_c) = 0.0265 \times 10^{-3}.$$

It can be seen that the modified likelihood methodology yields essentially the same results as the maximum likelihood; see also Chapter 4.

For estimation of parameters in a log-normal distribution, see Finney (1941), Hill (1963) and Tiku (1968a).

Robustness: Symmetric censored samples are used to formulate robust estimators of the location (mean) and scale (standard deviation) μ and σ of LTS distributions. In fact, Tiku (1980a) in J. Statistical Planning and Inference 4, page 132, made the following statement:

"Nonnormality essentially comes from the tails and once the extreme observations representing the tails are censored, there is hardly any difference between a normal sample and a nonnormal sample and in that situation whatever is good for normal is automatically good for nonnormal samples."

This statement is in the spirit of the following colourful remark due to Edgeworth, F.Y. (1987). Philosophical Magazine 24, page 269:

"The method of Least Squares is seen to be our best course when we have thrown overboard a certain portion of our data—a sort of sacrifice which has often to be made by those who sail upon the stormy seas of probability."

In the next chapter (Chapter 8), we discuss in detail the robustness features of the MML estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ based on symmetric censored samples. The robustness of the statistic

$$t = \sqrt{m}\,(\hat{\mu}_c/\hat{\sigma}_c) \tag{7.4.5}$$

used for testing the null hypothesis $H_0 : \mu = 0$ is discussed in Tiku (1980a). It is shown that for long-tailed symmetric (LTS) distributions with finite mean and variance, and outlier and contamination models, t is as robust as other tests, e.g., the test based on Huber M-estimators; see also Dunnett (1982).

Example 7.5: Consider the following Darwin's well-known data (Fisher, 1966) which represents the differences (in heights) between cross—and self-fertilized plants of the same pair grown together in one pot:

49 –67 8 16 6 23 28 41 14 29 56 24 75 60 –48

By using a formal test of outliers, we show in Chapter 9 that the two smallest and one largest observations in this data are outliers. It is imperative that the influence of outliers be depleted to make efficient estimation of parameters possible (Huber, 1981; Tiku et al., 1986, Chapter 7). A simple and effective way of doing this is to censor all outliers in the data. In the present situation, we censor the two smallest and one largest observations. The resulting sample

6 8 14 16 23 24 28 29 41 49 56 60

is a censored sample with n = 15, $r_1 = 2$, $r_2 = 1$ and A = 12. The MML estimates obtained from (7.3.10)-(7.3.11) and their standard errors are

$$\hat{\mu}_c = 27.418, \quad \hat{\sigma}_c = 23.452; \quad SE(\hat{\mu}_c) = \pm\, \hat{\sigma}_c/\sqrt{m} = \pm\, 6.16,$$

$$SE(\hat{\sigma}_c) = \pm\, \hat{\sigma}_c/\sqrt{22} = \pm\, 5.00.$$

If we had assumed that the 15 observations are from normal $N(\mu, \sigma^2)$, we would use

$$\bar{y} = 20.933 \text{ and } s = 37.744; \; SE(\bar{y}) = \pm\, 9.75 \text{ and } SE(s) = \pm\, 6.89.$$

The sample mean and sample standard deviation are clearly inefficient if the sample contains outliers.

To use the estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ based on symmetric censored samples, we censor the second largest observation also. From the remaining 11 observations (a symmetric censored sample with n = 15, $r_1 = r_2 = 2$), we have ($q = 2/15$, $\alpha = 0.710$, $\beta = 0.820$)

$$\hat{\mu}_c = 27.713 \quad \text{and} \quad \hat{\sigma}_c = 25.498; \; SE(\hat{\mu}_c) = \pm\, 6.74, \; SE\,(\hat{\sigma}_c) = \pm\, 5.70$$

The MMLE and their standard errors calculated from the censored samples with n = 15, $r_1 = 2$ and $r_2 = 1$, and n = 15, $r_1 = 2$ and $r_2 = 2$, are not much different from one another. In fact, we show in Chapter 9 that it is very important to remove all outliers from the sample for efficient estimation of parameters. If in the process a small number of good observations are sacrificed, that does not affect the efficiencies of the estimators too adversely.

Tan (1985) studied from a Bayesian point of view the MML estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ above and the statistic $t = \sqrt{m}\,(\hat{\mu}_c/\hat{\sigma}_c)$ based on them. His results are in complete agreement: (i) the estimators are highly efficient, and (ii) the distribution of t is closely approximated by the Student t with $v = A - 1$ degrees of freedom; see also Tan and Balakrishnan (1986).

7.5 CENSORED SAMPLES IN EXPERIMENTAL DESIGN

As in Chapter 6 which deals with complete samples, the results above based on censored samples readily extend to Experimental Design (Tiku, 1973). Consider, for example, the one-way classification model

$$y_{ij} = \mu + g_i + e_{ij} \quad (1 \le i \le k, \quad 1 \le j \le n); \tag{7.5.1}$$

e_{ij} are assumed to be iid normal $N(0, \sigma^2)$, μ is a constant and g_i is the effect due to the i^{th} block. Consider, for simplicity, the symmetric censored samples

$$y_{i,(r_i+1)}, y_{i,(r_i+2)}, \ldots\ldots y_{i,(n_i-r_i)} \quad (1 \le i \le k) \tag{7.5.2}$$

available from the k blocks. Writing

$$z_{i,j} = (y_{i,(j)} - \mu - g_i)/\sigma (r_i + 1 \le j \le n - r_i),\ a_i = r_i + 1 \text{ and } b_i = n - r_i \quad (1 \le i \le k), \tag{7.5.3}$$

and proceeding as in Section 7.3, the modified likelihood equations are obtained. They are

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{k}\left\{\sum_{j=a_i}^{b_i} z_{i,j} + r_i\beta_i(z_{i,a_i} + z_{i,b_i})\right\} = 0$$

$$\frac{\partial \ln L}{\partial g_i} \cong \frac{\partial \ln L^*}{\partial g_i} = \frac{1}{\sigma}\left\{\sum_{j=a_i}^{b_i} z_{i,j} + r_i\beta_i(z_{i,a_i} + z_{i,b_i})\right\} = 0 \tag{7.5.4}$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{1}{\sigma}\sum_{i=1}^{k}$$

$$\left\{-(n - 2r_i) + \sum_{j=a_i}^{b_i} z_{i,j}{}^2 - r_i z_{i,a_i}(\alpha_i - \beta_i z_{i,a_i}) + r_2 z_{i,b_i}(\alpha_i + \beta_i z_{i,b_i})\right\} = 0.$$

The coefficient α_i and β_i are obtained from (7.3.6) with t_2 replaced by t_i and q_2 replaced by $q_i = r_i/n$; t_i is determined by $\int_{-\infty}^{t_i} f(t)dt = 1 - q_i$ $(1 \le i \le k)$.

Assuming the constraint

$$\sum_{i=1}^{k} m_i g_i = 0 \quad (m_i = n - 2r_i + 2r_i\beta_i) \tag{7.5.5}$$

in the model (7.5.1), we obtain the following MML estimators as solutions of the equations in (7.5.4):

$$\hat{\mu} = \sum_{i=1}^{k} m_i K_i/m, \quad \hat{g}_i = K_i - \hat{\mu} \quad (1 \le i \le k) \tag{7.5.6}$$

and

$$\hat{\sigma} = \{B + \sqrt{(B^2 + 4AC)}\}/2A \tag{7.5.7}$$

where

$$M = \sum_{i=1}^{k} m_i, \quad K_i = \left\{\sum_{j=r_i+1}^{k-r_i} y_{i,(j)} + r_i\beta_i(y_{i,(r_i+1)} + y_{i,(n-r_i)})\right\}/m_i,$$

$$A = \sum_{i=1}^{k}(n_i - 2r_i), \quad B = \sum_{i=1}^{k} r_i\alpha_i(y_{i,(n-r_i)} - y_{i,(r_i+1)}) \quad \text{and} \tag{7.5.8}$$

$$C = \sum_{i=1}^{k}\left\{\sum_{j=r_i+1}^{n-r_i} y_{i,(j)}^2 + r_i\beta_i(y_{i,(r_i+1)}^2 + y_{i,(n-r_i)}^2 - m_i K_i{}^2\right\}.$$

The divisor 2A in $\hat{\sigma}$ may be replaced by $2\sqrt{A(A-k)}$ as a bias correction.

The estimators $\hat{\mu}$ and $\hat{g}_i$ $(1 \le i \le k)$ are unbiased. This follows from symmetry. For large A (in fact, large n with r_i or $q_i = r_i/n$ fixed), the distribution of $\hat{\mu}_i = \hat{\mu} + \hat{g}_i = K_i$ is normal with mean $\mu_i = \mu + g_i$ and variance σ^2/m_i $(1 \le i \le k)$. Realize that $\hat{\mu}_i$ are independently distributed. Under the null hypothesis H_0: $g_1 = g_2 = \ldots = g_k = 0$ (i.e, $\mu_1 = \mu_2 \ldots = \mu_k = \mu$), the distribution of

$$\chi^2 = \sum_{i=1}^{k} m_i(\hat{\mu}_i - \hat{\mu})^2/\sigma^2 \quad \left(\hat{\mu} = \sum_{i=1}^{k} m_i\hat{\mu}_i/M\right) \tag{7.5.9}$$

is chi-square with k – 1 degrees of freedom.

For large A, $(A - k)\hat{\sigma}^2/\sigma^2$ is distributed as chi-square with A – k degrees of freedom. Under H_0, therefore,

$$F = \sum_{i=1}^{k} m_i(\hat{\mu}_i - \hat{\mu})^2/(k-1)\hat{\sigma}^2 \tag{7.5.10}$$

has a central F distribution with (k – 1, A – k) degrees of freedom. Large values of F lead to the rejection of H_0 in favour of $H_1 : \mu_i \ne \mu_j$ $(i \ne j)$. The F statistic has excellent efficiency and robustness properties (Tiku, 1973; Tiku, 1980).

It may be noted that the constraint (7.5.5) does not affect the value of $\mu + g_1$ in the linear model (7.5.1) or its estimator $\hat{\mu} + \hat{g}_i$ $(1 \le i \le n)$.

Linear contrast: One is often interested in estimating a linear contrast

$$\theta = \sum_{i=1}^{k} l_i g_i = \sum_{i=1}^{k} l_i\mu_i \quad (\mu_i = \mu + g_i), \quad \sum_{i=1}^{k} l_i = 0$$

and testing that it is equal to zero. The MML estimator of θ is

$$\hat{\theta} = \sum_{i=1}^{k} l_i\hat{\mu}_i, \quad \hat{\mu}_i = K_i; \tag{7.5.11}$$

$\hat{\theta}$ is unbiased and since $\hat{\mu}_i$ $(1 \le i \le k)$ are independent of one another, its variance for large n (r_i or $q_i = r_i/n$ fixed) is

$$V(\hat{\theta}) \cong \sigma^2 \sum_{i=1}^{k} (l_i^2/m_i). \tag{7.5.12}$$

The MML estimator of σ is $\hat{\sigma}$ (equation 7.5.7).

The null distribution of the statistic

$$t_1 = \left(\sum_{i=1}^{k} l_i\hat{\mu}_i\right) \Big/ \hat{\sigma}\sqrt{\left\{\sum_{i=1}^{k} (l_i^2/m_i)\right\}} \tag{7.5.13}$$

is the Student t with A – k degrees of freedom. The t_1 statistic has excellent efficiency and robustness properties (Chapter 8).

To test $\tau = \sum_{i=1}^{k} l_i g_i = 0$, the MML estimator is

$$\hat{\tau} = \sum_{i=1}^{k} l_i\hat{g}_i \text{ with variance } V(\hat{\tau}) \cong \sigma^2 \sum_{i=1}^{k} l_i^2\left(\frac{1}{m_i} - \frac{1}{M}\right), \tag{7.5.14}$$

since $Cov(\hat{\mu}_i, \hat{\mu}) \cong \sigma^2/M$ and $V(\hat{\mu}) \cong \sigma^2/M$, $M = \sum_{i=1}^{k} m_i$. The null distribution of

$$t_2 = \left(\sum_{i=1}^{k} l_i \hat{g}_i\right) \Big/ \hat{\sigma}\sqrt{\left\{\sum_{i=1}^{k} l_i^{\,2}\left(\frac{1}{m_i} - \frac{1}{M}\right)\right\}} \qquad (7.5.15)$$

is the Student t with A – k degrees of freedom (Tiku, 1973). See also Tiku (1978) who gives solutions when censoring occurs in linear regression models.

The estimators and tests above readily extend to situations when the number of observations in the blocks are not necessarily equal or the censoring is not symmetric.

7.6 INLIERS IN NORMAL SAMPLES

Inliers are “bad” observations located close to the mean. Outliers are “bad” observations located away from the mean. While outliers are created by a mechanism which pulls a few smallest and/or largest observations away from the mean (Tiku, 1975, 1977; Hawkins, 1977), inliers are created by a mechanism which pushes a few smallest and/or largest observations towards the mean (Tiku et al., 2001; Akkaya and Tiku, 2003). There are, of course, a number of ways to create outliers in a sample (Barnett and Lewis, 1978). An effective way of depleting the influence of long-tails or outliers is to censor a proportion of the smallest and largest observations in a sample. The resulting MMLE are remarkably robust to outliers (Chapter 8). Similarly, the influence of inliers is depleted by censoring a small proportion of observations in the middle of the ordered sample. We develop the MMLE from such samples as follows (Akkaya and Tiku, 2003):

Let y_1, y_2,... y_n be a random sample from a normal population $N(\mu, \sigma^2)$. Consider the censored sample

$$y_1 \le y_2 \le \ldots\ldots \le y_{(r_1)} \le \ldots\ldots \le y_{(n-r_2+1)} \le \ldots\ldots \le y_{(n)} \qquad (7.6.1)$$

with $n - r_1 - r_2$ observations censored in the middle of the ordered sample $y_{(i)}(1 \le i \le n)$. Although our method applies to the general situation when $r_1 \ne r_2$ but, for simplicity, we take $r_1 = r_2 = r$. In the framework of robustness, r is chosen to be equal to [0.4n + 1/2]. Thus, nearly twenty percent order statistics in the middle are censored. Writing $z_{(i)} = (y_{(i)} - \mu)/\sigma$, the likelihood function L is

$$L \propto \sigma^{-2r} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{r} z_{(i)}^{\,2} + \sum_{i=n-r+1}^{n} z_{(i)}^{\,2})\right)\right\} [F(z_{(n-r+1)}) - F(z_{(r)})]^{n-2r}, \qquad (7.6.2)$$

$F(z) = \int_{-\infty}^{z} f(z)dz$ and $f(z) = (2\pi)^{-1/2} \exp(-z^2/2)$. The likelihood equations for estimating μ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\left(\sum_{i=1}^{r} z_{(i)} + \sum_{i=n-r+1}^{n} z_{(i)}\right) - \frac{(n-2r)}{\sigma}[g(z_{(n-r+1)}) - g(z_{(r)})] = 0 \qquad (7.6.3)$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{2r}{\sigma} + \frac{1}{\sigma}\left(\sum_{i=1}^{r} z_{(i)}^{\,2} + \sum_{i=n-r+1}^{n} z_{(i)}^{\,2}\right) - \frac{(n-2r)}{\sigma}[z_{(n-r+1)}\, g(z_{(n-r+1)}) - z_{(r)}\, g(z_{(r)})] = 0; \qquad (7.6.4)$$

$$g(z_{(j)}) = \frac{f(z_{(j)})}{[F(z_{n-r+1)}) - F(z_{(r)})]}. \qquad (7.6.5)$$

The equations (7.6.3)-(7.6.4) have no explicit solutions. The ML estimators are, therefore, elusive.

To obtain the MML estimators, we first note that F(z) is distributed as uniform (0, 1). The ordered variates $u_i = F(z_{(i)})$ are, therefore, the order statistics of a random sample of size n from a uniform (0, 1). Consequently (equation 1A.6), the variances and the covariance of $u_{(r)}$ and $u_{(n-r+1)}$, r = [0.4 + 1/2], are of order $0(n^{-1})$. Thus, $F(z_{(n-r+1)}) - F(z_{(r)})$ converges to its expected value $P = F(t_2) - F(t_1)$ very quickly as n becomes large; $F(t_2) = 1 - r/(n+1)$ and $F(t_1) = r/(n+1)$; $t_2 = -t_1 = t$. We consider the linear approximations

$$g(z_{(r)}) \approx \alpha_1 + \beta_1 z_{(r)} \quad \text{and} \quad g(z_{(n-r+1)}) \approx \alpha_2 - \beta_2 z_{(n-r+1)}. \tag{7.6.6}$$

To determine the values of α_2 and β_2 we realize that $z_{(n-r+1)}$ for fixed q = r/n will be located in the interval $(0, t_2)$ at any rate for large n. Substituting 0 and t_2 in (7.6.6) gives

$$\alpha_2 = \alpha = f(0)/P \text{ and } \beta_2 = \beta = [f(0) - f(t)]/tP \quad (t = t_2) \tag{7.6.7}$$

where $f(0) = (2\pi)^{-1/2}$ and $f(t) = (2\pi)^{-1/2}\exp(-t^2/2)$. Realize that β is always positive. Similarly, we obtain $\alpha_1 = \alpha$ and $\beta_1 = \beta$.

Incorporating (7.6.6) in (7.6.3)-(7.6.4) gives the following modified likelihood equations:

$$\frac{\partial \ln L}{\partial \mu} \approx \frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\left(\sum_{i=1}^{r} z_{(i)} + \sum_{i=n-r+1}^{n} z_{(i)}\right) + \frac{(n-2r)}{\sigma}\beta(z_{(r+1)} + z_{(n-r+1)}) = 0 \tag{7.6.8}$$

and

$$\frac{\partial \ln L}{\partial \sigma} \approx \frac{\partial \ln L^*}{\partial \sigma} = -\frac{2r}{\sigma} + \frac{1}{\sigma}\left(\sum_{i=1}^{r} z_{(i)}^2 + \sum_{i=n-r+1}^{n} z_{(i)}^2\right) + \frac{(n-2r)}{\sigma}[z_{(r)}(\alpha + \beta z_{(r)}) - z_{(n-r+1)}(\alpha - \beta z_{(n-r+1)})] = 0 \tag{7.6.9}$$

The claim here is not that the modified likelihood equations (7.6.8)-(7.6.9) are asymptotically equivalent to the likelihood equations (7.6.3)-(7.6.4). However, the differences between the two are very small since (7.6.6) are close approximations.

MML estimators: The solutions of (7.6.8)-(7.6.9) are the MML estimators:

$$\hat{\mu}_c = \left[\sum_{i=1}^{r} y_{(i)} + \sum_{i=n-r+1}^{n} y_{(i)} + (n-2r)\beta(y_{(r)} + y_{(n-r+1)})\right]/m, \tag{7.6.10}$$

$$m = 2r + 2(n-2r)\beta$$

and

$$\hat{\sigma}_c = [B + \sqrt{(B^2 + 4AC)}]/2\sqrt{A(A-1)}, \quad A = 2r; \tag{7.6.11}$$

$$B = (n-2r)\alpha(y_{(n-r+1)} - y_{(r)})$$

and

$$C = \sum_{i=1}^{r}(y_{(i)} - \hat{\mu})^2 + \sum_{i=n-r+1}^{n}(y_{(i)} - \hat{\mu})^2 + (n-2r)\beta\{(y_{(r)} - \hat{\mu})^2 + (y_{(n-r+1)} - \hat{\mu})^2\}$$

$$= \sum_{i=1}^{r} y_{(i)}^2 + \sum_{i=n-r+1}^{n} y_{(i)}^2 + (n-2r)\beta(y_{(r)}^2 + y_{(n-r+1)}^2) - m\hat{\mu}^2. \tag{7.6.12}$$

It may be noted that $\hat{\sigma}_c$ is always positive since β is positive. As said earlier, r is chosen to be the integer value [0.4n + 1/2] to achieve robustness to inliers and/or short-tails (Chapter 8).

Efficiency: For the normal $N(\mu, \sigma^2)$, the sample mean $\bar{y} = (1/n)\sum_{i=1}^{n} y_i$ is the MVB estimator with variance $V(\bar{y}) = \sigma^2/n$. The MML estimator $\hat{\mu}_c$ is also an unbiased estimator of μ; follows from symmetry. Its variance is given by

$$V(\hat{\mu}_c) = (l'\,\Omega l)\sigma^2/m^2 \tag{7.6.13}$$

where $l' = (1, 1,..., 1 + (n - 2r)\beta, 0,...,0, 1 + (n - 2r)\beta, 1,...,1)$ is a vector with n elements and Ω is the variance-covariance matrix of the standardized normal variates $z_{(i)}$ $(1 \le i \le n)$. Given in Table 7.2 are the simulated means of $\hat{\sigma}_c$ and $s = \sqrt{\left\{\sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)\right\}}$ and the simulated variance of s, and the relative efficiencies of $\hat{\mu}_c$ and $\hat{\sigma}_c$,

$$E_1 = 100\{V(\bar{y})/V(\hat{\mu}_c)\} \quad \text{and} \quad E_2 = 100\{V(s)/V(\hat{\sigma}_c)\} \tag{7.6.14}$$

for n = 10, 20, 30, 50 and 100. The values of E_1 for $n \le 20$ are exact, calculated from (7.6.13). It is seen that $\hat{\mu}_c$ is only marginally less efficient than the MVB estimator $\bar{y}$. Like s, $\hat{\sigma}_c$ has negligible bias but is somewhat less efficient. For short-tailed symmetric distributions and in situations when the sample contains inliers , $\hat{\mu}_c$ and $\hat{\sigma}_c$ are considerably more efficient than $\bar{y}$ and s, respectively, (Chapter 8). See also Table 7.3.

Table 7.2: Means of $\hat{\sigma}_c$ and s and the relative efficiencies E_1 and E_2 of $\hat{\mu}_c$ and $\hat{\sigma}_c$.

	n = 6	10	20	30	50	100
$(1/\sigma)E(s)$	0.962	0.966	0.982	0.989	0.995	0.997
$(1/\sigma)E(\hat{\sigma}_c)$	0.952	0.969	0.984	0.989	0.995	0.997
$(n/\sigma^2)\,V(s)$	0.573	0.536	0.535	0.538	0.532	0.553
E_1	94.7	98.6	98.7	98.7	98.9	98.9
E_2	93.3	96.3	94.3	94.1	93.1	93.1

Remark: The modified likelihood equation (7.6.8) when re-organized assumes the form

$$\frac{\partial \ln L}{\partial \mu} \approx \frac{\partial \ln L^*}{\partial \mu} = \frac{\sigma^2}{m}(\hat{\mu}_c - \mu); \tag{7.6.15}$$

m is given in (7.6.10). Since the difference between $\partial \ln L/\partial\mu$ and $\partial \ln L^*/\partial\mu$ is very small but not exactly zero, σ^2/m turned out to be a little larger than the exact variance of $\hat{\mu}_c$. The asymptotic value of $-E(\partial^2 \ln L/\partial\mu^2)$ suggests the formula $V(\hat{\mu}_c) \approx \sigma^2/(1.06m)$ and this gives accurate approximations. For example, we have the following values:

	n = 6	10	20	30	40	50	100
n/(1.06m)	1.138	1.050	1.049	1.049	1.049	1.049	1.049
n(variance)*	1.055	1.014	1.034	1.036	1.038	1.039	1.044

* Simulated for n > 20.

Incidentally, for n = 6 the simulated values are $E(\hat{\sigma}) = 0.962\sigma$ and $E(\sigma) = 0.952\sigma$; $E_1 = 94$ and $E_2 = 93$.

Hypothesis testing: To test $H_0 : \mu = 0$, the proposed statistic is (Akkaya and Tiku, 2003)

$$T = \sqrt{1.06m}\,(\hat{\mu}_c/\hat{\sigma}_c); \qquad (7.6.16)$$

m, $\hat{\mu}_c$ and $\hat{\sigma}_c$ are given in (7.6.10)-(7.6.11). Large values of T lead to the rejection of H_0 in favour of $H_1 : \mu > 0$. Since $\hat{\sigma}_c$ converges to σ and $\hat{\mu}_c$ is a linear function of order statistics, the null distribution of T is asymptotically normal N(0, 1). For small n, the null distribution of T is closely approximated by the Student t with v = n – 1 degrees of freedom. For example, we have the following simulated values of the probabilities $Prob(T \geq t_{0.05}(v) \mid H_0)$ and $Prob(t \geq t_{0.05}(v) \mid H_0)$, $t = \sqrt{n}\,(\bar{y}/s)$ being the Student t statistic:

	n = 10	20	30	40	50	100
T	0.049	0.051	0.051	0.050	0.047	0.054
t	0.051	0.051	0.052	0.050	0.048	0.051

For the normal population, the T test has slightly less power than the t test. For short-tailed symmetric (STS) distributions and in situations when the sample contains inliers, the T test has considerably higher power than the t test. This is illustrated in Chapter 8; see also Table 7.3.

M-estimators: We now show that the Huber M-estimators are less efficient than $\bar{y}$ and s and much less efficient than $\hat{\mu}_c$ and $\hat{\sigma}_c$ for STS distributions. Consider, for example, the Tukey symmetric lambda distributions defined by the transformation (Joiner and Rosenblatt, 1971)

$$z = [u^l - (1 - u)^l]/l \qquad (7.6.17)$$

where u is uniform (0, 1). The variance of the distribution is

$$\mu_2 = 2[1 - (l/2)\,\beta(l, l)]/l^2(2l + 1). \qquad (7.6.18)$$

We take l = 0.585 and 1.45 in which case the distributions are STS with kurtosis μ_4/μ_2^2 equal to 2 and 1.75, respectively. Given in Table 7.3 are the simulated means of the MML estimator $\hat{\sigma}_c$, the sample standard deviation s, and the w24 estimator $\hat{\sigma}_w$. The means of $\hat{\mu}_c$, $\bar{y}$ and $\hat{\mu}_w$ are zero (follows from symmetry). Since $V(\bar{y}) = \sigma^2/n$ we only give the relative efficiencies $100\{V(\hat{\mu}_c)/V(\bar{y})\}$ and $100\{V(\hat{\mu}_c)/V(\hat{\mu}_w)\}$ of $\bar{y}$ and $\hat{\mu}_w$, relative to the MML estimator $\hat{\mu}_c$. We also give the values of the variance $V(\hat{\sigma}_c)$ and the relative efficiencies $100\{V(\hat{\sigma}_c)/V(s)\}$ and $100\{V(\hat{\sigma}_c)/V(\hat{\sigma}_w)\}$. The random values generated from (7.6.17) were divided by $\sqrt{\mu_2}$. Therefore, all the three estimators $\hat{\sigma}_c$, s and $\hat{\sigma}_w$ are estimating σ (taken to be 1 without loss of generality):

Table 7.3: Means and relative efficiencies for the STS distributions

	l = 0.585				l = 1.45			
	n = 10	20	40	100	10	20	40	100
$E(\hat{\sigma}_c)$	0.962	0.969	0.973	0.973	0.955	0.955	0.955	0.959
E(s)	0.984	0.993	0.999	1.000	0.987	0.993	0.996	0.999
$E(\hat{\sigma}_w)$	0.974	1.019	1.036	1.041	0.979	1.030	1.044	1.050

$nV(\hat{\sigma}_c)$	0.285	0.248	0.235	0.220	0.209	0.184	0.164	0.156
$RE(\bar{y})$	95.4	94.2	93.6	93.7	90.9	90.1	90.4	88.4
$RE(\hat{\mu}_w)$	75.9	81.8	83.9	84.2	67.3	75.2	79.6	79.7
RE(s)	89.0	89.6	92.5	92.6	82.6	79.8	77.5	85.6
$RE(\hat{\sigma}_w)$	66.1	70.5	75.7	78.2	54.7	59.8	63.3	73.4

The M-estimators BS82 and H22 have essentially the same relative efficiencies as w24 above. It can be seen that $\bar{y}$ and s are less efficient than $\hat{\mu}_c$ and $\hat{\sigma}_c$. The M-estimators are less efficient than $\bar{y}$ and s and much less efficient than $\hat{\mu}_c$ and $\hat{\sigma}_c$.

Akkaya and Tiku (2003) have similar results for numerous other STS distributions, for example, the family (3.61), and the symmetric distributions with cdf

$$\begin{aligned} F(z) &= 2^{k-1} z^k, \quad 0 < z < 0.5 \\ &= 1 - 2^{k-1}(1-z)^k, \quad 0.5 < z < 1 \end{aligned} \tag{7.6.19}$$

with k = 1.5, 2.0 and 3.0 in which case the kurtosis is 2.123, 2.400 and 2.856, respectively. They have also similar results for the STS distributions with cdf

$$\begin{aligned} F(z) &= 0.5 - 2^{k-1}(0.5-z)^k, \quad 0 < z < 0.5 \\ &= 0.5 + 2^{k-1}(z-0.5)^k, \quad 0.5 < z < 1 \end{aligned} \tag{7.6.20}$$

with k = 1.5 and 2.0 in which case the kurtosis is 1.486 and 1.330, respectively. Clearly, the M-estimators are inefficient for STS distributions as said earlier in Chapter 2.

7.7 RAYLEIGH DISTRIBUTION

We have shown that for censored samples from $N(\mu, \sigma^2)$, the MMLE are asymptotically the same as the MLE. We have also shown that for small n, the MMLE are numerically the same as the MLE (almost). The MMLE have the beauty that they are explicit functions of sample observations and are easy to compute. We now consider another important distribution, the Rayleigh distribution. The distribution has the density function

$$f(y) = \frac{2}{\sigma} y e^{-y^2/\sigma}, \quad 0 < y < \infty, \tag{7.7.1}$$

and arises in electromagnetic wave propagation through a scattering medium, communication engineering, and so on (Siddiqui, 1962; Ariyawansa and Templeton, 1984). Realize that $x = y^2/\sigma$ has the exponential distribution E(0, 1). It may be noted that the scale of y is $\sqrt{\sigma}$ and that is the parameter of particular interest.

Suppose that the censored sample (7.2.1) comes from the Rayleigh distribution above. The log-likelihood function is

$$\ln L = \text{Constant} - \frac{1}{2}(n - r_1 - r_2)\ln\sigma + \sum_{i=r_1+1}^{n-r_2} \ln z_{(i)} - \sum_{i=r_1+1}^{n-r_2} z_{(i)}^2 + r_1 \ln F(z_{(r_1+1)}) + r_2 \ln[1 - F(z_{(n-r_2)})] \tag{7.7.2}$$

where $z_{(i)} = y_{(i)}/\sqrt{\sigma}$, $F(z) = \int_0^z f(z)\,dz = 1 - e^{-z^2}$ and $f(z) = 2ze^{-z^2}$, $0 < z < \infty$. The likelihood equation for estimating σ is

$$\frac{d\ln L}{d\sigma} = \frac{1}{\sigma}\left[-(n-r_1-r_2) + \sum_{i=r_1+1}^{n-r_2} z_{(i)}{}^2 - \frac{1}{2} r_1 z_{(r_1+1)} g(z_{(r_1+1)}) + r_2 z_{(n-r_2)}{}^2\right] = 0 \qquad (7.7.3)$$

$$g(z) = f(z)/F(z) = 2ze^{-z^2}/(1-e^{-z^2}).$$

The equation (7.7.3) does not admit an explicit solution. Lee et al. (1980) adopted an iterative procedure to solve it but, as expected, found it too time consuming and slowly converging. With a little bit of heavy censoring, in fact, they encountered divergence of iterations even with the actual population parameter as a starting value (Lee et al., 1980, p.16).

To avoid these pitfalls, we derive the MMLE of σ by using the linear approximation

$$g(z_{(r_1+1)}) \cong \alpha - \beta z_{(r_1+1)}. \qquad (7.7.4)$$

From the first two terms of a Taylor series expansion around $h_1 = E(z_{(r_1+1)})$, we obtain (Tiku et al., 1986)

$$\beta = -\left[\frac{d}{dz} g(z)\right]_{z=h_1} = -\frac{2e^{-h_1{}^2}}{(1-e^{-h_1{}^2})}\left[1 - \frac{2h_1{}^2}{1-e^{-h_1{}^2}}\right]$$

and

$$\alpha = \frac{f(h)}{q_1} + \beta h_1 \ (q_1 = r_1/n). \qquad (7.7.5)$$

It is not difficult to show that (Lee et al., 1980, p.19)

$$E(z_{(i)}) = \frac{n!}{(i-1)!(n-i)!} \sum_{j=0}^{i-1} \binom{i-1}{j} (-1)^j \frac{\sqrt{\pi}}{2(n-i+j+1)^{3/2}} \qquad (7.7.6)$$

which is, in fact, the equation (2.8.6) with p = 2. For n $\geq$ 20, the approximate value of h_1 determined from the following equation is used:

$$F(h_1) = \frac{r_1}{n+1} \quad \text{which gives} \quad h_1 = \left[-\ln\left(1 - \frac{r_1}{n+1}\right)\right]^{1/2}. \qquad (7.7.7)$$

Incorporating (7.7.4) in (7.7.3) gives the modified likelihood equation:

$$\frac{\partial \ln L}{\partial\sigma} \cong \frac{\partial \ln L^*}{\partial\sigma} = \frac{1}{\sigma}\left[-(n-r_1-r_2) + \sum_{i=r_1+1}^{n-r_2} z_{(i)}{}^2 - \frac{1}{2} r_1 z_{(r_1+1)}(\alpha - \beta z_{(r_1+1)}) + r_2 z^2_{(n-r_2)}\right]$$

$$= -\frac{1}{2\sigma^2}(2A\sigma + B\sqrt{\sigma} - C) = 0 \quad (A = n - r_1 - r_2). \qquad (7.7.8)$$

The positive root of (7.7.8) gives the MML estimator of $\sqrt{\sigma}$,

$$\sqrt{\hat{\sigma}_c} = (-B + \sqrt{B^2 + 8AC})/4A; \qquad (7.7.9)$$

$B = r_1 \alpha y_{(r_1+1)}$ and $C = 2\sum_{i=r_{1+1}}^{n-r_2} y_{(i)}{}^2 + r_1 \beta y^2_{(r_1+1)} + 2r_2 y^2_{(n-r_2)}$. The MML estimator of σ is, therefore,

$$\hat{\sigma}_c = [(B^2 + 4AC) - B(B^2 + 8AC)^{1/2}]/8A^2. \qquad (7.7.10)$$

As n tends to infinity, $z_{(r_1+1)}$ converges to h_1(r_1 or $q_1 = r_1/n$ fixed) and since g(z) is a bounded function over $0 < z < \infty$, $g(z_{(r_1+1)})$ converges to $f(h_1)/q_1$. Consequently, (7.7.4) is an equality. The MML estimator $\hat{\sigma}_c$ is, therefore, asymptotically equivalent to the ML estimator.

Asymptotic properties: The estimator $\hat{\sigma}_c$ is asymptotically unbiased. This follows from the first two terms of the Taylor series expansion of E(d ln L*/dσ). That gives

$$E(\hat{\sigma}_c) = \sigma + E\left(\frac{d \ln L^*}{d\sigma}\right)/R^2(\sigma) \qquad (7.7.11)$$

where $R^2(\sigma) = -E(d^2 \ln L^*/d\sigma^2)$. From (7.7.8),

$$-E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right) = \frac{A}{\sigma^2}\left[1 + \frac{1}{4A} r_1 \alpha E(z_{(r_1+1)})\right]. \qquad (7.7.12)$$

Since d ln L*/dσ is asymptotically equivalent to d ln L/dσ and the expected value of the latter is zero, it follows that E(d ln L*/dσ) = 0 (asymptotically). It then follows from (7.7.11) that $\hat{\sigma}_c$ is asymptotically unbiased.

Since $E(z_{(r_1+1)}) \cong h_1$ for large n, we obtain the asymptotic variance

$$V(\hat{\sigma}_c) \cong \frac{\sigma^2}{A}\left[1 + \frac{1}{4A} r_1 \alpha h_1\right]^{-1}. \qquad (7.7.13)$$

The exact values of $E(z_{(r_1+1)})$ can be obtained from (7.7.6) and used in (7.7.12). The variances of $\hat{\sigma}_c$ so obtained are not much different from (7.7.13).

Realize that the variance of $\hat{\sigma}_c$ increases much faster with increasing r_1 than with r_2. Censoring of smallest observations should, therefore, be avoided as much as possible.

Lee et al. (1980) examined the accuracy of the linear approximation (7.7.4). They found it very satisfactory for large n but deficit for small n. For small n, n ≤ 10 particularly, they suggested a second-linear-approximation by using the MMLE. This procedure starts with using (7.7.4) and obtains the MMLE $\hat{\sigma}$. Now expand $g(z_{(r_1+1)})$ in a Taylor series around $\hat{z}_{(r_1+1)} = y_{(r_1+1)}/\sqrt{\hat{\sigma}}$ to obtain

$$g(z_{(r_1+1)}) \cong g(\hat{z}_{(r_1+1)}) + (z_{(r_1+1)} - \hat{z}_{(r_1+1)})\left[\frac{d}{dz} g(z)\right]_{z=\hat{z}_{(r_1+1)}}$$

$$= \alpha^* - \beta^* z_{(r_1+1)} \qquad (7.7.14)$$

where $\beta^* = -[dg(z)/dz]_{z=\hat{z}_{(r_1+1)}}$ and $\alpha^* = g(\hat{z}_{(r_1+1)}) + \hat{z}_{(r_1+1)}\beta^*$. (7.7.15)

The new coefficients α* and β* are then substituted in (7.7.9), replacing α by α* and β and β*, and the revised estimator $\sqrt{\hat{\sigma}_1}$ obtained. Realize that $\sqrt{\hat{\sigma}_1}$ is an explicit function of sample observations and is much easier to compute than the ML estimator. Lee et al. (1980) found the agreement between $\hat{\sigma}_1$ and the MML estimator $\hat{\hat{\sigma}}$ amazingly close. We reproduce their values in Table 7.4; $\hat{\sigma}_c$ is the first-linear-approximation estimator (see also Tiku et al, 1986, pp. 93-98):

Table 7.4: Comparing the improved MMLE $\hat{\sigma}_1$ with the MLE $\hat{\hat{\sigma}}$.

q_1	q_2	$\hat{\sigma}_1$	$\hat{\hat{\sigma}}$	$\hat{\sigma}_1$	$\hat{\hat{\sigma}}$	$\hat{\sigma}_c$
		n = 10			n = 30	
0.1	0.1	2.5133	2.5123	2.94108	2.94108	2.934
	0.2	2.0379	2.0369	2.39997	2.39997	2.380
0.2	0.1	2.5337	2.5335	2.95687	2.95687	2.938
	0.2	2.0605	2.0603	2.41740	2.41738	2.390
0.3	0.1	2.5335	2.5335	2.95151	2.95148	2.985
	0.2	2.0603	2.0603	2.41096	2.41096	2.411

Alternative estimator: Let $y_{(i)}$ be the order statistics of $y_i (1 \le i \le n)$. Since $x = y^2$ is distributed as exponential $E(0, \sigma)$, we write $x_{(i)} = y_{(i)}{}^2 (r_1 + 1 \le i \le n - r_2)$ and define the sample spacings as

$$D_i = (n - i)\{x_{(i+1)} - x_{(i)}\}, \; r_1 + 1 \le i \le n - r_2 - 1;$$

D_i are iid exponential $E(0, \sigma)$. The ML estimator of σ is

$$\hat{\sigma}_c{}^* = \sum_{i=r_1+1}^{n-r_2-1} D_i \,/\, (n - r_1 - r_2 - 1) \tag{7.7.16}$$

with $E(\hat{\sigma}_c{}^*) = \sigma$ and variance $V(\hat{\sigma}_c{}^*) = \sigma^2/(n - r_1 - r_2 - 1)$.

Asymptotically, $\sqrt{\hat{\sigma}_c}$ and $\sqrt{\hat{\sigma}_c{}^*}$ are both unbiased estimators of $\sqrt{\sigma}$ and

$$V(\sqrt{\hat{\sigma}_c}) \cong \frac{\sigma}{4A}\left[1 + \frac{1}{4}\,\frac{r_1 \alpha h_1}{A}\right]^{-1} \quad \text{and} \quad V(\sqrt{\hat{\sigma}_c{}^*}) \cong \frac{\sigma}{4(A-1)}; \tag{7.7.17}$$

this follows from equation (1.2.11) in Chapter 1. The first variance is smaller than the second since α and h_1 are both positive.

The estimator $\sqrt{\hat{\sigma}_c}$ is somewhat more efficient than $\sqrt{\hat{\sigma}_c{}^*}$ even asymptotically, although the latter is easier to compute.

7.8 CENSORED SAMPLES FROM LTS DISTRIBUTIONS

Consider now the situation when the censored sample (7.2.1) comes from a distribution in the LTS family (2.2.9). Here, the likelihood equations are (Tiku and Suresh, 1992)

$$\frac{\partial \ln L}{\partial \mu} = \frac{2p}{k\sigma} \sum_{i=r_1+1}^{n-r_2} g(z_{(i)}) - \frac{r_1}{\sigma} h_1(z_{(r_1+1)}) + \frac{r_2}{\sigma} h_2(z_{(n-r_2)}) = 0 \tag{7.8.1}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{A}{\sigma} + \frac{2p}{k\sigma} \sum_{i=r_1+1}^{n-r_2} z_{(i)} g(z_{(i)}) - \frac{r_1}{\sigma} z_{(r_1+1)} h_1(z_{(r_1+1)})$$

$$+ \frac{r_2}{\sigma} z_{(n-r_2)} h_2(z_{(n-r_2)}) = 0, \tag{7.8.2}$$

$A = n - r_1 - r_2$ and $z_{(i)} = (y_{(i)} - \mu)/\sigma$. The functions g, h_1 and h_2 are $g(z) = z/\{1 + (1/k)z^2\}$, $h_1(z) = f(z)/\{1 - F(z)\}$ and $h_2(z) = f(z)/F(z)$;

$$F(z) = \int_{-\infty}^{z} f(z)\,dz \text{ and } f(z) = \{1 + (1/k)z^2\}^{-p}/\sqrt{k}\beta(1/2, p - 1/2). \tag{7.8.3}$$

The equations (7.8.1)-(7.8.2) are almost impossible to solve even by iteration.

To obtain the modified likelihood equations, we use the linear approximation (2.3.13) for $g(z_{(i)})$. We also use the linear approximations

$$h_1(z_{(r_1+1)}) \cong a_1 - b_1 z_{(r_1+1)} \quad \text{and} \quad h_2(z_{(n-r_2)}) \cong a_2 + b_2 z_{(n-r_2)}. \tag{7.8.4}$$

The coefficients (a_1, b_1) and (a_2, b_2) are obtained from the first two terms of Taylor series expansions of $h_1(z)$ and $h_2(z)$ around $t_1 = E(z_{(r_1+1)})$ and $t_2 = E(z_{(n-r_2)})$, respectively. For $n \geq 10$, the approximate values of t_1 and t_2 obtained from the equations $F(t_1) = r_1/n$ and $F(t_2) = 1 - r_2/n$ may be used (Tiku and Suresh, 1992).

It is easy to show that (a_1, b_1) and (a_2, b_2) are given by the following equations with $q = r_1/n$ and $q = r_2/n$, respectively:

$$b = -\frac{f(t)}{q}\left\{\frac{2p}{k}g(t) - \frac{f(t)}{q}\right\} \quad \text{and} \quad a = \frac{f(t)}{q} - bt \tag{7.8.5}$$

where t is determined by $F(t) = \int_{-\infty}^{t} f(z)\,dz = 1 - q$; g(z) and f(z) are given in (7.8.3). It is easy to find the value of t since $\sqrt{v/k}Z$ has the Student t distribution with $v = 2p - 1$ degrees of freedom; $k = 2p - 3$ and $p \geq 2$. For $1 \leq p < 2$, k is equated to 1 in all the equations above in which case σ is simply a scale parameter.

MML estimators: Incorporating the linear approximations above in (7.8.1) – (7.8.2), we obtain modified likelihood equations which are exactly of the same form as (7.3.8) – (7.3.9). The solutions of these equations are the MML estimators:

$$\hat{\mu}_c = K + D\hat{\sigma}_c \quad \text{and} \quad \hat{\sigma}_c = (B + \sqrt{B^2 + 4AC})/2\sqrt{A(A-1)}; \tag{7.8.6}$$

$$A = n - r_1 - r_2, \quad M = (2p/k)\sum_{i=r_1+1}^{n-r_2}\beta_i + r_1 b_1 + r_2 b_2,$$

$$D = \left[(2p/k)\sum_{i=r_1+1}^{n-r_2}\alpha_i - r_1 a_1 + r_2 a_2\right]/M,$$

$$K = \left[(2p/k)\sum_{i=r_1+1}^{n-r_2}\beta_i y_{(i)} + r_1 b_1 y_{(r_1+1)} + r_2 b_2 y_{(n-r_2)}\right]/M, \tag{7.8.7}$$

$$B = (2p/k)\sum_{i=r_1+1}^{n-r_2}\alpha_i(y_{(i)} - K) - r_1 a_1(y_{(r_1+1)} - K) + r_2 a_2(y_{(n-r_2)} - K),$$

and

$$C = (2p/k)\sum_{i=r_1+1}^{n-r_2}\beta_i(y_{(i)} - K)^2 + r_1 a_1(y_{(r_1+1)} - K)^2 + r_2 b_2(y_{(n-r_2)} - K)^2$$

$$= (2p/k)\sum_{i=r_1+1}^{n-r_2}\beta_i y_{(i)}^2 + r_1 b_1 y_{(r_1+1)}^2 + r_2 b_2 y^2_{(n-r_2)} - MK^2;$$

$D = 0$ if $r_1 = r_2 = r$. Asymptotically (r_1 and r_2 or q_1 and q_2 fixed and n tends to infinity), the MML estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ are equivalent to the ML estimators.

Since $\partial \ln L^*/\partial\mu$ when reorganized assumes the form

$$\frac{\partial \ln L}{\partial\mu} \cong \frac{\partial \ln L^*}{\partial\mu} = \frac{M}{\sigma^2}(K + D\sigma - \mu), \tag{7.8.8}$$

$\hat{\mu}_c(\sigma) = K + D\sigma$ is conditionally (σ known) the MVB estimator for large $n - r_1 - r_2$ and

$$V(\hat{\mu}_c(\sigma)) \cong \sigma^2/M. \tag{7.8.9}$$

For symmetric censoring $D = 0$ in which case $\hat{\mu}_c = K$ and the expressions (7.8.7) simplify, i.e.,

$$A = n - 2r,\ M = (2p/k)\sum_{i=r+1}^{n-r}\beta_i + 2rb,$$

$$K = \left[(2p/k)\sum_{i=r+1}^{n-r}\beta_i y_{(i)} + rb(y_{(r+1)} + y_{(n-r)})\right]/M, \tag{7.8.10}$$

$$B = (2p/k)\sum_{i=r+1}^{n-r}\alpha_i y_{(i)} + ra(y_{(n-r)} - y_{(r+1)})$$

and

$$C = (2p/k)\sum_{i=r+1}^{n-r}\beta_i (y_{(i)} - K)^2 + rb\{(y_{(r+1)} - K)^2 + (y_{(n-r)} - K)^2\}$$

$$= (2p/k)\sum_{i=r+1}^{n-r}\beta_i y_{(i)}^2 + rb({y_{(r+1)}}^2 + {y_{(n-r)}}^2) - MK^2.$$

The following result is important for hypothesis testing. Realize that $\hat{\mu}_c$ is unbiased for all $n - 2r$; this follows from symmetry.

Theorem 7.1: For large A (r or r/n fixed), the estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ (based on symmetric censored samples) have the following asymptotic properties:

(a) $\hat{\mu}_c$ is the MVB estimator of μ with $V(\hat{\mu}_c) \cong \sigma^2/M$ (M given in the equation 7.8.10) and is normally distributed,

(b) $\hat{\mu}_c$ and $\hat{\sigma}_c$ are independently distributed, and

(c) $(A - 1){\hat{\sigma}_c}^2/\sigma^2$ is distributed as chi-square with $A - 1$ degrees of freedom.

Proof: The results follow from the following representations of the modified likelihood equations:

$$\frac{\partial \ln L}{\partial\mu} \cong \frac{\partial \ln L^*}{\partial\mu} = \frac{\sigma^2}{M}(\hat{\mu}_c - \mu) \tag{7.8.11}$$

with $E(\partial^{r+s} \ln L^*/\partial\mu^r\partial\sigma^s) = 0$ for all $r \geq 1$ and $s \geq 1$, and

$$\frac{\partial \ln L}{\partial\sigma} \cong \frac{\partial \ln L^*}{\partial\sigma} = \frac{A}{\sigma^3}\left[\frac{C(\mu)}{A} - \sigma^2\right] \tag{7.8.12}$$

where $C(\mu)$ is exactly the same as (7.8.10) with K replaced by μ and

$$C(\mu) = M(\hat{\mu} - \mu)^2 + C. \tag{7.8.13}$$

Remark: The chi-square approximations similar to (2.11.17) and (2.11.19) are more accurate than (c) above.

To have an idea about how efficient $\hat{\mu}_c$ is and how closely σ^2/M approximates the true variance, we give below in Table 7.5 the values of (a) 1/M, (b) $(1/\sigma^2)V(\hat{\mu}_c)$ and (c) $(1/\sigma^2)V(\mu_c{*})$; $r_1 = r_2 = r$, p = 2. The exact variances of $\hat{\mu}_c$ and the BLUE $\mu_c{*}$ were obtained from equations exactly similar to (7.6.13):

Table 7.5: Exact variances of MMLE and BLUE for symmetric censored samples, p=2.

	n = 10		n = 20			
	r = 1	r = 2	r = 1	r = 2	r = 3	r = 4
(a)	0.0513	0.0508	0.0255	0.0255	0.0255	0.0254
(b)	0.0550	0.0549	0.0262	0.0263	0.0263	0.0263
(c)	0.0549	0.0549	0.0262	0.0263	0.0263	0.0263

For p > 2 the values (a), (b) and (c) are closer to one another than in Table 7.5. The MML estimator is clearly highly efficient and σ^2/M provides close approximations to the true variances.

It is interesting to note that in Table 7.5 the variance of $\hat{\mu}_c$ calculated from a sample with a proportion q = 0.2 of extreme observations censored is only marginally bigger than the variance of $\hat{\mu}$ calculated from a complete sample. For p = 2, n = 10 and $q = q_1 = q_2 = 0.2$, for example, $(1/\sigma^2)V(\hat{\mu}_c) = 0.0549$. For p=2, n = 10 and q = 0, $(1/\sigma^2)V(\hat{\mu}_c) = 0.0546$. We show in Chapter 8 that $\hat{\mu}_c$ and $\hat{\mu}$ are both robust, the former being robust to more extreme deviations from an assumed distribution in the LTS family (2.2.9).

In an interesting paper, Vaughan (1992a) evaluated the efficiencies of the MML estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$. Since for small p (≤ 3 particularly), the first few and the last few coefficients β_i in (2.3.14) can be negative as said earlier, he determined the value of r such that β_i are positive for all $r + 1 \le i \le n - r$. This ensures that $\hat{\sigma}_c$ is always real and positive. He also noted that censoring these r smallest and largest observations does not affect too adversely the efficiencies either of the MMLE or the BLUE μ^* and σ^*. He showed that both $\hat{\mu}_c$ and $\hat{\sigma}_c$ are highly efficient. He gave two examples to illustrate the closeness of the MLE and MMLE. We reproduce his results in Examples 7.5 and 7.6; see also Tiku et al. (1986, p.86).

7.9 VARIANCES AND COVARIANCES

The asymptotic variances and the covariance of $\hat{\mu}_c$ and $\hat{\sigma}_c$ in (7.8.6) are given by $I^{-1}(\mu, \sigma)$, where the elements of I are

$$I_{11} = -E\left(\frac{\partial^2 \ln L^*}{\partial \mu^2}\right) = \frac{M}{\sigma^2}$$

$$I_{12} = I_{21} = -E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}\right) = -\frac{M}{\sigma^2} D$$

$$= -\frac{1}{\sigma^2}\left[\frac{2p}{k}\sum_{i=r_1+1}^{n-r_2}\alpha_i - r_1 a_1 + r_2 a_2\right] \tag{7.9.1}$$

$$I_{22} = -E\left(\frac{\partial^2 \ln L^*}{\partial\sigma^2}\right) = \frac{2A}{\sigma^2}\left[1 - \frac{1}{2A}E\left\{\frac{2p}{k}\sum_{i=r_1+1}^{n-r_2}\alpha_i z_{(i)} - r_1 a_1 z_{(r_1+1)} + r_2 a_2 z_{(n-r_2)}\right\}\right];$$

(7.9.1) are computed from the expected values of $z_{(i)}$ given in Tiku and Kumra (1981). It is interesting indeed to note that $E(z_{(i)}{}^2)$ are not required to calculate (7.9.1). For $r_1 = r_2$ (symmetric censored samples), of course, $I_{12} = 0$.

For symmetric censored samples, the MVB(σ) is taken to be (Vaughan, 1992a, p.463; Hill, 1995)

$$I_{22}{}^{-1} = \text{MVB}(\sigma) \cong (p + 1)\sigma^2/\, 2A(p - 1/2)\ (p \geq 1), \quad A = n - 2r; \tag{7.9.2}$$

see also the equation (2.4.9). A straightforward extension of this result for $r_1 \neq r_2$ gives

$$I_{22}{}^{-1} = 1/\{-E(\partial^2 \ln L^*/\partial\sigma^2)\} \cong (p+1)\sigma^2/2A(p - 1/2), \quad A = n - r_1 - r_2. \tag{7.9.3}$$

Writing $N = 2A(p - 1/2)/(p + 1)$, the information matrix is

$$I(\mu, \sigma) = \frac{1}{\sigma^2}\begin{pmatrix} M & -MD \\ -MD & N \end{pmatrix}. \tag{7.9.4}$$

The matrix (7.9.4) gives the following variances and the covariance for large $A = n - r_1 - r_2$:

$$V(\hat{\mu}_c) \cong \frac{\sigma^2}{M}\left(1 - \frac{M}{N}D^2\right)^{-1}, \quad V(\hat{\sigma}_c) \cong \frac{\sigma^2}{N}\left(1 - \frac{M}{N}D^2\right)^{-1}$$

and

$$\text{Cov}(\hat{\mu}_c, \hat{\sigma}_c) \cong \frac{D}{N}\left(1 - \frac{M}{N}D^2\right)^{-1}. \tag{7.9.5}$$

Example 7.5: Sarhan and Greenberg (1962, p.212) present the results of an experiment measuring the concentration of Strontium-90 in milk, which was assumed to contain 9.22 μC/liter. Ten measurements were made, with the 3 largest and 2 smallest discarded as being unreliable. The remaining observations in order are

8.2 8.4 9.1 9.8 9.9

The underlying distribution is assumed to be logistic (see also Tiku et al., 1986, p.86). As said earlier, the logistic distribution when standardized with respect to its standard deviation $\sqrt{3.29}$ is virtually indistinguishable from the distribution (2.2.9) with p = 5 (i.e., a constant multiple of the Student t with 9 degrees of freedom). Here, we have the following estimates and their standard errors. The ML estimate and its standard error are based on the results of Harter and Moore (1968, p.682). The MML estimate based on the logistic distribution are from Tiku (1968, p.74); the standard errors are worked out by interpolation in his Table 3. The first three estimates and their standard errors (within brackets) are based on the logistic distribution. The last estimate is based on the Student t distribution with 9 degrees of freedom:

Estimate	ML	MML	BLUE	MML
μ	9.27 (±0.491)	9.27 (±0.512)	9.29 (±0.532)	9.24 (±0.515)
σ	1.57 (±0.618)	1.63 (±0.634)	1.69(±0.723)	1.62 (±0.693)

Even for this small data set, the ML and the MML estimates and their standard errors are close to one another. Realize that the computation of the MML estimates (being explicit functions of sample observations) is much easier than the ML estimates. The computation of the BLUE requires expected values and the variances and covariances of order statistics (and matrix inversions) and is much involved. The MLE have to be computed iteratively.

Example 7.6: Vaughan (1992a, pp.463-464) has 20 observations from the Cauchy distribution $1/\{\pi(1 + y^2)\}$, $-\infty < y < \infty$. Arranged in ascending order, the observations are

– 5.137	– 3.709	– 3.245	– 3.061	– 2.928	– 2.127	– 0.334
– 0.282	– 0.199	– 0.072	0.186	0.319	0.332	0.556
0.696	0.848	0.961	1.165	3.220	12.621	

He calculates five different estimators and their standard errors. The results are given below:

Estimate of	Median	MMLE	BLUE
μ	0.057(±0.568)	0.088(±0.353)	0.056(±0.265)
σ	0.518(±1.47)	0.863(±0.501)	0.722(±0.330)
Estimate of	Quantile	T-S MMLE	
μ	– 0.297(±0.682)	0.051(±0.247)	
σ	1.656(±1.566)	0.663(±0.267)	

For the median estimator of σ, he uses $(y_{11} - y_{10})/(t_{(11)} - t_{(10)})$ which is unbiased for σ. The MMLE are those in (7.4.8) with r = 6. The BLUE are based on the middle 16 observations as recommended by Barnett (1966b). The Quantile estimates are from Chan (1970) and are very unsatisfactory. And the T – S MMLE are those in (7.8.10) with r = 7. He concludes that the T – S MMLE have the smallest standard errors and are, therefore, most efficient. Barnett (1966b) discusses the difficulties with the maximum likelihood estimation for such data.

It may be noted that the BLUE and the T – S MMLE are close to one another, the latter having somewhat smaller standard errors. Like the Quantile estimates, the Median estimates have large standard errors; both are highly inefficient.

7.10 HYPOTHESIS TESTING

To test H_0: $\mu = 0$ against H_1: $\mu > 0$, we define for symmetric censored samples the statistic

$$T = \sqrt{M}(\hat{\mu}_c / \hat{\sigma}_c)\,; \tag{7.10.1}$$

M, $\hat{\mu}_c$ and $\hat{\sigma}_c$ are given in (7.8.10). Large values of T lead to rejection of H_0 in favour of H_1. To ensure that $\hat{\sigma}_c$ is real and positive when p is small ($p \le 3$), r in (7.8.10) is chosen as in Vaughan (1992a, Table III) unless it is already larger than the values given below:

p	n = 10	15	20	p	n = 10	15	20	p	n = 10	15	20
		r				r				r	
1.0	3	5	7	1.5	2	3	4	2.0	1	2	3
2.5	1	1	1	3.0	0	1	1	3.5	0	0	0

For large A = n – 2r (r or r/q fixed), the null distribution of T is normal N(0, 1). For $A \ge 20$, in fact, the standard normal provides accurate approximations for the percentage points of T.

To find the null distribution of T for small A, we use the chi-square approximation (2.11.17). In view of the fact that $E(\hat{\sigma}_c/\sigma) \cong 1$ and from the equation (7.9.3)

$$h = 1 + \{(p + 1)/2A(p - 1/2)\}, \tag{7.10.2}$$

the null distribution of

$$T = \sqrt{M}(\hat{\mu}_c \sqrt{h})/\hat{\sigma}_c \tag{7.10.3}$$

is referred to the Student t with A-1 degrees of freedom. See also Vaughan (1992a, p.465) who gives a very accurate approximation in terms of the Student t distribution.

7.11 TYPE I CENSORING

In a random sample of size $N = n_1 + n + n_2$ from a normal population $N(\mu, \sigma^2)$, n observations assume values between two known limits u' and u'' and their numerical values $y_i (1 \le i \le n)$ are available. However, n_1 observations are less than u' and n_2 observations are greater than u'' and their numerical values are not available; n_1 and n_2 are random (N is fixed). The likelihood function of this Type I censored sample is

$$L \propto \sigma^{-n} [F(z')]^{n_1} [1 - F(z'')]^{n_2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right); \tag{7.11.1}$$

$z_i = (y_i - \mu)/\sigma$, $z' = (u' - \mu)/\sigma$ and $z'' = (u'' - \mu)/\sigma$. The likelihood equations for estimating μ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\sum_{i=1}^{n} z_i - \frac{n_1}{\sigma} g_1(z') + \frac{n_2}{\sigma} g_2(z'') = 0 \tag{7.11.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma}\sum_{i=1}^{n} z_i^2 - \frac{n_1}{\sigma} z' g_1(z') + \frac{n_2}{\sigma} z'' g_2(z'') = 0; \tag{7.11.3}$$

$$g_1(z) = f(z)/F(z) \quad \text{and} \quad g_2(z) = f(z)/[1 - F(z)]; \tag{7.11.4}$$

$f(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ and $F(z) = \int_{-\infty}^{z} f(z)\,dz$. The equations (7.11.2) – (7.11.3) have no explicit solutions. Specialized tables and nomographs are required to compute the ML estimators from these equations; see, for example, Cohen (1957; 1991) and Schneider (1986, pp.230-242).

To obtain the modified likelihood equations we realize that for large n, z' is very likely to be covered by the interval (a_1, b_1) where

$$a_1 = \{u' - (\bar{y} + s/\sqrt{n})\}/s \quad \text{and} \quad b_1 = \{u' - (\bar{y} - s/\sqrt{n})\}/s, \tag{7.11.5}$$

$$\bar{y} = \sum_{i=1}^{n} y_i/n \quad \text{and} \quad s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1).$$

Of course, $\bar{y}$ and s^2 are biased estimators of μ and σ^2, but we are only interested in z' being covered by (a_1, b_1), not that it should necessarily be the mid-point of this interval. Similarly, z'' is very likely to be covered by the interval (a_2, b_2) where a_2 and b_2 are exactly the same as a_1 and b_1, respectively, with u' replaced by u''.

To obtain modified likelihood equations, we use the linear approximations

$$g_1(z') \cong \alpha_1 - \beta_1 z' \quad \text{and} \quad g_2(z'') = \alpha_2 + \beta_2 z'', \tag{7.11.6}$$

where $\beta_1 = -\{g_1(b_1) - g_1(a_1)\}/(b_1 - a_1)$ and $\alpha_1 = g_1(a_1) + \beta_1 a_1$, (7.11.7)

and $\beta_2 = -\{g_2(b_2) - g_2(a_2)\}/(b_2 - a_2)$ and $\alpha_2 = g_2(a_2) - \beta_2 a_2$. (7.11.8)

Realize that as n tends to infinity, $a_1 \cong b_1 = (u' - \bar{y})/s$ in which case $\beta_1 = -\left\{\frac{d}{dz} g_1(z)\right\}$ evaluated at $z = (u' - \bar{y})/s$ and, similarly, β_2.

In practice, however, the values of z' and z'' will hardly ever be known since μ and σ are not known.

Substituting (7.11.6) in (7.11.2) – (7.11.3) gives the modified likelihood equations,

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma}\left[\sum_{i=1}^{n} z_i - n_1(\alpha_1 - \beta_1 z') + n_2(\alpha_2 + \beta_2 z'')\right] \quad (7.11.9)$$

$$= \frac{m}{\sigma^2}(K + D\sigma - \mu)$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{1}{\sigma}\left[-n + \sum_{i=1}^{n} z_i^2 - n_1 z'(\alpha_1 - \beta_1 z') + n_2 z''(\alpha_2 + \beta_2 z'')\right] \quad (7.11.10)$$

$$= -\frac{1}{\sigma^3}\left[(n\sigma^2 - B\sigma - C) - m(K - \mu)(K + D\sigma - \mu)\right] = 0.$$

Calculations show that (7.11.6) are close approximations, at any rate for large n. Consequently, the differences between the likelihood and the modified likelihood equations are very small. See also Section 7.13.

The solutions of (7.11.9) – (7.11.10) are the MML estimators:

$$\hat{\mu} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = [B + \sqrt{(B^2 + 4nC)}]/2n; \quad (7.11.11)$$

$$K = \left(\sum_{i=1}^{n} y_i + n_1\beta_1 u' + n_2\beta_2 u''\right)/m, \quad D = (n_2\alpha_2 - n_1\alpha_1)/m,$$

$$m = n + n_1\beta_1 + n_2\beta_2, \quad B = n_2\alpha_2(u'' - K) - n_1\alpha_1(u' - K)$$

and

$$C = \sum_{i=1}^{n} (y_i - K)^2 + n_1\beta_1(u' - K)^2 + n_2\beta_2(u'' - K) \quad (7.11.12)$$

$$= \sum_{i=1}^{n} y_i^2 + n_1\beta_1 u'^2 + n_2\beta_2 u''^2 - mK^2.$$

Realize that n_1 and n_2 are random variables with $(N = n_1 + n + n_2)$

$$E(n) = N[F(z'') - F(z')], \; E(n_1) = NF(z') \quad \text{and} \quad E(n_2) = N[1 - F(z'')]. \quad (7.11.13)$$

For large n, $\hat{\mu}$ and $\hat{\sigma}$ are unbiased; see also Section 7.13.

Covariance matrix: Differentiating the modified likelihood equations (7.11.9) – (7.11.10), we obtain

$$-(\partial^2 \ln L^*/\partial\mu^2) = m/\sigma^2 \quad (m = n + n_1\beta_1 + n_2\beta_2)$$

$$-(\partial^2 \ln L^*/\partial\mu\partial\sigma) = -(n_2\alpha_2 - n_1\alpha_1)/\sigma^2 \quad (7.11.14)$$

and

$$-(\partial^2 \ln L^*/\partial\sigma^2) = \frac{2n}{\sigma^2}\left[1 - \frac{1}{2n}(n_2\alpha_2 z'' - n_1\alpha_1 z')\right].$$

Realize that in practice n_1/n and n_2/n are small. Following Cox (1961, 1962), Schneider (1968) and Tiku (1968b), an estimate of the asymptotic covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is given by I^{-1} evaluated at $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$. See the next section for an example illustrating the remarkable closeness of the MML and the ML estimates. Also, the expected values of I_{ij} in (7.11.14) are exactly similar to (7.3.15).

As in Section 7.5, the technique above readily generalizes to k Type I censored samples in the framework of Experimental Design; see also Section 7.13.

7.12 PROGRESSIVELY CENSORED SAMPLES

To quote Cohen (1963, p.328), a progressively censored sample is defined as follows:

"Let N designate the total sample size and n the number of sample specimens which fail and therefore result in completely determined life spans. Suppose that censoring occurs progressively in k stages at times T_i $(l \le i \le k)$ and that at the i^{th} stage of censoring r_i sample specimens selected randomly from the survivors at time T_i are censored from further observations. It follows that

$$N = n + \sum_{i=1}^{k} r_i .$$

In Type I censoring, the T_i are fixed and the number of survivors at these times are random variables. In Type II censoring, the T_i coincide with times of failure and are random variables, whereas the number of survivors at these times are fixed. For both types, r_i is fixed".

Let $y_1, y_1, \ldots, y_N$ be a k-stage progressively Type I censored sample from a normal population $N(\mu, \sigma^2)$. The samples of this kind are encountered in life and fatigue studies in the context of life-test experiments (Cohen, 1963). Here,

$$\ln L = \text{Constant} - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 + \sum_{i=1}^{k} r_i \ln[1 - F(w_i)] \qquad (7.12.1)$$

where $w_j = (T_j - \mu)/\sigma$ $(1 \le j \le k)$ and $F(w) = (2\pi)^{-1/2} \int_{-\infty}^{w} \exp(-z^2/2)dz$.

Writing $z_i = (y_i - \mu)/\sigma$,

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma}\left[\sum_{i=1}^{n} z_i + \sum_{j=1}^{k} r_j g(w_j)\right] = 0$$

and

$$\frac{\partial \ln L}{\partial \sigma} = \frac{1}{\sigma}\left[-n + \sum_{i=1}^{n} z_i^2 + \sum_{j=1}^{k} r_j w_j g(w_j)\right] = 0 . \qquad (7.12.2)$$

It is not easy to solve these equations (Cohen, 1963).

To obtain the modified likelihood equations, we use the linear approximations

$$g(w_j) \cong \alpha_j + \beta_j w_j \quad (1 \le j \le k) \qquad (7.12.3)$$

where α_j and β_j are obtained from (7.11.8) with a_2 and b_2 replaced by

$$a_{2j} = \{ T_j - (\bar{y} + s/\sqrt{n})\}/s \quad \text{and} \quad b_{2j} = \{T_j - (\bar{y} - s/\sqrt{n})\}/s \quad (1 \le j \le k). \qquad (7.12.4)$$

Incorporating (7.12.4) in (7.12.2)-(7.12.3) gives the modified likelihood equations

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{1}{\sigma} \sum_{i=1}^{n} z_i^2 + \sum_{j=1}^{k} r_j w_j(\alpha_j + \beta_j w_j)$$

$$= \frac{m}{\sigma^2}(K + D\sigma - \mu) = 0$$

and

$$\frac{\partial \ln L}{\partial \sigma} \cong \frac{\partial \ln L^*}{\partial \sigma} = \frac{1}{\sigma}\left[-n + \sum_{i=1}^{n} z_i^2 + \sum_{j=1}^{k} r_j w_j(\alpha_j + \beta_j w_j)\right]$$

$$= \frac{-1}{\sigma^3}\,[(n\sigma^2 - B\sigma - C) - m(K - \mu)(K + D\sigma - \mu)] = 0. \qquad (7.12.5)$$

These equations are similar to those in (7.11.9)-(7.11.10). Their solutions are the following MML estimators:

$$\hat{\mu} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \{B + \sqrt{(B^2 + 4nC)}\}/2n; \tag{7.12.6}$$

$$K = \left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{k} r_j\beta_j T_j\right)\Big/m, \quad m = n + \sum_{j=1}^{k} r_j \beta_j, \quad D = \left(\sum_{j=1}^{k} r_j\alpha_j\right)\Big/m,$$

$$B = \sum_{j=1}^{k} r_j\alpha_j\,(T_j - K) \text{ and } C = \sum_{i=1}^{n}(y_i - K)^2 + \sum_{j=1}^{k} r_j\beta_j(T_j - K)^2 \tag{7.12.7}$$

$$= \sum_{i=1}^{n} y_i^2 + \sum_{j=1}^{k} r_j\alpha_j T_j^2 - mK^2 .$$

Notice the beauty of these expressions. They are, in fact, amended versions of those given in Tiku et al. (1986, p.100).

Covariance matrix: From (7.12.5), we obtain the following second derivatives

$$-E(\partial^2 \ln L^*/\partial\mu^2) = m/\sigma^2, \quad -E(\partial^2 \ln L^*/\partial\mu\partial\sigma) = -\sum_{j=1}^{k} r_j\alpha_j/\sigma^2$$

and

$$-E(\partial^2 \ln L^*/\partial\sigma^2) = \frac{2n}{\sigma^2}\left[1 - \frac{1}{2n}\sum_{j=1}^{k} r_j\alpha_j\left(\frac{T_j - \mu}{\sigma}\right)\right]. \tag{7.12.8}$$

Their values are evaluated at $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$ to give the estimated information matrix. The inverse of this matrix gives the asymptotic variances and the covariance (estimates).

Remark: The MML estimators for progressive Type II censored samples are exactly the same as (7.12.6) – (7.12.7) with (α_j, β_j) calculated from (7.3.6) with q_2 replaced by $q_{2j} = j/n$ or $j/(n + 1)$, $1 \le j \le k$.

Example 7.7: A total of 316 specimens were placed under observation and their life spans recorded in days (Cohen, 1963, p.337). Ten specimens were censored after 36.5 days and 10 more censored after 44.5 days. Data for this progressive Type I censored sample are as follows:

$$N = 316,\ n = 296,\ T_1 = 36.5,\ T_2 = 44.5,\ \bar{y} = 39.2703 \text{ and } s^2 = 20.1634.$$

Here (using the values of the normal density f(z) and its cdf F(z) given in Biometrika Tables: Vol.2, pp.153-155)

$$\alpha_1 = 0.803,\ \beta_1 = 0.478;\ \alpha_2 = 0.820,\ \beta_2 = 0.800;\ m = 308.78,$$
$$K = 296(39.2703 + 0.5894 + 1.2027)/308.78 = 39.3629;$$
$$D = 0.0526,\ B = 19.1351,\ C = 6221.4168;$$
$$\hat{\mu} = 39.606 \quad \text{and} \quad \hat{\sigma} = 4.617$$

with the following variances and the covariance:

$$V(\hat{\mu}) = 0.0691 \quad Cov(\hat{\mu}, \hat{\sigma}) = 0.0019 \quad V(\hat{\sigma}) = 0.0363.$$

Cohen (1963) obtained the following ML estimates iteratively:

$$\hat{\mu} = 39.583 \quad \text{and} \quad \hat{\sigma} = 4.611, \qquad \text{with}$$
$$V(\hat{\mu}) = 0.069 \quad Cov(\hat{\mu}, \hat{\sigma}) = 0.002 \quad V(\hat{\sigma}) = 0.036$$

The agreement between the two methods is remarkable, as expected. The MMLE are, however, very easy to calculate. Unlike the MLE, no specialized tables and nomographs are required to compute the MMLE.

Some of the interesting references dealing with estimation problems in life testing are the following: Bain (1972, 1978), Barlow and Proschan (1975), Bartholomew (1957), Basu and Ghosh (1980), Billman et al. (1972), Cox (1959), Davis (1952), Dixon (1960), Engelhardt and Bain (1974b, 1979), Epstein (1960a, b), Epstein and Tsao (1953), Fleming and Harrington (1980), Gehan and Thomas (1969), Govindarajulu (1964), Harter (1969), Harter and Moore (1966, 1967, 1968, 1976), Irwin (1942), Johnson (1974), Lagakos (1979), Lawless (1971, 1977), Lawless and Singhal (1980), Mann (1969, 1972), Mann and Singapurwalla (1983), McCool (1982), Mantel and Myers (1971), Mendenhall and Hader (1958), Nelson (1972), Tallis and Light (1968), Wolynetz (1974), Wright et al. (1978), and Zacks (1971).

7.13 TRUNCATED NORMAL DISTRIBUTION

A normal variate assumes values on the real line IR: $-\infty < y < \infty$. In some practical situations, however, y is restricted to the range $u' < y < u''$; u' and u'' are known limits. The distribution in this situation is the truncated normal

$$f_T(y) \propto e^{-(y-\mu)^2/2\sigma^2}/[F(z'') - F(z')], \quad u' < y < u'', \tag{7.13.1}$$

where $z' = (u' - \mu)/\sigma$ and $z'' = (u'' - \mu)/\sigma$ and $F(z) = \int_{-\infty}^{z} f(z)\,dz$ is the cdf of the standard normal $f(z) = (2\pi)^{-1/2}\exp(-z^2/2)$, $-\infty < z < \infty$.

Given a random sample $y_1, y_2, \ldots, y_n$, one wants to estimate μ and σ. To find the ML estimators, the likelihood function is

$$L \propto \sigma^{-n}\, e^{-\sum_{i=1}^{n} z_i^2/2}\, [F(z'') - F(z')]^{-n},$$

$z_i = (y_i - \mu)/\sigma$. The likelihood equations for estimating μ and σ are

$$\frac{\partial \ln L}{\partial \mu} = \frac{n}{\sigma}\left[\frac{1}{n}\sum_{i=1}^{n} z_1 - g_1(z') + g_2(z'')\right] = 0 \tag{7.13.2}$$

and

$$\frac{\partial \ln L}{\partial \sigma} = \frac{n}{\sigma}\left[-1 + \frac{1}{n}\sum_{i=1}^{n} z_i^2 - z' g_1(z') + z'' g_2(z'')\right] = 0$$

where

$$g_1(z') = f(z')/[F(z'') - (F(z')] \quad \text{and} \quad g_2(z'') = f(z'')/[F(z'') - F(z')]. \tag{7.13.3}$$

Specialized tables and nomographs are needed to solve these equations; see, for example, Schneider (1968, pp. 230-242).

To obtain the modified likelihood equations, we linearize the intractable functions $g_1(z')$ and $g_2(z'')$:

$$g_1(z') \cong \alpha_1 - \beta_1 z' \quad \text{and} \quad g_2(z'') \cong \alpha_2 + \beta_2 z''. \tag{7.13.4}$$

The coefficients (α_1, β_1) and (α_2, β_2) are obtained as in (7.11.6). Here, $a = \bar{y} + s/\sqrt{n}$ and $b = \bar{y} - s/\sqrt{n}$, $\bar{y} = \sum_{i=1}^{n} y_i/n$ and $s^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$, and

$$\beta_1 = \frac{s}{b-a}\left[\frac{f\left(\frac{u'-b}{s}\right)}{D(b)} - \frac{f\left(\frac{u'-a}{s}\right)}{D(a)}\right] \quad \text{and} \quad \alpha_1 = \frac{f\left(\frac{u'-a}{s}\right)}{D(a)} + \beta_1\left(\frac{u'-a}{s}\right) \tag{7.13.5}$$

and, similarly,

$$\beta_2 = \frac{s}{b-a}\left[\frac{f\left(\frac{u''-b}{s}\right)}{D(b)} - \frac{f\left(\frac{u''-a}{s}\right)}{D(a)}\right] \quad \text{and} \quad \alpha_2 = \frac{f\left(\frac{u''-a}{s}\right)}{D(a)} - \beta_1\left(\frac{u''-a}{s}\right) \qquad (7.13.6)$$

$$D(h) = F\left(\frac{u''-h}{s}\right) - F\left(\frac{u'-h}{s}\right);$$

D(a) and D(b) are equal to D(h) with h replaced by a and b, respectively. It is easy to compute (7.13.5) – (7.13.6) since tables of f(z) and F(z) are readily available. Alternatively, an IMSL subroutine may be used to evaluate the cdf F(z) of a standard normal N(0, 1).

Incorporating (7.13.4) in (7.13.2) gives the modified likelihood equations:

$$\frac{\partial \ln L}{\partial \mu} \approx \frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma}\left[\frac{1}{n}\sum_{i=1}^{n} z_i - (\alpha_1 - \beta_1 z') + (\alpha_2 + \beta_2 z'')\right] \qquad (7.13.7)$$

$$= \frac{m}{\sigma^2}(K + D\sigma - \mu) = 0$$

and

$$\frac{\partial \ln L}{\partial \sigma} \approx \frac{\partial \ln L^*}{\partial \sigma} = \frac{m}{\sigma}\left[-1 + \frac{1}{n}\sum_{i=1}^{n} z_i^2 - z'(\alpha_1 - \beta_1 z') + z''(\alpha_2 + \beta_2 z'')\right] \qquad (7.13.8)$$

$$= -\frac{1}{\sigma^3}[(\sigma^2 - B\sigma - C) - m(K-\mu)(K + D\sigma - \mu)] = 0.$$

The solutions of these equations are the following MML estimators:

$$\hat{\mu} = K + D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \{B + \sqrt{(B^2 + 4C)}\}/2; \qquad (7.13.9)$$

$K = (\bar{y} + \beta_1 u' + \beta_2 u'')/m$, $m = 1 + \beta_1 + \beta_2$, $D = (\alpha_2 - \alpha_1)/m$,

$B = \alpha_2(u'' - K) - \alpha_1(u' - K)$,

$$C = (1/n)\sum_{i=1}^{n}(y_i - K)^2 + \beta_1(u' - K)^2 + \beta_2(u'' - K)^2 \qquad (7.13.10)$$

$$= (1/n)\sum_{i=1}^{n} y_i^2 + \beta_1 u'^2 + \beta_2 u''^2 - mK^2.$$

Notice the beauty and simplicity of these equations.

Covariance matrix: The elements of the asymptotic covariance matrix are given by

$$-E(\partial^2 \ln L^*/\partial\mu^2) = m/\sigma^2, \quad -E(\partial^2 \ln L^*/\partial\mu\partial\sigma) = -(\alpha_2 - \alpha_1)/\sigma^2$$

and

$$-E(\partial^2 \ln L^*/\partial\sigma^2) = \frac{2n}{\sigma^2}\left[1 - \frac{1}{2n}(\alpha_2 z'' - \alpha_1 z')\right]. \qquad (7.13.11)$$

Their values are evaluated at $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$. The inverse of the matrix gives estimates of the variances and the covariance of $\hat{\mu}$ and $\hat{\sigma}$.

Accuracy: To appreciate the accuracy of the linear approximations (7.13.4) we note that as n becomes large, b – a tends to zero in which case

$$\beta_1 \cong \beta_1^* = -\left\{\frac{d}{dz} g_1(z)\right\}_{z=z^*} = \frac{z^* f(z^*)}{F(z^{**}) - F(z^*)}$$

and

$$\alpha_1^* = \frac{f(z^*)}{F(z^{**}) - F(z^*)} + \beta_1^* z^* \qquad (7.13.12)$$

and, similarly, $\beta_2 \cong \beta_2^*$ and $\alpha_2 \cong \alpha_2^*$. Now, we have the equations

$$\lim_{n\to\infty} \frac{1}{n} E\left(\sum_{i=1}^{n} z_i\right) = \int_{z'}^{z''} zh(z)\,dz = g_1(z') - g_2(z'')$$

and
$$\lim_{n\to\infty} \frac{1}{n} E\left(\sum_{i=1}^{n} z_i^2\right) = \int_{z'}^{z''} z^2 h(z)\,dz = 1 + z'g_1(z') - z''g_2(z''). \tag{7.13.13}$$

Consequently,

$$z^* = \frac{z' - [g_1(z') - g_2(z'')]}{\{1 + z'g_1(z') - z''g_2(z'') - [g_1(z') - g_2(z'')]^2\}^{1/2}}$$

and z** is similar and has z″ in the numerator replacing z′. Given below are the values of the differences

$$(1)\ \ g_1(z') - (\alpha_1^* - \beta^* z') \quad (2) \quad g_2(z'') - (\alpha_2^* + \beta_2^* z'') \tag{7.13.14}$$

for numerous truncation points z′ and z″ (Tiku and Stewart, 1977, p. 1492):

Table 7.6: Values of the difference between g(z) and its linear approximation.

z′	z″	F(z″) – F(z′)	$g_1(z')$	(1)	$g_2(z')$	(2)
– ∞	2.32	0.99	0	0	0.0273	0.001
	1.96	0.975	0	0	0.0569	– 0.000
	1.64	0.95	0	0	0.1095	0.012
	1.28	0.90	0	0	0.1954	0.032
	0.84	0.80	0	0	0.3506	0.084
– 2.32	2.32	0.98	0.0276	0.001	0.0276	0.001
	1.64	0.94	0.0278	0.005	0.1107	0.015
	0.84	0.79	0.0342	0.014	0.3551	0.096
– 1.64	1.28	0.85	0.1225	0.034	0.2070	0.055

The differences (1) and (2) are small if the truncated area T(z′, z″) =1– [F(z″) – F(z′)] ≤ 0.20. Consequently, the MML estimators are almost identical to the ML estimators.

If the truncated area T(z′, z″) is greater than 0.20, the MML estimators may be sharpened as follows:

In the first calculation use $\bar{y}$ and s in (7.13.5) – (7.13.6) and obtain the MML estimates $\hat{\mu}$ and $\hat{\sigma}$ from (7.13.9) – (7.13.10). Now use $\hat{\mu}$ and $\hat{\sigma}$ in (7.13.5) – (7.13.6) and obtain the revised estimates from (7.13.9) – (7.13.10). The process might be repeated one more time. The resulting estimates are almost identical to the ML estimates.

In practice, one might not know whether the truncated area is small or not. It is, therefore, advisable to do one or two iterations always. These iterations are easy to carry out since the MML estimators are explicit functions of sample observations.

To illustrate the closeness of the MML and the ML estimates, we have the following examples. The MMLE are the result of the first calculation. Second and third iteraritons make no substantive difference either in their values or their variances and covariance.

Example 7.8. Right Truncation: To ensure meeting a maximum weight specification of 12 oz on a certain radio component for an aircraft installation, all production of this component is weighed and those units that exceed the maximum weight are discarded. For a random sample of 50 units selected from the expected production, $\bar{y}$ = 9.35 oz and s^2 = 1.1494; n = 50 and u'' = 12.

Assuming that the weights of the components are normally distributed, we want to estimate its mean μ and standard deviation σ.

Here $\alpha_1 = 0$, $\beta_1 = 0$; $\alpha_2 = 0.12872$ and $\beta_2 = -0.04443$ (from equations 7.13.6).

Substituting these in (7.13.9) – (7.13.10), we obtain the MML estimates:

$$\hat{\mu} = 9.3737 \quad \text{and} \quad \hat{\sigma} = 1.0905$$

Their variances and the covariance obtained by inverting the information matrix evaluated at $\mu = \hat{\mu}$ and $\sigma = \hat{\sigma}$ are

$$V(\hat{\mu}) \cong 0.0252 \quad \text{Cov}(\hat{\mu}, \hat{\sigma}) \cong 0.0019 \quad V(\hat{\sigma}) \cong 0.0142$$

These may be compared to the following MLE given by Cohen (1959):

$$\hat{\hat{\mu}} = 9.38 \quad \text{and} \quad \hat{\hat{\sigma}} = 1.091$$

with $$V(\hat{\hat{\mu}}) \cong 0.027 \quad \text{Cov}(\hat{\hat{\mu}}, \hat{\hat{\sigma}}) \cong 0.004 \quad V(\hat{\hat{\sigma}}) \cong 0.016.$$

There is close agreement between the two.

Example 7.9. Double Truncation: The entire production of a certain bushing is sorted through go/no go gauges with the result that items of diameter in excess of 0.6015 inches and those less than 0.5985 inches are discarded. For a random sample of 75 bushings selected from the screened production, $\bar{y}$ = 0.60015 and $s^2 = 0.3762 \times 10^{-6}$; n = 75, u' = 0.5985 and u'' = 0.6015

Assuming the distribution to be normal $N(\mu, \sigma^2)$, we have

$$\alpha_1 = 0.08974, \beta_1 = -0.02918; \alpha_2 = 0.21326, \beta_2 = -0.08021.$$

The MML estimates are

$$\hat{\mu} = 0.60018 \quad \text{and} \quad \hat{\sigma} = 0.00068$$

with $$V(\hat{\mu}) \cong 0.6922 \times 10^{-8} \quad \text{Cov}(\hat{\mu}, \hat{\sigma}) \cong 0.063 \times 10^{-8} \quad V(\hat{\sigma}) \cong 0.452 \times 10^{-8}.$$

Cohen (1957) gives the following estimates

$$\hat{\hat{\mu}} = 0.60018 \quad \text{and} \quad \hat{\hat{\sigma}} = 0.00066$$

with $$V(\hat{\hat{\mu}}) \cong 0.6819 \times 10^{-8} \quad \text{Cov}(\hat{\hat{\mu}}, \hat{\hat{\sigma}}) \cong 0.099 \times 10^{-8} \quad V(\hat{\hat{\sigma}}) \cong 0.5471 \times 10^{-8}.$$

The agreement between the two methods is remarkably close, as in all the previous examples.

It may be noted that the modified likelihood equations based on Type I censored samples are exactly of the same form as those based on Type II censored samples. The information matrices are similar, both being asymptotically equivalent to the information matrices of the MLE. The MMLE have the clear advantage of being explicit and, therefore, easy to compute. They can also be implemented in more complex situations. We illustrate it by considering the truncated normal distribution in the context of Experimental Design as follows.

7.14 EXPERIMENTAL DESIGN WITH TRUNCATED NORMAL

Consider the one-way classification experimental design model (7.5.1). Here, the distribution of e_{ij} $(1 \le j \le n)$ is normal, truncated at u_i' and u_i'' $(1 \le i \le k)$, i.e., the distribution of e_{ij} $(1 \le j \le n)$ is

$$\frac{(\sqrt{2\pi}\sigma)^{-1}\exp\{-(y_{ij}-\mu-g_i)^2/2\sigma^2\}}{F(z_i'')-F(z_i')}, \quad u_1' < y_{ij} < u_i''. \tag{7.14.1}$$

The truncation points u_i' and u_i'' $(1 \le i \le k)$ are known. Realize that $\mu + g_i$ is the mode of the distribution (7.14.1); g_i may be called the ith group-effect. Writing

$$z_{ij} = (y_{ij} - \mu - g_i)/\sigma,\ z_1' = (u_i' - \mu - g_1)/\sigma \quad \text{and} \quad z_1'' = (u_i'' - \mu - g_i)/\sigma, \quad \text{and}$$
$$g_1(z') = f(z')/[F(z'') - F(z')] \quad \text{and} \quad g_2(z'') = f(z'')/[F(z'') - F(z')], \tag{7.14.2}$$

$$f(z) = (2\pi)^{-1/2}\exp(-z^2/2),\ F(z) = \int_{-\infty}^{z} f(z)\,dz, \tag{7.14.3}$$

the likelihood equations work out in terms of the intractable functions $g_1(z_i')$ and $g_2(z_i'')$, $1 \le i \le k$. Pearson and Lee (1908), Fisher (1931), Cohen (1955, 1963), Cohen and Woodword (1953), Halperin (1952), Samford and Taylor (1959), and Taylor (1973) devised procedures (involving some complicated functions) which are iterative in nature. They are indeed laborious and time consuming (Tiku and Stewart, 1977, pp.1486-1489).

To obtain the modified likelihood equations, we utilize the linear approximations

$$g_1(z_i') \cong \alpha_{1i} - \beta_{1i}z_i' \quad \text{and} \quad g_2(z_i'') \cong \alpha_{2i} + \beta_{2i}z_i'' \quad (1 \le i \le k) \tag{7.14.4}$$

where $(\alpha_{1i}, \beta_{1i})$ and $(\alpha_{2i}, \beta_{2i})$ are given by the equations (7.13.5) – (7.13.6) with (a, b) replaced by (a_i, b_i), and u' and u'' replaced by u_i' and u_i'', respectively:

$$a_i = \bar{y}_i + s_i/\sqrt{n} \quad \text{and} \quad b_i = \bar{y}_i - s_i/\sqrt{n},\ 1 \le i \le k; \tag{7.14.5}$$

$\bar{y}_i = \sum_{j=1}^{n} y_{ij}/n$ and $s_i^2 = \sum_{j=1}^{n}(y_{ij} - \bar{y}_i)/(n-1)$ are the sample means and variances.

For estimating μ, g_i and σ, the modified likelihood equations are

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{n}{\sigma}\sum_{i=1}^{k}\left\{\frac{1}{n}\sum_{j=1}^{n} z_{ij} - (\alpha_{1i} - \beta_{1i}z_i') + (\alpha_{2i} + \beta_{2i}z_i'')\right\} = 0$$

$$\frac{\partial \ln L}{\partial g_i} \cong \frac{\partial \ln L^*}{\partial g_i} = \frac{n}{\sigma}\left\{\frac{1}{n}\sum_{j=1}^{n} z_{ij} - (\alpha_{1i} - \beta_{1i}z_i') + (\alpha_{2i} + \beta_{2i}z_i'')\right\} = 0 \tag{7.14.6}$$

and
$$\frac{\partial \ln L}{\partial \sigma} = \frac{\partial \ln L^*}{\partial \sigma} = \frac{n}{\sigma}\sum_{i=1}^{k}\left\{-1 + \frac{1}{n}\sum_{j=1}^{n} z_{ij}^2 - z_i'(\alpha_{1i} - \beta_{1i}z_i') + z_1''(\alpha_{2i} - \beta_{2i}z_i'')\right\} = 0.$$

Under the constraint

$$\sum_{i=1}^{k} m_i g_i = 0 \quad (m_i = 1 + \beta_{1i} + \beta_{2i}) \tag{7.14.7}$$

the modified likelihood equations above give the following MMLE:

$$\hat{\mu} = K + D\hat{\sigma}, \quad \hat{g}_i = K_i + D_i\hat{\sigma} - \hat{\mu} \quad (1 \le i \le k) \tag{7.14.8}$$

and
$$\hat{\sigma} = \{B + \sqrt{(B^2 + 4kC)}\}/2k, \tag{7.14.9}$$

where $K = \sum_{i=1}^{k} m_i K_i/m \quad \left(m = \sum_{i=1}^{k} m_i\right), \quad D = \sum_{i=1}^{k} m_i D_i/m,$

$$K_i = (\bar{y}_i + \beta_{1i}u_i' + \beta_{2i}u_i'')/m_i,\ D_i = (\alpha_{2i} - \alpha_{1i})/m,$$
$$B_i = \alpha_{2i}(u_i'' - K_i) - \alpha_{1i}(u_i' - K_i) \ \text{and}$$
$$C_i = (1/n)\sum_{j=1}^{n}(y_{ij} - K_i)^2 + \beta_{1i}(u_i' - K_i)^2 + \beta_{2i}(u_i'' - K_i)^2 \tag{7.14.10}$$

$$= (1/n) \sum_{j=1}^{n} y_{ij}^{2} + \beta_{1i}u_i'^2 + \beta_{2i}u_i''^2 - m_i K_i^2;$$

$$B = \sum_{i=1}^{n} B_i \quad \text{and} \quad C = \sum_{i=1}^{k} C_i .$$

Remark: Since in practice the sizes of the truncated areas $F(z_i'') - F(z_i')$, $1 \le i \le k$, are not known, whether they are less than 0.20 or not as mentioned in the previous section, it is advisable to sharpen the MMLE by doing one or two iterations as explained earlier.

Asymptotic covariance matrix: In view of the equations (7.3.8) – (7.3.9), it is easy to work out the elements of the Fisher information matrix:

$$I_{11} = -E\left(\frac{\partial^2 \ln L^*}{\partial \mu^2}\right) = \frac{nm}{\sigma^2} \qquad I_{12} = -E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial g_i}\right) = \frac{nm_i}{\sigma^2} \tag{7.14.11}$$

$$I_{13} = -E\left(\frac{\partial^2 \ln L^*}{\partial \mu \partial \sigma}\right) = \frac{n}{\sigma^2}\left(\sum_{i=1}^{k} m_i D_i\right) \qquad I_{22} = -E\left(\frac{\partial^2 \ln L^*}{\partial g_i^{\,2}}\right) = \frac{nm_i}{\sigma^2}$$

$$I_{23} = -E\left(\frac{\partial^2 \ln L^*}{\partial g_i \partial \sigma}\right) = \frac{nm_i D_i}{\sigma^2}$$

and

$$I_{33} = -E\left(\frac{\partial^2 \ln L^*}{\partial \sigma^2}\right) = \frac{n}{\sigma^2}\sum_{i=1}^{k} (2 + \alpha_{1i}u_i' - \alpha_{2i}u_i''). \tag{7.14.12}$$

It is interesting to note the simplicity of these elements. The asymptotic covariance matrix is given by I^{-1}.

The bias in the MMLE $\hat{\mu}$, $\hat{g}_i$ and $\hat{\sigma}$ is negligible for large n, as illustrated in Section 7.13.

Linear contrast: To test the linear contrast

$$\theta = \sum_{i=1}^{k} l_i g_i = \sum_{i=1}^{k} l_i (\mu + g_i) = 0, \quad \sum_{i=1}^{k} l_i = 0, \tag{7.14.13}$$

the MML estimator of θ is

$$\hat{\theta} = \sum_{i=1}^{k} l_i \hat{\mu}_i, \quad \hat{\mu}_i = \hat{\mu} + \hat{g}_i = K_i + D_i\hat{\sigma}. \tag{7.14.14}$$

Realize that $\hat{\mu}_i$ $(1 \le i \le k)$ are independently distributed.

The asymptotic variances and covariance of $\hat{\mu}_i$ and $\hat{\sigma}$ are given by I^{-1}, where

$$I = \frac{n}{\sigma^2}\begin{bmatrix} m_i & \alpha_{2i} - \alpha_{1i} \\ \alpha_{2i} - \alpha_{1i} & \sum_{i=1}^{k} (2 + \alpha_{1i}z_i' - \alpha_{2i}z_1'') \end{bmatrix}. \tag{7.14.15}$$

Let ν_{ii} be the first element of I^{-1} evaluated at $\mu_i = \hat{\mu}_i$ and $\sigma = \hat{\sigma}$. To test H_0: $\theta = 0$, we define the statistic

$$W = \left(\sum_{i=1}^{k} l_i \hat{\mu}_i\right) \Big/ \hat{\sigma}\sqrt{\left(\sum_{i=1}^{k} l_i^{\,2}\, v_{ii}\right)}. \tag{7.14.16}$$

Large values of $|W|$ lead to the rejection of H_0 in favour of H_1: $\theta \neq 0$. For large n, the null distribution of W is referred to normal N(0,1).

Example 7.9: All production of a certain component is weighed and those units which exceed the maximum weight specification of 12 ounces are discarded. Two machines are used in the production and for random samples of 50 units selected from the two accepted productions,

$$\text{Machine I} : \bar{y}_1 = 5.3500 \quad \text{and} \quad s_1^2 = 1.1493$$
$$\text{Machine II} : \bar{y}_2 = 8.9800 \quad \text{and} \quad s_2^2 = 1.1512.$$

The problem is to test that $\mu_1 = \mu_2$, μ_1 and μ_2 being the modal values of the two productions. Here, we have the following ($\alpha_{1i} = \beta_{1i} = 0$; i = 1, 2)

	α_{2i}	β_{2i}	m_i	K_i	D_i	B_i	C_i
I:	0.1379	– 0.04775	0.95225	9.2171	0.14481	0.38381	0.77108
II:	0.0706	– 0.02220	0.97780	8.9114	0.07220	0.21808	0.92170

Substituting these values in (7.14.8) – (7.14.10), we obtain the estimates

$$\hat{\mu}_1 = 9.3739,\ \hat{\mu}_2 = 8.9896,\ \hat{\sigma} = 1.0827$$

The statistic ($l_1 = 1$ and $l_2 = -1$)

$$W = 0.3843/0.2209 = 1.74$$

which is bigger than 1.64. At 5% significance level, we reject H_0 in favour of H_1: $\mu_1 > \mu_2$ (Tiku and Stewart, 1977, p. 1499).

SUMMARY

In life-test experiments, Type I and Type II censored samples occur frequently. Type I censoring occurs when a sample of size n is drawn and observations below or above certain pre-determined limits are discarded. Type II censoring occurs if a pre-determined number of smallest or largest observations are discarded. Another important situation is when the sample comes from a truncated distribution. In this Chapter, we consider the three situations and work out the MMLE of the parameters. For illustration we consider the exponential, normal, Rayleigh, and the family (2.2.9) of long-tailed symmetric distributions. Other than for the exponential, the MLE are intractable; nomographs and specialized tables are required to compute them even under a normal distribution let alone non-normal distributions. The MMLE have expressions as beautiful as those for complete samples (Chapter 2). They are asymptotically fully efficient and have high efficiencies for small sample sizes. The MMLE are shown to be numerically the same (almost) as the MLE. We develop hypothesis testing procedures and show that they are straightforward analytically and computationally. We extend the results to experimental designs and to estimating and testing linear contrasts. Several real-life applications are given to illustrate the usefulness of the MMLE.

APPENDIX 7A

EXPONENTIAL SAMPLE SPACINGS

The joint distribution of the order statistics $y_{(i)}$ ($1 \le i \le n$) of a random sample of size n from the exponential distribution E(θ, σ) is

$$n!\sigma^{-n} e^{-\sum_{i=1}^{n}(y_{(i)} - \theta)/\sigma}, \quad \theta < y_{(1)} \le y_{(2)} \le \ldots.. \le y_{(n)} < \infty. \tag{7A.1}$$

Make the transformation

$$D_i = (n - i + 1)\,[y_{(i)} - y_{(i-1)}],\ y_{(0)} = \theta. \quad (7A.2)$$

The Jacobian is

$$1/J = |\partial D_i/\partial y_{(i)}| = n!$$

Substituting (7A.2) in (7A.1) and multiplying by J and realizing that

$$\sum_{i=1}^{n} (y_{(i)} - \theta) = \sum_{i=1}^{n} D_i\,,$$

the joint probability density function of $D_1, D_2, \ldots, D_n$ is

$$e^{-\sum_{i=1}^{n} D_i/\sigma}, \quad 0 < D_i < \infty, \quad (7A.3)$$

and that proves the result.

APPENDIX 7B

To obtain the null distribution of U, we use the following results:

(i) The statistics $y_{(r_1+1)}$, $y_{(r_2+1)}$ and $\hat{\sigma}_c$ are independently distributed.

(ii) The distribution of $2d\hat{\sigma}_c/\sigma$ is chi-square with $d = f - s_1 - s_2 - 2$ degrees of freedom; $f = n_1 + n_2 - r_1 - r_2$.

Theorem: The null distribution of $U(0 < U < \infty)$ is given by

$$f(u) = D\left[\sum_{l=0}^{r_2} (-1)^l \binom{r_2}{l} \beta(f + l, r_1 + 1)\left\{1 + \frac{(n_2 - r_2 + l)}{d} u\right\}^{-(d+1)} + \sum_{l=0}^{r_1} (-1)^l \binom{r_1}{l} \beta(f + l, r_2 + 1)\left\{1 + \frac{(n_1 - r_1 + l)}{d} U\right\}^{-(d+1)}\right];$$

$$D = 1/\{\beta(n_1 - r_1, r_1 + 1)\,\beta(n_2 - r_2, r_2 + 1)\},\ \beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b).$$

Proof. Write

$$a = r_1 + 1,\ b = r_2 + 1,\ A_1 = (y_{1,a} - \theta)/\sigma,\ A_2 = (y_{1,b} - \theta)/\sigma,$$

$$S_1 = \min(A_1, A_2),\ S_2 = \max(A_1, A_2),\ R = S_2 - S_1 \text{ and } W = d\hat{\sigma}_c/\sigma.$$

Then, $U = d(R/W)$. The joint distribution of S_1 and S_2 is from (i) above

$$De^{-\{(n_1 - r_1)S_1 + (n_2 - r_2)S_2\}}(1 - e^{-A_1})^{r_1}(1 - e^{-A_2})^{r_2}, \quad 0 < A_1, A_2 < \infty .$$

The joint distribution of S_1 and S_2 is

$$D\Big[e^{-\{(n_1 - r_1)S_1 + (n_2 - r_2)S_2\}}(1 - e^{-S_1})^{r_1}(1 - e^{-S_2})^{r_2} + e^{-\{(n_1 - r_1)S_2 + (n_2 - r_2)S_1\}}(1 - e^{-S_2})^{r_1}(1 - e^{-S_1})^{r_2}\Big], \quad 0 < S_1 < S_2 < \infty.$$

The joint distribution of R and $S = S_1$ is

$$[D[e^{-\{fS + (n_2 - r_2)R\}}(1 - e^{-S})^{r_1}\{1 - e^{-(S+R)}\}^{r_2} + e^{-\{fS + (n_1 - r_1)R\}}(1 - e^{-S})^{r_2}\{1 - e^{-(S+R)}\}^{r_1}], \quad 0 < R, S < \infty .$$

The joint distribution of R and $Z = \exp(-S)$ is

$$(D/Z)\Big\{Z^f e^{-(n_2 - r_2)R}(1 - Z)^{r_1}(1 - Ze^{-R})^{r_2} + Z^f e^{-(n_1 - r_1)R}(1 - Ze^{-R})^{r_1}(1 - Z)^{r_2}\Big\}; \quad 0 < R < \infty,\ 0 < Z < 1.$$

Integrating over Z from zero to 1 we obtain the distribution of R,

$$f(R) = D\Big[e^{-(n_2 - r_2)R} \sum_{l=0}^{r_2} (-1)^l \binom{r_2}{l} e^{-lR} \beta(f + l, r_1 + 1)$$

$$+ e^{-(n_1 - r_1)R} \sum_{l=0}^{r_1} (-1)^l \binom{r_1}{l} e^{-lR} \beta(f + l, r_2 + l)\Big], \quad 0 < R < \infty.$$

Since W is independently distributed as Gamma with parameter d, from (ii) above the joint distribution of R and W is

$$f(R)e^{-W}W^{d-1}/\Gamma(d), \quad 0 < R, W < \infty.$$

The joint distribution of U and W is

$$f(UW/d)e^{-W} W^{d-1}/\Gamma(d), \quad 0 < U, W < \infty.$$

Integrating out W, we get the distribution of U.

For $r_1 = r_2 = 0$, the distribution reduces to that of Kumar and Patel (1971).

Corollary: For $r_1 = r_2 = r$ and $n_1 = n_2 = n$, the results simplify, i.e.,

$$f(u) = \frac{2}{\{\beta(n - r, r + 1)\}^2} \left\{ \sum_{l=0}^{r} (-1)^l \binom{r}{l} \beta(2n - 2r + l, r + 1) \times (1 + hu)^{-(d+1)} \right\}, \quad 0 < u < \infty;$$

$$\text{Prob}(U \le x \mid H_0) = \frac{2}{\{\beta(n - r, r + 1)\}^2} \left\{ \sum_{l=0}^{r} (-1)^l \binom{r}{l} \beta(2n - 2r + l, r + 1) \times (1/hd)\{1 - (1 + hx)^{-d}\} \right\},$$

$$h = (n - r + l)/d \quad \text{and} \quad d = 2(n - r - 1) - s_1 - s_2.$$

Khatri (1981) gives the power function of the U test. The derivations are a little too complicated, however. We do not reproduce it here.

APPENDIX 7C

THE COEFFICIENTS FOR CENSORED NORMAL SAMPLES

For $q_1 = q$, $\alpha_1 = \alpha$ and $\beta_1 = \beta$. For $q_2 = q$, $\alpha_2 = \alpha$ and $\beta_2 = \beta$:

q	α	β	q	α	β
0	0	1	0.22	0.7497	0.7721
0.02	0.5954	0.8888	0.24	0.7564	0.7629
0.04	0.6319	0.8696	0.26	0.7626	0.7538
0.06	0.6562	0.8549	0.28	0.7682	0.7446
0.08	0.6748	0.8423	0.30	0.7733	0.7355
0.10	0.6902	0.8309	0.32	0.7779	0.7262
0.12	0.7033	0.8202	0.34	0.7820	0.7169
0.14	0.7147	0.8100	0.36	0.7856	0.7075
0.16	0.7249	0.8003	0.38	0.7888	0.6979
0.18	0.7340	0.7907	0.40	0.7915	0.6882
0.20	0.7422	0.7814	0.50	0.7979	0.6366

CHAPTER 8

Robustness of Estimators and Tests

8.1 INTRODUCTION

Two major problems in statistics are: (i) estimation of parameters and (ii) hypothesis testing. Traditionally, a particular mathematical form of the underlying distribution is assumed and optimal solutions of (i) and (ii) sought. Since in practice deviations are very common, one cannot feel comfortable with assuming a particular distribution and believing it to be exactly correct even if a great deal of thought is given to identifying the underlying distribution. That brings the issue of robustness in focus. An estimator is called robust if it is fully efficient (or nearly so) for an assumed distribution but maintains high efficiency for plausible alternatives. A fully efficient estimator has already been defined in Section 1.1 (Chapter 1). We would like our estimator(s) to be fully efficient for an assumed ditribution, at any rate for large n. We anticipate two types of alternatives to an assumed ditribution: (a) distributions which differ by moderate amounts and the differences are rather difficult to detect by graph plotting techniques or goodness-of-fit tests and (b) distributions which differ drastically and the differences are easy to detect (Chapter 9). Distributions of type (a) will be called plausible alternatives. We are not going to mix (a) with (b) since our concept of robustness is not that of a person who imagines walking over a rough terrain in darkness and has no feel where the cliff is: assume normal but imagine it can be as far apart as Cauchy or worse normal/uniform (slash distribution). Moreover, if an experiment is designed and executed well as it should be, there is no reason for extreme deviations and so called breakdown points to occur. We are, therefore, primarily interested in seeking robustness with respect to alternatives (a). We also prefer to use all the observations in a sample rather than censor some of them implicitly or explicitly as in Huber (1981) and Tiku et al. (1986), respectively. It is indeed very desirable to avoid censoring of observations since practitioners of statistics are generally dissuaded from deliberately censoring observations. Should, however, robustness against alternatives of type (b) be sought, censoring a substantial number of observations might be necessary to achieve robustness. The reason is that the extreme order statistics in random samples from such distributions have large or infinite variances and their inclusion inflates the variance of an estimator making it inefficient.

8.2 ROBUST ESTIMATORS OF LOCATION AND SCALE

In the first place, we want to find estimators of a location parameter (mean) μ and a scale parameter (standard deviation) σ which are robust to long-tailed symmetric distributions (kurtosis $\mu_4/\mu_2^2 > 3$). Huber M-estimators are particulary useful here. Alternatively, we recommend taking p small in the family of distributions (2.2.9) and calculating the MMLE $\hat{\mu}$ and $\hat{\sigma}$ from (2.4.8) and using them as robust estimators. If for a sample, the value of C in (2.4.8) is negative (this happens very rarely if $p > 3$), the values of $\hat{\mu}$ and $\hat{\sigma}$ are calculated from the sample with α_i replaced by $\alpha_i^* = 0$ and β_i replaced by $\beta_i^* = 1/\{1 + (1/k)t_{(i)}^2\}$. This ensures that $\hat{\sigma}$ is always real and positive. The values of $t_{(i)}$ $(1 \le i \le n)$ are obtained from (2.4.13). Thus, no tables of the expected values of order statistics are needed to calculate the MMLE $\hat{\mu}$ and $\hat{\sigma}$. We, particularly, recommend taking p = 3.5; see also Lange et al. (1989) and Tiku et al. (2000). The reason is that (2.2.9) with p = 3.5 is the scaled Student t distribution with $\nu = 2p - 1 = 6$ degrees of freedom (its kurtosis is finite but all its even moments of order six and higher do not exist). The distribution is strategically located between the Student t with $\nu = 3$ (its even moments of order greater than two do not exist) and normal (its even moments are all finite).

As plausible alternatives to the assumed distribution $f(p, \sigma)$ in (2.2.9) with p = 3.5 we consider the following, all having finite mean and finite variance, i.e., alternatives of type (a) above.

Misspecification of the distribution:

$$(1)\ p = 2 \qquad (2)\ p = 2.5 \qquad (3)\ p = 5 \qquad (4)\ p = \infty \text{ (normal)}. \tag{8.2.1}$$

Dixon's outlier model:

$(n - k_1)$ observations come from $f(3.5, \sigma)$ and k_1 (we do not know which) come from (5) $f(3.5, 2\sigma)$, (6) $f(3.5, 4\sigma)$, (7) $f(\infty, 4\sigma)$; (8.2.2)

$$k_1 = [0.5 + 0.1n] \text{ (integer value)}.$$

Mixture model:

$$(8)\ 0.90\, f(3.5, \sigma) + 0.10\, f(3.5, 4\sigma). \tag{8.2.3}$$

Contamination model:

$$(9)\ 0.90\, f(3.5, \sigma) + 0.10 \text{ Uniform } (-0.5, 0.5). \tag{8.2.4}$$

The distribution $f(3.5, \sigma)$ will be called population model and the distributions (1) – (9) will be called sample models. Note that $f(\infty, \sigma)$ is the normal distribution $N(\mu, \sigma^2)$.

Without loss of generality we take $\mu = 0$ in which case the mean of all the distributions above is zero. The estimators $\hat{\mu}$ and $\hat{\mu}_w$ are both unbiased for μ. The estimators $\hat{\sigma}$ and $\hat{\sigma}_w$ both estimate $\tau\sigma$ $(\tau > 0)$; τ is, in fact, the square root of the ratio of the variance of the sample model to the variance of the population model. The coefficient τ has no role to play in the computation of the estimators. Its values are given in Table 8A.1 only for determination of the bias and the MSE (mean square error) of $\hat{\sigma}$ and $\hat{\sigma}_w$. Since the assumed population model is $f(3.5, \sigma)$, the coefficients α_i and β_i (or α_i^* and β_i^*) are computed from (2.3.14) or (2.4.12) with p = 3.5 and used for all the sample models (1) – (9) above. It may be noted that the estimators $(\hat{\mu}, \hat{\sigma})$, $(\hat{\mu}_w, \hat{\sigma}_w)$ etc., are matching estimators, i.e., for large n

$$V\{\sqrt{m}\hat{\mu}/E(\hat{\sigma})\} \cong 1 \text{ (for the assumed distribution), and}$$

$$V\{\sqrt{n}\hat{\mu}_w/E(\hat{\sigma}_w)\} \cong 1.$$

This gives the asymptotic distribution of $\sqrt{m}\hat{\mu}/\hat{\sigma}$ and $\sqrt{n}\ \hat{\mu}_w/\hat{\sigma}_w$ as normal with variance 1; σ^2/m is the asymptotic variance of $\hat{\mu}$.

The simulated variances of $\hat{\mu}$ and $\hat{\mu}_w$, and the simulated means and MSE of $\hat{\sigma}$ and $\hat{\sigma}_w$, are given in Appendix 8A (Table 8A.1). The simulated means of $\hat{\mu}$ and $\hat{\mu}_w$ are not reproduced since they are equal to zero (almost) as expected. It may be noted that $\hat{\mu}$ is uncorrelated with $\hat{\sigma}$ and, similary, $\hat{\mu}_w$ is uncorrelated with $\hat{\sigma}_w$. This follows from symmetry. The values for the M-estimators BS82 and H22 are not reproduced since they turned out to be essentially the same as w24; see also Dunnett (1982) and Tiku et al. (1986). The simulated values given in Table 8A.1 are based on [100,000/n] (integer value) Monte Carlo runs and their standard errors are well within ± 0.005. Türker (2002) has similar results.

It may be noted that for p = 3.5 in (2.2.9), the minimum variance bounds for estimating μ and σ calculated from (2.3.19) and (2.4.9) are

$$MVB(\mu)=0.857\sigma^2/n \quad \text{and} \quad MVB(\sigma) = 1.5\sigma^2/2n. \tag{8.2.5}$$

For the population model f(3.5, σ), it is clear from Table 8A.1 that the MMLE $\hat{\mu}$ and $\hat{\sigma}$ are very highly efficient. In fact, they are the MVB estimators for large n. They are also unique and explicit and are easy to compute. Maximum likelihood estimators have enormous computational difficulties and are, therefore, elusive (Chapter 2); see also Barnett (1966a) and Vaughan (1992 a; 2002, p. 230) who discuss this issue in detail.

It can be seen from Table 8A.1 that $\hat{\mu}$ is on the whole as efficient as $\hat{\mu}_w$. However, there is a problem with $\hat{\sigma}_w$: it can have substantial downward bias and has on the whole larger MSE than $\hat{\sigma}$. We conclude, therefore, that the MMLE $\hat{\mu}$ and $\hat{\sigma}$ are jointly more efficient than the M-estimators $\hat{\mu}_w$ and $\hat{\sigma}_w$ for the very wide range of long-tailed symmetric distributions represented by the models in Table 8A.1. It may also be noted that no observation is censored in calculating the MMLE $\hat{\mu}$ and $\hat{\sigma}$. In calculating the w24 estimators $\hat{\mu}_w$ and $\hat{\sigma}_w$ (and BS82 and H22) an unknown number of smallest and largest observations are censored, and the number is random depending on how many observations satisfy the constraint $|z_i| < \pi$ $(1 \le i \le n)$, $z_i = (y_i - T_0)/(hS_o)$ as in (2.11.10).

The reason for the inherent robustness of the MMLE is that $B/\sqrt{nC} \cong 0$ at any rate for large n. Consequently,

$$\hat{\sigma}^2 \cong C/(n-1); \; C = (2p/k)\sum_{i=1}^{n} \beta_i \,(y_{(i)} - \hat{\mu}^2)$$

and

$$\hat{\mu} = \sum_{i=1}^{n} \beta_i y_{(i)}/m \left(m = \sum_{i=1}^{n} \beta_i \right). \tag{8.2.6}$$

But the coefficients β_i $(1 \le i \le n)$ have umbrella ordering, i.e., they increase until the middle value and then decrease in a symmetric fashion. For p = 3.5 and n = 20, for example, the first ten β_i coefficients are given in (2.10.5); $\beta_{n-i+1} = \beta_i$. Thus, the extreme observations $y_{(i)}$ in $\hat{\mu}$ and extreme deviation squares $(y_{(i)} - \hat{\mu})^2$ in $\hat{\sigma}^2$ automatically receive small weights. This depletes the influence of extreme observations and gives the MMLE the inherent robustness property they have. They are also highly efficient for the assumed distribution; in fact, they are the MVB estimators for large n. In other words, $\hat{\mu}$ and $\hat{\sigma}$ are robust estimators, robust to plausible deviations from the assumed distribution.

Robustness of the MMLE $\hat{\mu}$ and $\hat{\sigma}$ with respect to the symmetric models (1) – (9) above (all having finite mean and finite variance) was achieved by using a strategically chosen value of p in (2.2.9), i.e., p=3.5. Consider now alternatives of type (b) above, e.g., the following distributions (called disaster distributions in Tiku et al., 1986, p. 60);

Outlier model: (10) $(n - k_1)N(0, \sigma^2)$ and $k_1N(0, 9\sigma^2)$,

(11) $(n - k_1)N(0, \sigma^2)$ and $k_1N(0, 100\,\sigma^2)$, $k_1 \geq [0.5+0.2n]$, (8.2.7)

(12) Student t_2, (13) Cauchy, and

(14) normal/uniform. (8.2.8)

It may be noted that the variances of (10) and (11) are very large, the variance of (12) does not exist, and both the mean and variance of (13) and (14) do not exist. The extreme order statistics in random samples from (12) – (14) have infinite variances.

The models (10) – (14) are rare but have been considered in robustness studies (Huber, 1981; Tiku et al., 1986). Such extreme models can easily be detected by plotting techniques (Lawless, 1982, Section 2.4; Cleveland, 1984) or Q – Q plots and/or formal goodness-of-fit tests (Chapter 9) and remedial action taken and the process that produces such extreme deviations corrected. However, the Huber M-estimators are very useful here in estimating the location parameter μ. Computation of the w24 estimators reveals that for models (10) – (14) a large proportion of observations in a sample do not satisfy the condition $|z| < \pi$ in (2.11.10) and are censored implicitly and, similarly, for the BS82 and H22 estimators. For achieving robustness with respect to extreme deviations represented by (10) – (14), we recommend the use of the MMLE $\hat{\mu}_c$ and $\hat{\sigma}_c$ based on the censored sample (Tiku, 1980; Dunnett, 1982)

$$y_{(r+1)} \leq y_{(r+2)} \leq \ldots \leq y_{(n-r)} \tag{8.2.9}$$

with $r = [0.5 + 0.3n]$; see also Tiku et al. (1986) and Vaughan (1992a). Since a normal distribution provides an adequate fit to the middle fourty or so percent order statistics (8.2.9), in conformity with the notion that non-normality essentially comes from the tails (Section 7.4 in Chapter 7), we take $p = \infty$ in (2.2.9). Realize that for $p = \infty$, (2.2.9) reduces to normal $N(\mu, \sigma^2)$. Of course, any other value of p in (2.2.9) may be chosen but that does not necessarily enhance the efficiencies in any substantial way since a majority of observations are censored.

The expressions for the MMLE $\hat{\mu}_c$ and $\hat{\sigma}_c$ are given in (7.3.11), $r_1 = r_2 = r$. They are easy to compute and $\hat{\sigma}_c$ is always real and positive. Given in Table 8.1 are the simulated variances of $\hat{\mu}_c$ and $\hat{\mu}_w$ reproduced from Tiku et al. (1986, p. 60), both being unbiased for μ. Also given are the simulated means of $\hat{\sigma}_c$ and $\hat{\sigma}_w$, both presumed to estimate σ. The values for the M-estimators BS82 and H22 are not reproduced since they are essentially the same as w24. We take $\sigma = 1$ without loss of generality.

It can be seen that the MMLE $\hat{\mu}_c$, like $\hat{\mu}_w$, has good robustness properties with respect to models (10) – (14). The MMLE $\hat{\sigma}_c$ is on the whole closer to σ than $\hat{\sigma}_w$ is. Incidentally, $\hat{\mu}_c$ is considerably more efficient than the sample median for all the distributions in Table 8A.1 and Table 8.1. We are, however, primarily interested in more realistic models such as (1) – (9) above. Here, the MMLE $\hat{\mu}$ and $\hat{\sigma}$ based on complete samples are very useful as robust estimators.

Table 8.1: Means and variances of robust estimators; $n = 20$, $k_1 = 4$.

	Estimator	(10)	(11)	(12)	(13)	(14)
Variance	$\hat{\mu}_w$	0.098	0.105	0.106	0.185	0.362
	$\hat{\mu}_c$	0.092	0.113	0.095	0.167	0.329
Mean	$\hat{\sigma}_w$	1.33	1.39	1.37	1.80	2.53
	$\hat{\sigma}_c$	1.15	1.26	1.15	1.41	2.09

It may be noted that the simulated means and variances of the robust estimators above are good measures of their bias and efficiency since their distributions are normal or near-normal.

8.3 COMPARISON FOR SKEW DISTRIBUTIONS

Although the M-estimators are not particulary suited to skew distributions, but let us venture to evaluate their performance and that of the MMLE used in Section 8.2 (for symmetrical distributions) for two representative skew distributions from the Generalized Logistic family GL(b, σ), b = 0.5 and b = 4. The skewness and kurtosis of these two distributions are

	b = 0.5	b = 4
Skewness	– 0.855	0.868
Kurtosis	5.400	4.758

The estimators $\hat{\mu}$ and $\hat{\mu}_w$ both are estimating $\mu^* = \mu + \{\psi(b) - \psi(1)\}\sigma$, and $\hat{\sigma}$ and $\hat{\sigma}_w$, both are estimating $\tau\sigma$, $\tau = \sqrt{\{\psi'(b) + \psi'(1)\}}$. The values of $\psi(b)$ and $\psi'(b)$ are given in Appendix 2D (Chapter 2). Given in Table 8.2 are the means and the MSE (mean square errors) of the estimators. Also given are the values of μ^* and τ; μ and σ are taken to be equal to 0 and 1, respectively, without loss of generality. Note that μ^* and τ have absolutely no role to play in the computation of the estimators.

It can be seen that the M-estimators $\hat{\mu}_w$ and $\hat{\sigma}_w$ are trailing behind: they have generally larger bias and larger MSE than the MMLE $\hat{\mu}$ and $\hat{\sigma}$. See also Table 2.4 (Chapter 2) which gives the means and MSE of the MML and the bias-corrected w24 estimators. The results for other skew distributions, e.g. the Weibull family, were found to be similar. The conclusion is that the MMLE perform better than the M-estimators for skew distributions. We have also shown in Chapter 7 that for symmetric short-tailed distributions, the M-estimators are less efficient than $\bar{y}$ and s and much less efficient than the MMLE (based on samples censored in the middle). See also Section 8.6 for additional results on the MMLE.

Table 8.2: Means and MSE of the MML and w24 estimators for skew distributions.

	Mean				MSE			
n	$\hat{\mu}$	$\hat{\mu}_w$	$\hat{\sigma}$	$\hat{\sigma}_w$	$\hat{\mu}$	$\hat{\mu}_w$	$\hat{\sigma}$	$\hat{\sigma}_w$
				b = 0.5				
				μ^* = – 1.386, τ = 2.565				
10	– 1.310	– 1.243	2.719	2.161	0.604	0.647	0.716	0.669
20	– 1.256	– 1.220	2.662	2.268	0.307	0.334	0.321	0.328
40	– 1.234	– 1.224	2.612	2.310	0.166	0.178	0.155	0.192
60	– 1.225	– 1.223	2.594	2.324	0.122	0.127	0.099	0.141
				b = 4				
				μ^* = 1.833, τ = 1.513				
10	1.781	1.746	1.493	1.207	0.188	0.205	0.198	0.171
20	1.757	1.746	1.459	1.262	0.095	0.102	0.090	0.081
40	1.746	1.749	1.436	1.284	0.050	0.052	0.040	0.042
60	1.739	1.750	1.433	1.294	0.040	0.040	0.028	0.030

8.4 COMPARISON WITH TUKEY AND TIKU ESTIMATORS

For models such as the ones considered in Section 8.2, the Tukey estimators $\hat{\mu}_{Trim}$ and $\hat{\sigma}_{Trim}$ (given in 1.2.5 and 2E.1) and the Tiku estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ (given in 2E.2 – 2E.3) are used as robust estimators. Both are based on censored samples (8.2.9) with r = [0.5+0.1n]. To have an idea about their efficiencies as compared to the MMLE $\hat{\mu}$ and $\hat{\sigma}$ based on complete samples (Section 8.2), we give in Table 8.3 the simulated variances of $\hat{\mu}_c$ and $\hat{\mu}_{Trim}$ (both being unbiased for μ) and the simulated means and MSE of $\hat{\sigma}_c$ and $\hat{\sigma}_{Trim}$. It suffices to reproduce values only for four models as in Table 8.3 since the values for other models are similar. It can be seen that, like the M-estimator $\hat{\sigma}_w$, $\hat{\sigma}_c$ and particularly $\hat{\sigma}_{Trim}$ can have substantial downward bias. Moreover, they are not on the whole as efficient as the MMLE $\hat{\mu}$ and $\hat{\sigma}$ (based on complete samples). It is interesting to note that ($\hat{\mu}_c$, $\hat{\sigma}_c$) are essentially as efficient as ($\hat{\mu}_w$, $\hat{\sigma}_w$); see also Tiku et al. (1986, Chapter 2).

The Tukey estimators are on the whole a little less efficient (jointly) than the Tiku estimators; see also Tiku (1980) and Dunnett (1982).

It should be stated that the Tiku estimators have a rigorous foundation and are defined for a censored sample

$$y_{(r_1+1)} \le y_{(r_1+2)} \le \ldots \le y_{(y_{n-r_2})} \tag{8.4.1}$$

with $r_1 \ge 0$ and $r_2 \ge 0$, r_1 and r_2 not necessarily equal. If the underlying distribution is normal, they are fully efficient for large n; for small n, they are at least as efficient as the BLUE (Chapter 7). These results, of course, extend to other distributions (Chapter 7). See also Bhattacharyya (1985).

8.5 HYPOTHESIS TESTING

In practice it is advantageous, since the underlying distribution is not known exactly, to use a hypothesis testing procedure which has both the criterion robustness as well as the efficiency robustness (Preface). Consider in the first place long-tailed symmetric (LTS) distributions of the type $(1/\sigma)f((y-\mu)/\sigma)$. To test H_0: $\mu=0$, we have the following robust statistics.

The statistic based on the MMLE is

$$T = \sqrt{M}\,(\hat{\mu}/\hat{\sigma}) \tag{8.5.1}$$

Table 8.3: Means, variances and mean square errors of Tukey and Tiku estimators.

	Var		Mean		MSE	
n	$\hat{\mu}_c$	$\hat{\mu}_T$	$\hat{\sigma}_c$	$\hat{\sigma}_T$	$\hat{\sigma}_c$	$\hat{\sigma}_T$
			True model: $\tau = 1$			
10	0.092	0.090	0.874	0.812	0.095	0.103
20	0.046	0.044	0.884	0.818	0.051	0.066
40	0.022	0.021	0.894	0.825	0.029	0.047
60	0.015	0.014	0.902	0.831	0.022	0.038
			Model (1): $\tau = 1$			
10	0.064	0.059	0.708	0.658	0.153	0.176
20	0.032	0.029	0.712	0.659	0.116	0.144

40	0.016	0.014	0.715	0.660	0.097	0.129
60	0.011	0.009	0.715	0.660	0.092	0.125
			Model (3): $\tau = 1$			
10	0.098	0.097	0.912	0.847	0.086	0.091
20	0.048	0.047	0.925	0.856	0.043	0.053
40	0.024	0.024	0.933	0.861	0.022	0.034
60	0.016	0.016	0.934	0.862	0.016	0.029
			Model (5): $\tau = 1.140$			
10	0.108	0.103	0.940	0.873	0.138	0.156
20	0.053	0.050	0.965	0.893	0.075	0.099
40	0.026	0.024	0.963	0.889	0.053	0.082
60	0.018	0.017	0.965	0.890	0.045	0.074

where $M = (2p/k) \sum_{i=1}^{n} \beta_i$; β_i are replaced by β_1^* $(1 \le i \le n)$ whenever the latter are used in the computation of $\hat{\mu}$ and $\hat{\sigma}$ as explained in Section 8.2. In the robustness framework using $\sqrt{M}$ in (8.5.1), in place of $\sqrt{\{np(p-1/2)/(p-3/2)(p+1)\}}$ in (2.10.3), gives type I errors generally a little closer to a pre-assigned level.

The statistic based on the wave estimators w24 is

$$t_w = \sqrt{n}\,(\hat{\mu}_w/\hat{\sigma}_w). \tag{8.5.2}$$

The statistic based on the BS82 and H22 estimators are exactly similar to t_w. For example, the statistic based on the BS82 estimators is

$$t_B = \sqrt{n}\,(\hat{\mu}_B/\hat{\sigma}_B). \tag{8.5.3}$$

Large values of the statistics lead to the rejection of H_0 in favour of H_1: $\mu > 0$. The null distribution of T and t_w (and t_B) are referred to the Student t with n – 1 degrees of freedom, the asymptotic distributions being normal N(0, 1).

Given in Table 8.4 are the simulated values of the probabilities (type I errors)

$$\text{Prob}\{T \ge t_\alpha(\nu) \mid H_0\} \quad \text{and} \quad \text{Prob}\{t_w \ge t_\alpha(\nu) \mid H_0\} \tag{8.5.4}$$

for eleven models considered in Table 8A.1; $t_\alpha(\nu)$ is the $100(1-\alpha)$ percent point of the Student t distribution with $\nu = n - 1$ degrees of freedom. The Student t provides accurate values except for the null distribution of t_w when n is small (≤ 20). Here, the percentage points of t_w need correcting since its type I error is a little too high; see also Tiku et al. (1986, p.119).

When the percentage points of t_w for $n \le 20$ are corrected, its power is not on the whole higher than that of T in any substantial way. Consider, for example, the LTS distribution f(p, σ) with p = 3.5. For n = 10 and 20, the 95 % points of t_w are 2.10 and 1.816, respectively. Given in Table 8.5 are the simulated values of the power for some of the models considered in Table 8A.1. It suffices to reproduce values only for four models. The values for other models are similar. For n > 20, the differences in the values are of the same order.

It can be seen that the T test is essentially as powerful as the t_w test. The T test also extends to short-tailed symmetric and skew distributions as follows. See also Akkaya and Tiku (2003).

8.6 ROBUSTNESS FOR STS DISTRIBUTIONS

Consider the family of STS (short-tailed symmetric) distributions

$$f(y) \propto \frac{1}{\sqrt{2\pi}\sigma}\left\{1+\frac{\lambda}{2r}\left(\frac{y-\mu}{\sigma}\right)^2\right\}^r \exp\left\{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right\}, \quad -\infty < y < \infty; \tag{8.6.1}$$

$\lambda = r/(r-d)$, $d < r$. The MMLE of μ and σ, obtained exactly along the same lines as in Section 2.4 (Chapter 2), are

$$\hat{\mu} = \sum_{i=1}^{n} \beta_i y_{(i)}/m \left(m = \sum_{i=1}^{n} \beta_i \right) \quad \text{and}$$

$$\hat{\sigma} = \left(-\lambda B + \sqrt{(\lambda B)^2 + 4nC}\right)/2\sqrt{n(n-1)}; \tag{8.6.2}$$

Table 8.4: Values of type I errors of the T and t_w tests, presumed value is 0.050.

n	T	t_w	T	t_w	T	t_w
	True model		Model (1)		Model (2)	
10	0.044	0.064	0.042	0.067	0.044	0.065
20	0.046	0.053	0.043	0.055	0.051	0.064
40	0.045	0.047	0.044	0.050	0.039	0.047
60	0.052	0.053	0.057	0.056	0.050	0.056
	Model (4)		Model (5)		Model (6)	
10	0.053	0.067	0.045	0.065	0.036	0.066
20	0.055	0.059	0.043	0.060	0.034	0.054
40	0.054	0.051	0.048	0.059	0.030	0.047
60	0.056	0.052	0.048	0.050	0.050	0.050
	Model (7)		Model (8)		Model (9)	
10	0.039	0.066	0.038	0.065	0.046	0.066
20	0.034	0.052	0.033	0.053	0.046	0.058
40	0.036	0.052	0.031	0.047	0.038	0.047
60	0.041	0.056	0.048	0.045	0.042	0.047

$$B = \sum_{i=1}^{n} \alpha_i y_{(i)} \quad \text{and} \quad C = \sum_{i=1}^{n} \beta_i (y_{(i)} - \hat{\mu})^2 = \sum_{i=1}^{n} \beta_i y_{(i)}^2 - m\hat{\mu}^2. \tag{8.6.3}$$

For $\lambda \le 1$, the coefficients α_i and β_i (given in 3.6.14) are used. For $\lambda > 1$, α_i and β_i are replaced by α_i^* and β_i^* (given in 3.7.5), respectively.

For large n, $\hat{\mu}$ is the MVB estimator of μ with variance $V(\hat{\mu}) \cong \sigma^2/m$. This follows from the asymptotic equivalence of likelihood and modified likelihood equations and the representation

$$\frac{\partial \ln L}{\partial \mu} \cong \frac{\partial \ln L^*}{\partial \mu} = \frac{m}{\sigma^2}(\hat{\mu} - \mu). \tag{8.6.4}$$

For achieving robustness with respect to STS distributions and to inliers, we choose a strategically located member of the family (8.6.1); $r = 4$ and $d = 0$ ($\lambda = 1$) is a good choice since the kurtosis of the distribution is 2.370 and is a value between 1.80 (kurtosis of a uniform distribution) and 3 (kurtosis of a normal distribution).

Table 8.5: Values of the power of T and t_w tests, $\sigma = 1$.

	$\mu = 0.0$	0.4	0.8	1.2	$\mu = 0.0$	0. 2	0.4	0.6	0.8
		n = 10					n = 20		
				True model					
T	0.045	0.34	0.78	0.97	0.045	0.24	0.57	0.86	0.98
t_w	0.047	0.32	0.76	0.96	0.050	0.24	0.56	0.85	0.97
				Model (1)					
T	0.040	0.44	0.86	0.97	0.040	0.28	0.69	0.92	0.98
t_w	0.046	0.45	0.88	0.99	0.047	0.32	0.74	0.95	0.99
				Model (3)					
T	0.049	0.34	0.77	0.97	0.048	0.23	0.57	0.84	0.98
t_w	0.045	0.30	0.73	0.95	0.044	0.23	0.54	0.83	0.96
				Model (5)					
T	0.042	0.36	0.81	0.97	0.046	0.23	0.60	0.88	0.98
t_w	0.046	0.34	0.79	0.97	0.049	0.24	0.62	0.88	0.98
				Model (8)					
T	0.036	0.47	0.86	0.96	0.034	0.30	0.71	0.92	0.98
t_w	0.047	0.50	0.92	0.99	0.049	0.37	0.81	0.97	1.00

To compare the MMLE $\hat{\mu}$ and $\hat{\sigma}$ with the sample mean $\bar{y}$ and the sample standard deviation s, we give in Table 8A.2 the simulated variances of $\hat{\mu}$ and $\bar{y}$ for a variety of models (STS distributions); both estimators are unbiased for μ. Also given are the means and variances of $\hat{\sigma}$ and s. The alternatives we consider are the following sample models (1) – (4); the population model being (8.6.1) with r = 4 and d = 0. Realize that $\hat{\sigma}$ and s both estimate $\tau\sigma$; τ is the square root of the ratio of the variance of the sample model to the variance of the population model.

Misspecification of the distribution:

(1) The normal N(0, 1) with kurtosis 3.

(2) Tukey lambda family $z = \{u^1 - (1 - u)^1\}/1$ with l = 0.585, l = 1 and l = 1.45, the kurtosis being 2, 1.80 and 1.75, respectively.

(3) The symmetric short-tailed family with cdf (cumulative distribution function)

$$\begin{aligned} F(z) &= 2^{k-1} z^k, \quad 0 < z < 0.5 \\ &= 1 - 2^{k-1} (1 - z)^k, \quad 0.5 < z < 1 \quad (k > 1). \end{aligned} \tag{8.6.5}$$

We take k = 1.5, 2.0 and 3.0 as in Dudewicz and Meulen (1981, p.969) with kurtosis 2.123, 2.400 and 2.856, respectively.

(4) The symmetric short-tailed family with cdf

$$\begin{aligned} F(z) &= 0.5 - 2^{k-1} (0.5 - z)^k, \quad 0 < z < 0.5 \\ &= 0.5 + 2^{k-1}(z - 0.5)^k, \quad 0.5 < z < 1 \quad (k > 1). \end{aligned} \tag{8.6.6}$$

We take k = 1.5 and 2.0 (Dudewicz and Meulen, 1981) with kurtosis 1.486 and 1.330, respectively.

It can be seen that the MMLE $\hat{\mu}$ and $\hat{\sigma}$ obtained by assuming a strategically located distribution from the STS family of distributions (8.6.1) are remarkably efficient and robust.

This is due to the inverted umbrella ordering of the coefficients β_i(or β_1*): they constitute a sequence of values which decrease until the middle value and then increase in a symmetric fashion. For r = 4 and d = 0 and n = 20, for example, the first ten β_i coefficients are, $\beta_{n-i+1} = \beta_i$:

0.957 0.831 0.707 0.580 0.453 0.329 0.214 0.116 0.043 0.005 (8.6.7)

Thus, the middle observations automatically receive small weights. This depletes their influence and makes $\hat{\mu}$ and $\hat{\sigma}$ robust to STS distributions and data anomalies in the middle of the sample (inliers). Realize that $(\hat{\mu}, \hat{\sigma})$ are overall considerably more efficient than $(\bar{y}, s)$.

The statistic $T = \sqrt{m}\,(\hat{\mu}/\hat{\sigma})$ provides a test of H_0: $\mu = 0$ which has criterion as well as efficiency robustness with respect to short-tailed distributions and inliers in a sample; see Akkaya and Tiku (2003).

Comment: Since the MML estimators use all the observations in a sample and are defined for all the three types of distributions (skew, STS and LTS), we particularly focus on them in this book. For comprehensive details about the M-estimators, one may refer to the authoritative books by Huber (1981), Hoaglin et al. (1983), Rey (1983), and Rousseuw and Leroy (1987). For details about the MMLE $\hat{\mu}_c$ and $\hat{\sigma}_c$ based on censored normal samples, one may refer to Tiku et al. (1986), Schneider (1986) and Cohen (1991). Some of the key results for censored normal and non-normal samples are given in Chapter 7.

Remark: Robustness with respect to skew distributions can similary be achieved by choosing a strategically located distribution and calculating the MMLE $\hat{\mu}$ and $\hat{\sigma}$ and the Student type statistics based on them, along the same lines as in the next section.

8.7 ROBUSTNESS OF REGRESSION ESTIMATORS

Consider the linear model (3.2.1) and suppose that the errors e_i $(1 \le i \le n)$ are iid and have the Generalized Logistic distribution $(e = y - \theta_0 - \theta_1 x)$

$$GL(b, \sigma) : (b/\sigma)\exp(-e/\sigma)/\{1 + \exp(-e/\sigma)\}^{b+1}, \quad -\infty < e < \infty; \tag{8.7.1}$$

θ_0 and σ in the model are essentially location and scale parameters and their efficiency and robustness properties are discussed in Chapter 2 and Sections 8.2 – 8.6. See also Islam et al. (2001). We will, therefore, focus on the MML and the LS estimators of θ_1. The latter is used extensively in statistical applications in numerous areas.

Modified likelihood: The MMLE of θ_1 is (Chapter 2)

$$\hat{\theta}_1 = K - D\hat{\sigma}; \quad \hat{\sigma} = \{-B + \sqrt{B^2 + 4nC}\}/2\sqrt{n(n-1)}, \tag{8.7.2}$$

$$K = \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})y_{[i]} \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2,$$

$$D = \sum_{i=1}^{n} \Delta_i (x_{[i]} - \bar{x}_{[.]})y_{[i]} \Big/ \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2,$$

$$B = (b+1)\sum_{i=1}^{n} \Delta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\} \quad \text{and}$$

$$C = (b+1)\sum_{i=1}^{n} \beta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\}^2 \tag{8.7.3}$$

$$= (b+1)\left\{\sum_{i=1}^{n} \beta_i \{y_{[i]} - \bar{y}_{[.]}\}^2 - KQ\right\}, \quad Q = \sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})\, y_{[i]},$$

where the coefficients α_i and β_i are given in (2.5.5); $\Delta_i = \alpha_i - (b+1)^{-1}$. Since $\beta_i > 0$ $(1 \le i \le n)$, the MMLE $\hat{\sigma}$ is always real and positive. The MMLE $\hat{\theta}_1$ is unbiased for large n. Its computation is explained in Chapter 3.

Least squares: The LSE of θ_1 is as usual

$$\tilde{\theta}_1 = \sum_{i=1}^{n} (x_i - \bar{x})y_i / \sum_{i=1}^{n} (x_i - \bar{x})^2$$

with variance

$$V(\tilde{\theta}_1) = \{\psi'(b) + \psi'(1)\}\,\sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{8.7.4}$$

Realize that $\tilde{\theta}_1$ is unbiased for θ_1.

First, consider the situation when the value of the shape parameter b in (8.7.1) is known exactly. Islam et al. (2001) generate the design points x_i from Uniform (0, 1) and give the simulated variances of the MMLE $\hat{\theta}_1$ and the relative efficiency

$$E = 100\{V(\hat{\theta}_1)/V(\tilde{\theta}_1)\} \tag{8.7.5}$$

Table 8.6: Variance of the MMLE $\hat{\theta}_1$ and the relative efficiency of the LSE $\tilde{\theta}_1$.

	b = 0.5		b = 2.0		b = 4.0		b = 8.0	
n	$nV(\hat{\theta}_1)$	E	$nV(\hat{\theta}_1)$	E	$nV(\hat{\theta}_1)$	E	$nV(\hat{\theta}_1)$	E
10	53.86	85	20.48	93	16.27	86	13.55	78
20	56.80	82	21.56	89	16.60	81	13.82	74
30	57.09	79	22.16	91	16.16	77	14.31	74
50	63.67	79	25.30	89	19.26	79	15.07	72
100	57.65	75	24.85	88	20.30	79	14.86	70

of the LSE $\tilde{\theta}_1$. We reproduce their values in Table 8.6. The means are not given since the bias in $\hat{\theta}_1$ is negligibly small even for small n. It can be seen that the MMLE $\hat{\theta}_1$ is considerably more efficient than the LSE $\tilde{\theta}_1$. In fact, the relative efficiency of $\tilde{\theta}_1$ decreases as n increases and stabilizes at values considerably less than 100 percent. The relative efficiency of $\tilde{\theta}_1$ is essentially the same for other designs (Islam et al., 2001).

Robustness: Consider the situation when in (8.7.1), b = 0.5 (population model); the skewness and kurtosis of the distribution are $\mu_3/\mu_2^{3/2} = -0.855$ and $\mu_4/\mu_2^{\,2} = 5.400$, respectively. As plausible alternatives, we consider the following sample models representing a wide range of skew distributions:

In GL(b, σ): (1) b = 0.2,

(2) $(n - k_1)$ GL (0.5, σ) and k_1GL(0.5, 4σ), $k_1 = [0.5 + 0.1n]$ (outlier model), (8.7.6)

(3) 0.90GL(0.5, σ) + 0.10GL(0.5, 4σ) (mixture model),

(4) 0.90GL(0.5, σ) + 0.10 Uniform (– 05, 0.5) (contamination model).

The simulated means and variances are given in Table 8.7. Also given are the values of the relative efficiency of the LSE $\tilde{\theta}_1$. It can be seen that the bias in both is negligible (except for the outlier model for which the LSE has enormous bias) but the MMLE is efficient and robust which is due to half-umbrella ordering of the β_i $(1 \le i \le n)$ coefficients: they decrease in the direction of the long tail. Thus, the effect of the long tail is automatically depleted.

Table 8.7: Relative efficiency of the LSE; true model is GL(0.5, σ).

	n = 10				n = 50			
	Mean		Var	RE	Mean		Var	RE
	$\tilde{\theta}_1$	$\hat{\theta}_1$	$\hat{\theta}_1$	$\tilde{\theta}_1$	$\tilde{\theta}_1$	$\hat{\theta}_1$	$\hat{\theta}_1$	$\tilde{\theta}_1$
True model	1.01	0.97	0.897	85	1.00	1.00	0.168	75
Model (1)	1.00	0.98	0.194	66	1.00	1.00	0.032	52
(2)*	1.24	1.04	0.553	72	1.33	1.02	0.098	41
(3)	1.00	1.00	0.709	84	0.99	1.00	0.083	80
(4)	0.99	0.97	1.058	84	1.00	1.00	0.238	78

* The LSE has enormous bias.

Hypothesis testing: To test the null hypothesis H_0: θ_1=0, we define the statistic

$$T = \sqrt{\left\{(b+1)\sum_{i=1}^{n} \beta_i\, (x_{[i]} - \bar{x}_{[.]})^2\right\}}\quad (\hat{\theta}_1/\hat{\sigma}). \tag{8.7.7}$$

Large values of T lead to the rejection of H_0 in favour of H_0: $\theta_1 > 0$. In view of the fact that ln $L^*/\partial\theta_1$ assumes a form exactly similar to (3.5.7), the asymptotic null distribution of T is normal N(0, 1). For small n (≤ 30), the null distribution of T is referred to the Student t with ν = n – 1 degrees of freedom. The Student t and normal N(0, 1) distributions give remarkably accurate values for the probability (Islam et al., 2001)

$$P\{T \ge t_\alpha(\nu) \mid H_0\} \tag{8.7.8}$$

for all b ≥ 0.5. For example, we have the following simulated values:

Type I error of the T test, presumed value is 0.050.

	b						
n	0.2	0.5	1.0	2.0	4.0	6.0	8.0
10	0.062	0.051	0.052	0.051	0.049	0.055	0.051
20	0.066	0.050	0.051	0.046	0.046	0.046	0.045
50	0.060	0.058	0.050	0.051	0.049	0.045	0.052

For b = 0.2, GL(b, σ) has enormous amount of skewness and that explains why the type I error of the T test is considerably larger than the presumed value. For large n (≥ 100), however, the type I error is close to 0.050 even for b = 0.2.

The statistic based on the LSE is

$$G = \sqrt{\textstyle\sum_{i=1}^{n}(x_i - \bar{x})^2}\ (\tilde{\theta}_1/s_e); \tag{8.7.9}$$

$$s_e^{\,2} = \textstyle\sum_{i=1}^{n}\{y_i - \bar{y} - \tilde{\theta}_1\,(x_i - \bar{x})\}^2/(n-2).$$

The asymptotic null distribution of G is normal N(0, 1). For small n (≤ 30), the null distribution of G is referred to the Student t with ν = n – 1 degrees of freedom. Gayen (1949) and Tiku (1971a) among others study the distribution of G under non-normality and show that the distribution is very sensitive to population skewness (Chapter 1).

The asymptotic power functions of the T and G tests are exactly similar to those in (3.13.3) with noncentrality parameters $\lambda_1^{\,2}$ and $\lambda_0^{\,2}$ respectively. Here,

$$\lambda_1^2 = \frac{b}{(b+2)}\left(\frac{\theta_1}{\sigma}\right)^2 \quad \text{and} \quad \lambda_0^2 = \frac{1}{\{\psi'(b)+\psi'(1)\}}\left(\frac{\theta_1}{\sigma}\right)^2.$$

Since b/(b+2) is greater than $\{\psi'(b) + \psi'(1)\}^{-1}$, the T tests is more powerful than the G test. For b = 1, for example, the values are 0.333 and 0.304; for b = 0.5, the values are 0.200 and 0.152, respectively. In fact, the T test is more powerful than the G test for all n (Chapter 3). See also Islam and Tiku (2004).

To illustrate the robustness properties, assume that the true distribution is (8.7.1) with b = 0.5. Consider the plausible alternatives (1) – (4) as in (8.7.6). Given in Table 8.8 are the simulated values of the power

$$P(T \geq t_\alpha(\nu) \mid H_1\} \quad \text{and} \quad P\{G \geq t_\alpha(\nu) \mid H_1\} \tag{8.7.10}$$

of the two tests for models (1), (2) and (4). The values for model (3) are similar to those for (2). The random numbers generated were standardized to have variance 1.

Table 8.8: Values of the power, true model is GL(0.5, σ); n = 30.

		θ_1 = 0.0	0.4	0.8	1.2	1.6	2.0	2.4
True model	MML	0.047	0.17	0.42	0.70	0.89	0.97	0.99
	LS	0.047	0.16	0.38	0.62	0.83	0.93	0.98
Model (1)	MML	0.042	0.45	0.93	1.00	1.00	1.00	1.00
	LS	0.053	0.35	0.78	0.96	0.99	1.00	1.00
Model (2)	MML	0.042	0.21	0.56	0.85	0.96	0.99	1.00
	LS	0.056	0.19	0.42	0.64	0.80	0.89	0.99
Model (4)	MML	0.051	0.16	0.34	0.59	0.79	0.92	0.97
	LS	0.047	0.14	0.30	0.51	0.71	0.86	0.94

The superiority of the T test over the G test is apparent. Islam et al. (2001) report similar results for the family of Weibull distributions.

Symmetric family: Consider now the family of LTS distributions (3.9.1). The MMLE of θ_0, θ_1 and σ are given in (3.9.5) – (3.9.7). If for a sample, C in (3.9.7) assumes a negative value, the MMLE are calculated by replacing α_i and β_i by 0 and β_i*, respectively, as explained earlier.

First, consider the situation when the shape parameter p in (3.9.1) is known exactly. We give in Table 8.9 the simulated variances of the MMLE $\hat{\theta}_1$ and the relative efficiency of the LSE $\tilde{\theta}_1$, the design points x_i $(1 \leq i \leq n)$ generated from the Uniform (0,1). The bias in both $\hat{\theta}_1$ and $\tilde{\theta}_1$ is negligible and their means are not, therefore, reported. It can be seen that $\hat{\theta}_1$ is enormously more efficient than $\tilde{\theta}_1$. The relative efficiency of $\tilde{\theta}_1$ decreases as n increases.

Table 8.9: Variances of the MMLE $\hat{\theta}_1$ and the relative efficiency of the LSE $\tilde{\theta}_1$; θ_1=1 and σ = 1.

	p = 2		p = 2.5		p = 3.5		p = 5	
n	Var	RE	Var	RE	Var	RE	Var	RE
10	0.802	80	1.164	90	1.224	93	1.376	97
20	0.368	68	0.657	83	0.533	90	0.661	96
30	0.237	65	0.304	79	0.301	88	0.442	95
50	0.126	61	0.176	77	0.208	85	0.229	95
100	0.076	54	0.109	73	0.112	84	0.118	94

Robustness: Consider the situation when the population model is (3.9.1) with p = 3.5. To study the robustness of $\hat{\theta}_1$ and $\tilde{\theta}_1$, we simulated the means and variances of $\hat{\theta}_1$ and $\tilde{\theta}_1$, and the relative efficiency of $\tilde{\theta}_1$, for the nine alternatives (1) – (9) given in (8.2.1) – (8.2.4), and the outlier and mixture models

(10) $(n - k_1)$ observations come from normal $N(\mu, \sigma^2)$ and k_1 from $N(\mu, 16\sigma^2)$, and

(11) $0.90N(\mu, \sigma^2) + 0.10N(\mu, 16\sigma^2)$: (8.7.11)

$\mu = 0$ and $\sigma = 1$ without any loss of generality.

The simulated values are given in Table 8A.3. It can be seen that $\hat{\theta}_1$ is enormously more efficient than the LSE $\tilde{\theta}_1$, except for model (4) in which case it is a little less efficient as expected. Although we generated the design points x_i $(1 \le i \le n)$ from the Uniform(0, 1) but the relative efficiencies are essentially the same for other designs, e.g., x_i $(1 \le i \le n)$ generated from normal N(0, 1). See also Tiku et al. (2001). Note again that the performance of the LSE for the outlier model is disappointing indeed.

Huber method of M-estimation has been extended to linear regression models; see, for example, Andrews (1974), and Rousseuw and Leroy (1987). There seems to be no overwhelming advantage in using this computer intensive method in view of the comparisons given in Tables 8A.1 and 8.2 and in Table 7.3 (Chapter 7). See also Lange et al. (1989) who show that the MLE based on the Student t distributions with small degrees of freedom are for symmetric long-tailed distributions as efficient and robust as the M-estimators. See also Islam and Tiku (2004).

Hypothesis testing: To test H_0: $\theta_1 = 0$ we note the following representation of $\partial \ln L^*/\partial\theta_1$:

$$\frac{\partial \ln L}{\partial \theta_1} \cong \frac{\partial \ln L^*}{\partial \theta_1} = \frac{1}{\sigma^2}\frac{2p}{k}\{\textstyle\sum_{i=1}^{n} \beta_i\,(x_{[i]} - \bar{x}_{[.]})^2\}(\hat{\theta}_1 - \theta_1), \tag{8.7.12}$$

$k = 2p - 3$, $p \ge 2$. For large n, therefore, the MMLE $\hat{\theta}_1$ is the MVB estimator of θ_1 with variance

$$V(\hat{\theta}_1) \cong \sigma^2/(2p/k)\textstyle\sum_{i=1}^{n} \beta_i\,(x_{[i]} - \bar{x}_{[.]})^2. \tag{8.7.13}$$

To test H_0, we define the statistic

$$T = \sqrt{(2p/k)\textstyle\sum_{i=1}^{n} \beta_i (x_{[i]} - \bar{x}_{[.]})^2}\;(\hat{\theta}_1/\hat{\sigma}). \tag{8.7.14}$$

The estimator $\hat{\sigma}$ is given in (3.9.6); β_i is replaced by β_1^* $(1 \le i \le n)$ if need be. Large values of T lead to the rejection of H_0 in favour of H_1: $\theta_1 > 0$.

The asymptotic null distribution of T is normal N(0, 1). For small n, the null distribution of T is referred to the Student t with $\nu = n - 1$ degrees of freedom.

The statistic based on the LSE $\tilde{\theta}_1$ is

$$G = \sqrt{\textstyle\sum_{i=1}^{n} (x_i - \bar{x})^2}\;(\tilde{\theta}_1/s_e). \tag{8.7.15}$$

The null distribution of G is exactly the same as that of T.

To evaluate the robustness properties of the T and G tests for LTS distributions, we choose p = 3.5 (population model) in the family (3.9.1). As plausible alternatives (sample models) we choose for illustration the models (1), (3), (5) and (8) for e_i $(1 \le i \le n)$, as in (8.2.1) – (8.2.4). Given in Table 8.10 are the simulated values of the probabilities (power)

$$P\{T \ge t_{0.05}(\nu) \mid H_1\} \quad \text{and} \quad P\{G \ge t_{0.05}(\nu) \mid H_1\}; \tag{8.7.16}$$

$t_{0.05}(n)$ is the 95 percent point of Student t distribution with $\nu = n - 1$ degrees of freedom. It can be seen that the T test has the power superiority. See also Tiku et al. (2001), and Islam and Tiku (2004).

In calculating the power, the random numbers generated were standardized to have variance 1. This is important for a power comparison when distributions with different variances are considered.

Table 8.10: Power of the T and G tests for long-tailed symmetric distributions, n = 20.

	θ_1 =	0.0	0.4	0.8	1.2	1.6	2.0	2.4
True model	T	0.049	0.14	0.29	0.50	0.70	0.85	0.94
	G	0.050	0.13	0.28	0.48	0.68	0.83	0.92
Model (1)	T	0.044	0.17	0.41	0.68	0.86	0.94	0.98
	G	0.051	0.17	0.40	0.64	0.80	0.90	0.95
Model (3)	T	0.051	0.12	0.24	0.42	0.60	0.75	0.87
	G	0.050	0.12	0.23	0.39	0.58	0.74	0.86
Model (5)	T	0.048	0.14	0.33	0.56	0.76	0.89	0.96
	G	0.049	0.15	0.32	0.53	0.73	0.85	0.93
Model (8)	T	0.037	0.14	0.36	0.62	0.80	0.91	0.96
	G	0.049	0.15	0.34	0.56	0.73	0.84	0.90

The differences in the values for other values of n are of the same order.

STS distributions: For the family (8.6.1), the MMLE of θ_1 and σ are given in Tiku et al. (2001). To test H_0: $\theta_1 = 0$, the statistic T is exactly similar to that in (8.7.14). The null distribution of T is asymptotically normal N(0, 1) and for $n \le 30$ is well approximated by the Student t with n = n – 1 degrees of freedom. Unlike the G test based on the LSE, the T test is robust to STS distributions and to inliers in a sample. This is due to the inverted-umbrella ordering of the coefficients β_i (or β_i*). Values of the power of the two tests are given in Tiku et al. (2001, Table 4). It is shown that the T test has a double advantage: its type I error is generally smaller but its power is higher than the G test. This is an interesting phenomenon indeed. See also Islam and Tiku (2004).

8.8 ROBUSTNESS OF ESTIMATORS IN BINARY REGRESSION

The binary regression model is (Chapter 4)

$$\pi(x) = E(Y \mid X = x) = \int_{-\infty}^{z} f(u)\, du\,;$$

Y = 0 or 1 and $z = \gamma_0 + \gamma_1 x$ ($\gamma_1 > 0$). Consider the situation when f(u) is one of the distributions in the Generalized Logistic family (4.11.1) denoted here by GL(b, γ_1). The function h(x) in (4.11.1) has in particular the same form as f(u).

The MMLE of γ_0 and γ_1 are given in (4.13.1). Since γ_1 quantifies the effect of the risk factor X on Y, the MMLE $\hat{\gamma}_1$ is of particular interest.

The distribution of $\hat{\gamma}_1$ and its efficiency properties are studied by Oral (2002). Since $\hat{\gamma}_1$ is asymptotically equivalent to the MLE, it is fully efficient (asymptotically) and the distribution of $\hat{\gamma}_1/\gamma_1$ is normal with mean 1 and variance given by the last element in (4.14.1). However, it takes a large sample size (> 100) to attain near-normality. To study the distribution of $\hat{\gamma}_1/\gamma_1$, Oral (2002) proceeds as follows.

Realize that $GL(b, \gamma_1)$ is assumed to be the same as $GL(b, \sigma)$ in (2.5.1) with $\gamma_0 = -\mu/\sigma$ and $\gamma_1 = 1/\sigma$ (> 0). Let x_i $(1 \le i \le n)$ be a random sample of size n from $GL(b, \gamma_1)$ and

$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)} \tag{8.8.1}$$

be the order statistics. The corresponding binary observations $w_i = y_{[i]}$ are obtained by calculating the probabilities

$$\hat{P}_i = \{1 + \exp(-\hat{z}_{(i)})\}^{-b}, \quad \hat{z}_{(i)} = \hat{\gamma}_{00} + \hat{\gamma}_{10} x_{(i)}, \tag{8.8.2}$$

and defining $$w_i = 0 \text{ if } u_i \le \hat{P}_i$$
$$= \text{otherwise}, \tag{8.8.3}$$

u_i $(1 \le i \le n)$ being n independent Uniform (0, 1) variates; $\hat{\gamma}_{00}$ and $\hat{\gamma}_{10}$ are the MMLE calculated from (8.8.1), i.e.,

$$\hat{\gamma}_{00} = (1/m)\sum_{i=1}^{n} \Delta_i - \hat{\gamma}_{10}\bar{x}_{(..)} \quad \text{and} \quad \hat{\gamma}_{10} = \{-B_0 + \sqrt{B_0^{\,2} + 4nC_0}\}/2C_0; \tag{8.8.4}$$

$$m = \sum_{i=1}^{n} \beta_i, \quad \bar{x}_{(..)} = (1/m)\sum_{i=1}^{n} \beta_i x_{(i)}, \quad \Delta_i = \alpha_i - (b+1)^{-1},$$

$$B_0 = (b+1)\sum_{i=1}^{n} \Delta_i (x_{(i)} - \bar{x}_{(.)}) \quad \text{and} \quad C_0 = (b+1)\sum_{i=1}^{n} \beta_i (x_{(i)} - \bar{x}_{(.)})^2.$$

The coefficients α_i and β_i $(1 \le i \le n)$ are obtained from (2.5.5).

Oral (2002) gives the simulated values of the skewness $\beta_1^* = \mu_3^2/\mu_2^3$ and kurtosis $\beta_2^* = \mu_4/\mu_2^2$ of $\hat{\gamma}_1$ when $\gamma_1 = 0.001, 0.01, 0.10$ and 1.00, n = 10, 60, 80 and 100, and b = 0.5, 1, 2 and 4. She shows that β_1^* and β_2^* steadily tend to 0 and 3, respectively, as n increases. Moreover, μ_3 is always positive and β_1^* and β_2^* satisfy the condition (5.15.4) in Chapter 5.

Table 8.11: Skewness and kurtosis of the MMLE $\hat{\gamma}_1$ in binary regression; (a) skewness β_1^*, (b) kurtosis β_2^* and (c) the absolute difference $|\beta_2^* - (3 + 1.5\beta_1^*)|$.

n		b = 0.5	1	2	b = 0.5	1	2
				$\gamma_0 = 0,\ \gamma_1 = 0.001$			
30	(a)	0.399	0.524	0.464	0.523	0.421	0.287
	(b)	3.578	3.909	3.830	3.992	3.687	3.412
	(c)	0.020	0.123	0.134	0.208	0.056	0.018
60	(a)	0.331	0.100	0.103	0.178	0.139	0.239
	(b)	3.850	3.134	3.157	3.513	3.122	3.544
	(c)	0.354	0.016	0.002	0.025	0.086	0.186
100	(a)	0.096	0.209	0.075	0.262	0.139	0.069
	(b)	2.996	3.263	2.932	3.469	3.153	3.408
	(c)	0.148	0.050	0.180	0.076	0.056	0.304

We reproduce her results in Table 8.11 for $\gamma_1 = 0.001$ and n = 30, 60 and 100, for illustration. Thus, the 3-moment chi-square approximation (5.15.5) is applicable. Oral (2002) shows that the approximation provides accurate values for the percentage points of the distribution of $\hat{\gamma}_1/\gamma_1$. In particular, she shows that the mean of $\hat{\gamma}_1/\gamma_1$ is almost 1 (i.e., $\hat{\gamma}_1$ is almost unbiased for γ_1) and its variance is free of γ_0 and γ_1, in agreement with (4.14.1). For b = 0.5, for example, the simulated values of the mean and variance are given in Table 8.12.

Table 8.12: Mean and variance of $\hat{\gamma}_1/\gamma_1$ in binary regression.

		n = 30		n = 60		n = 100	
γ_0	γ_1	Mean	Var	Mean	Var	Mean	Var
0.0	0.001	1.039	0.033	1.019	0.015	1.011	0.009
	0.10	1.035	0.033	1.020	0.016	1.011	0.009
	1.00	1.035	0.033	1.019	0.015	1.011	0.009
2.0	0.001	1.030	0.034	1.021	0.016	1.012	0.008
	0.10	1.037	0.033	1.015	0.015	1.013	0.009
	1.00	1.030	0.033	1.017	0.017	1.014	0.009

For n > 100, of course, the last element in the matrix (4.14.1) gives values close to the simulated variance of $\hat{\gamma}_1/\gamma$. This was to be expected since the MMLE $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are asymptotically fully efficient.

To illustrate the accuracy of the 3-moment chi-square approximation, Oral (2002) gives the following simulated values of the probability $P\{(\hat{\gamma}_1/\gamma_1) \geq \chi_{0.05}{}^2\}$, the presumed value being 0.050; n = 100:

γ_0	$\gamma_1 = 0.001$	0.01	0.1	0.5	1.0
0.0	0.049	0.048	0.052	0.045	0.050
2.0	0.052	0.049	0.050	0.047	0.049

The chi-square approximation is amazingly accurate.

Since X is a genuine risk factor, γ_1 is always positive howsoever small. To test $H_0 = \gamma_1 = \gamma_{10}$ (> 0), she defines the statistic

$$W = \hat{\gamma}_1/\gamma_{10}. \tag{8.8.5}$$

Large values of W lead to the rejection of H_0 in favour of H_0: $\gamma_1 > \gamma_{10}$. The test is asymptotically most powerful since $\hat{\gamma}_1$ is fully efficient (asymptotically). The estimator $\hat{\gamma}_1$ being as efficient as the ML estimator even for small n (Chapter 4), it will be difficult to improve over this test so far its power is concerned.

Robustness: Oral (2002) shows that the W test is remarkably robust to plausible deviations from an assumed density h(x) and to numerous data anomalies (outliers, contaminations and mixtures). Consider, for example, the situation when h(x) is the logistic density, i.e. b = 1 in (4.11.1), and the alternatives are the following.

Misspecification of the density:

In (2.2.9) (y replaced by x), (1) p = 5 and (2) p = 10.

Outlier model: (8.8.6)

(3) $(n - k_1)GL(1, \sigma)$ and $k_1GL(1, 2\sigma)$,

(4) $(n - k_1)GL(1, \sigma)$ and $k_1GL(1, 4\sigma)$; $k_1 = [0.5+0.1n]$.

Mixture model:

(5) $0.90GL(1, \sigma)+0.10GL(1, 2\sigma)$,

(6) $0.90GL(1, \sigma)+0.1GL(1, 4\sigma)$.

Contamination model:

(7) $0.90GL(1, \sigma)+0.10$Uniform $(-0.5, 0.5)$.

Assuming that h(x) is the logistic density $GL(1, \sigma)$, Oral (2002) determines from a 3-moment chi-square approximation the percentage point of the null distribution of the statistic W. Denote this percentage point by χ_α^2. She simulates the values of the probability $P\{W \geq \chi_{0.05}^2 \mid \gamma_1 = \gamma_{10}\}$ for the models (1) – (8) above. We reproduce her values in Table 8.13; $\gamma_0 = 0$, $\gamma_{10} = 0.001$.

Table 8.13: Values of the type I error in binary regression.

Model	n = 50	60	80	100
(1)	0.049	0.051	0.048	0.048
(2)	0.049	0.047	0.048	0.051
(3)	0.051	0.050	0.049	0.051
(4)	0.049	0.049	0.048	0.052
(5)	0.049	0.052	0.048	0.048
(6)	0.051	0.047	0.046	0.050
(7)	0.050	0.053	0.047	0.053

Clearly, the W test has criterion robustness.

Oral (2002) also studies the robustness of the W test when the population model is one of the members of the Generalized Logistic family $GL(b, \gamma_1)$ $(b \neq 1)$ and the Student t family. Her results are similar to those in Table 8.13. She also simulates the values of the power. They are not affected in any substantial way when the underlying distribution constitutes a plausible alternative to the one assumed. Thus, the W test has both criterion as well as efficiency robustness.

It is also of interest to note that if $\hat{\gamma}_0$ and $\hat{\gamma}_1$ in (8.8.2) – (8.8.3) are replaced by the true values of γ_0 and γ_1, the results above do not change in any substantial way.

Oral and Tiku (2003) study the estimator $\hat{\gamma}_1$ given in (4.15.10). They show that the estimator is remarkably efficient and robust.

8.9 ROBUSTNESS OF AUTOREGRESSION ESTIMATORS

Consider the autoregressive model (5.2.2). To evaluate the robustness properties of the MMLE of the parameters μ, δ, ϕ and σ, we first consider skew distributions represented by the Gamma family $G(k, \sigma)$ as in (5.3.1). For k known, the MMLE are given in (5.4.7) – (5.4.9). The corresponding LSE are given in (5.6.1) – (5.6.2). Consider, for illustration, $k = 3$. Thus, the assumed population model is that a_t in (5.2.2) are iid and have the Gamma distribution $G(3, \sigma)$. Because of the intricate nature of the autoregressive model (5.2.2), particularly in situations when the innovations a_t have a skew distribution, we confine to a smaller departure from an assumed value of k than was possible for the simple linear regression model (3.2.1); this is prudent in the application of both the MMLE as well as the LSE of μ, δ, ϕ and σ. We consider the following alternatives to the assumed $G(3, \sigma)$ innovations.

Gamma (k, σ): (1) $k = 2$, (2) $k = 3.5$,

Outlier model: (3) $(n - k_1)G(3, \sigma)$ and $k_1G(3, 4\sigma)$, $k_1 = [0.5+0.1n]$,

Mixture model: (4) $0.90G(3, \sigma) + 0.10G(3, 4\sigma)$, and (8.9.1)

Contamination model: (5) $0.90G(3, \sigma) + 0.1U(-0.5, 0.5)$.

It may be noted that the MMLE $\hat{\sigma}$ and LSE $\tilde{\sigma}$ are both estimating $\tau\sigma$. The values of τ are given in Table 8A.4.

The simulated means and variances of the MMLE and the LSE are given in Table 8A.4 (Appendix A). The design points x_t are the same as in (5.5.4) and the model considered for y_0 is

$$\text{Model B: } y_0 = a_0/\sqrt{1-\phi^2}\,. \tag{8.9.2}$$

The random error a_0 has the same distribution as that of a_t $(1 \le t \le n)$. The results under Vinod-Shenton Model A: $y_0 = 0$, are not different in any substantial way for the MMLE and the LSE of δ, ϕ and σ; the bias in $\hat{\mu}$ and $\tilde{\mu}$ reduces by a margin, however.

Akkaya and Tiku (2001a) work with model B and generate the design points x_t only once and hold them common to all the N = [100,000/n] number of y-samples generated. For the values in Table 8A.4, we generate (x_t, y_t) simultaneously (this may be more meaningful in numerous applications in business and economics). That explains the slight differences in the values given in their Table 3 and those in Table 8A.4. Realize that $\hat{\sigma}$ and $\tilde{\sigma}$ both are estimating $\tau\sigma$; τ has absolutely no role to play in the computations as said earlier.

It can be seen that the MMLE $\hat{\mu}$, $\hat{\delta}$, $\hat{\phi}$ and $\hat{\sigma}$ are enormously more efficient than the corresponding LSE.

Hypothesis testing: We consider testing H_0: $\delta = 0$ against H_1: $\delta \ne 0$, and evaluate the robustness properties of the T_1 test (equation 5.7.3) and the G_1 test (equation 5.7.6). We give in Table 8.14 the values of the probabilities (power)

$$P\{|T_1| \ge z_\alpha \mid \delta\} \quad \text{and} \quad P\{|G_1| \ge z_\alpha \mid \delta\}. \tag{8.9.3}$$

The population model is Gamma $G(3, \sigma)$ and the alternatives are the ones given in (8.9.1). The random numbers geneated were standardized to have variance 1. It can be seen that the T_1 test has considerably higher power than the G_1 test and is robust.

Turker (2002) generalizes the results above to k autoregressive models. In particular, she develops a test for H_0: $\delta_1 = \delta_2 = \ldots = \delta_k$ against H_1: $\delta_i \ne \delta_j$ and shows that the T test based on the MMLE is robust and considerably more powerful than the G test based on the LSE. She considers both situations: σ_i $(1 \le i \le k)$ are all equal, and σ_i are not necessarily equal.

Table 8.14: Power of the tests; presumed type I error is 0.050; $\phi = 0.5$.

		$\delta = 0.0$	0.05	0.10	0.15	0.20	0.25	0.30
					n = 50			
True model	T	0.048	0.09	0.24	0.46	0.70	0.85	0.93
	G	0.047	0.08	0.15	0.29	0.45	0.62	0.78
Alternative (1)	T	0.033	0.08	0.25	0.52	0.76	0.91	0.96
	G	0.062	0.09	0.16	0.29	0.46	0.63	0.76
(2)	T	0.056	0.11	0.25	0.44	0.67	0.84	0.94
	G	0.054	0.08	0.16	0.29	0.46	0.63	0.77
(3)	T	0.049	0.06	0.10	0.18	0.29	0.41	0.54
	G	0.050	0.05	0.08	0.12	0.18	0.24	0.32
(4)	T	0.050	0.06	0.10	0.17	0.28	0.39	0.54
	G	0.049	0.06	0.07	0.11	0.16	0.23	0.32

(5)	T	0.066	0.08	0.10	0.16	0.25	0.36	0.47
	G	0.055	0.05	0.08	0.10	0.16	0.23	0.31
				n = 200				
True model	T	0.034	0.24	0.74	0.98	1.00	1.00	1.00
	G	0.042	0.12	0.43	0.77	0.95	1.00	1.00
Alternative (1)	T	0.036	0.29	0.84	0.99	1.00	1.00	1.00
	G	0.054	0.14	0.42	0.77	0.95	0.99	1.00
(2)	T	0.054	0.24	0.74	0.98	1.00	1.00	1.00
	G	0.070	0.19	0.45	0.76	0.94	0.99	1.00
(3)	T	0.048	0.11	0.33	0.61	0.84	0.95	1.00
	G	0.052	0.08	0.18	0.30	0.49	0.69	0.85
(4)	T	0.046	0.11	0.31	0.63	0.86	0.96	0.99
	G	0.036	0.08	0.16	0.34	0.51	0.68	0.84
(5)	T	0.036	0.08	0.25	0.53	0.78	0.94	0.98
	G	0.060	0.08	0.16	0.29	0.49	0.66	0.81

The differences in the values of the power for other values of ϕ are more or less of the same order**.** The performance of the MMLE as compared to the LSE is compelling.

STS distributions: Consider the situation when a_t in (5.2.2) are iid and have one of the short-tailed symmetric distributions (5.8.1). The MMLE of μ, δ, ϕ and σ are given in (5.8.4) – (5.8.5). To evaluate the robustness of the MMLE as compared to the LSE we consider, for illustration, the situation when r = 2 and d = 0 (λ=1) in (8.6.1) with kurtosis $\mu_4/\mu_2^2 = 2.44$; this will be called population model. As plausible alternatives (sample models), we consider the Tukey lambda-family (7.6.17) with

$$(1)\ l = 0.585, \quad (2)\ l = 1.00, \quad (3)\ l = 1.45, \text{ and } (4)\ l = 2.82$$

with kurtosis 2, 1.80, 1.75 and 2, respectively. The lambda-family, or for that matter (5.8.1), can also be used for modelling samples containing inliers (Tiku et al., 2001). The lambda-family (7.6.17) with

$$u_t = \{(y_t - \phi y_{t-1}) - \mu - \delta(x_t - \phi x_{t-1})\}/\sigma$$

is not amenable to maximum likelihood or modified likelihood estimation of the parameters. We are, therefore, using it only for an assessment of robustness.

Given in Table 8A.5 (Appendix A) are the means and the variances of the MMLE. Also given are the relative efficiencies (100 times the ratio of the variance of the MMLE to the variance of the LSE). The means of the LSE are essentially the same as those of the MMLE and are not, therefore, reported. The estimators $\hat{\sigma}$ and $\tilde{\sigma}$ are both estimating $\tau\sigma$, τ being the square root of the variance of the sample model to the variance of the population model as said earlier. For the population model, $\tau = 1$. The design points x_t were taken to be the same as in (5.5.4) and $y_0 = a_0/\sqrt{(1-\phi^2)}$, a_0 having the same STS distribution as a_t $(1 \le t \le n)$. Realize that $E(a_t)=0$.

The MMLE and LSE of μ and δ have negligible bias; the means of these estimators are not, therefore, reported. The bias in $\tilde{\phi}$ and $\tilde{\sigma}$ is essentially the same as in the MMLE $\hat{\phi}$ and $\hat{\sigma}$. It can be seen that the LSE are considerably less efficient, and their efficiency decreases as n increases. For illustration, we give in Table 8A.5 the values for $\mu = 0$, $\delta = 1$ and $\phi = 0.5$; $\sigma = 1$ without loss of generality. Similar results, of course, are true for other STS distributions, e.g., the family (5.8.1) with r = 4 and d = 0. We conclude that the MMLE are robust (Akkaya and Tiku, 2002b) which is due to the inverted-umbrella ordering of the weights given to the ordered residuals $a_{(i)}$, $1 \le i \le n$. For r = 2 and d = 0 in (8.6.1), for example, the first ten β_i coefficients for n = 20 are, $\beta_{n-i+1} = \beta_i$:

$$1.06,\ 0.97,\ 0.87,\ 0.74,\ 0.61,\ 0.46,\ 0.31,\ 0.17,\ 0.070, 0.008. \qquad (8.9.4)$$

The middle ordered residuals, therefore, receive small weights. This depletes the influence of inliers and short tails (Akkaya and Tiku, 2002a, b, c).

Hypothesis testing: In the robustness framework, to test H_0: $\delta = 0$, the statistic based on the MMLE is defined as (ϕ in $u_{[i]}$ is replaced by the MMLE $\hat{\phi}$)

$$T_2 = \sqrt{\sum_{i=1}^{n} \beta_i \,(u_{[i]} - \bar{u}_{[.]})^2}\ (\hat{\delta}/\hat{\sigma}). \qquad (8.9.5)$$

The MMLE $\hat{\delta}$ and $\hat{\sigma}$ are given in (5.8.4) – (5.8.5). Large values of $|T_2|$ lead to the rejection of H_0 in favour of H_1: $\delta \ne 0$. For d > 0 ($\lambda > 1$), α_i and β_i are replaced by α_i^* and β_i^*, respectively.

Since $\hat{\phi}$ converges to ϕ as n tends to infinity, in view of Lemma 5.2, the asymptotic null distribution of T_2 is normal N(0, 1). For small n (≤ 60) the null distribution of $T_2/\sqrt{V_2}$, V_2 being the simulated variance of T_2 (exact variance is difficult to determine) is closely approximated by the normal N(0, 1). Realize that for the design points (5.5.4), $\bar{u}_{[.]} \cong 0$.

The statistic based on the LSE is G_2 as in (5.10.5). The asymptotic null distribution of G_2 is normal N(0, 1). For $n \le 60$ the null distribution of $G_2/\sqrt{V}$, V being the simulated variance of G_2 (exact variance is again difficult to determine) is closely approximated by N(0, 1). For the design points (5.5.4), $\bar{u} \cong 0$.

The T_2 test is robust and has considerably higher power than the G_2 test (Akkaya and Tiku, 2002b).

LTS distributions: Suppose that the distribution of a_t ($1 \le t \le n$) is one of the members of the long-tailed symmetric family of distributions (5.11.1). The MMLE are given in (5.11.5)-(5.11.6).

Consider the situation when p = 3.5 in (5.11.1). This is the assumed population model. The sample models are taken to be (1) – (9), given in (8.2.1) – (8.2.4). We give the simulated means and variances of the MMLE in Table 8.15 for models (1), (3), (6), (8) and (9) for illustration. The design points are as in (5.5.4) and $y_0 = a_0/\sqrt{(1-\phi^2)}$. It can be seen that the MMLE are considerably more efficient than the LSE. The relative efficiencies of the LSE are essentially the same for other designs, e.g., x_t generated from a normal N(0, 1). See also Tan and Lin (1993).

Table 8.15: Values of (a) Mean and (b) n(Variance), of the MML and the LS estimators; $\delta = 1$, $\phi = 0.5$, $\sigma = 1$.

		n = 50				n = 100			
		μ	δ	ϕ	σ	μ	δ	ϕ	σ
					True model, $\tau = 1$				
(a)	MML	0.002	1.000	0.459	1.014	0.004	1.002	0.473	1.008
	LS	0.001	0.999	0.454	0.989	0.004	1.000	0.474	0.995

(b)	MML	1.068	0.553	0.752	0.828	1.070	0.569	0.769	0.975
	LS	1.225	0.632	0.826	1.114	1.181	0.636	0.854	1.138
					Model (1), $\tau = 1$				
(a)	MML	– 0.009	1.001	0.450	0.978	0.001	0.997	0.483	0.947
	LS	0.003	1.005	0.457	0.944	0.003	1.000	0.480	0.947
(b)	MML	1.634	0.490	0.721	5.449	1.220	0.406	0.575	4.061
	LS	1.122	0.642	0.774	4.770	1.026	0.597	0.727	4.843
					Model (3), $\tau = 1$				
(a)	MML	0.006	1.000	0.459	1.040	0.001	1.001	0.484	1.003
	LS	– 0.001	0.995	0.460	0.992	0.008	0.999	0.477	0.999
(b)	MML	1.124	0.578	0.790	0.748	1.109	0.575	0.722	0.727
	LS	1.271	0.658	0.791	0.818	1.037	0.590	0.758	0.885
					Model (6), $\tau = 1.581$				
(a)	MML	0.001	0.992	0.435	1.491	– 0.003	1.001	0.484	1.469
	LS	– 0.005	1.000	0.439	1.506	0.004	0.999	0.477	1.524
(b)	MML	1.124	0.578	0.790	0.748	1.109	0.575	0.722	6.675
	LS	1.271	0.658	0.791	0.818	1.037	0.590	0.758	8.584
					Model (8), $\tau = 1.581$				
(a)	MML	– 0.006	1.000	0.455	1.524	0.009	0.998	0.482	1.468
	LS	– 0.008	1.004	0.459	1.507	– 0.007	1.002	0.477	1.556
(b)	MML	3.775	1.050	0.641	8.058	2.631	0.905	0.498	10.226
	LS	2.892	1.590	0.744	8.994	2.730	1.671	0.743	11.716
					Model (9), $\tau = 0.953$				
(a)	MML	0.005	0.997	0.459	0.952	0.006	0.997	0.478	0.941
	LS	– 0.003	0.997	0.456	0.943	– 0.003	1.002	0.476	0.947
(b)	MML	0.911	0.466	0.686	0.870	0.801	0.427	0.615	0.800
	LS	1.105	0.566	0.774	1.106	1.018	0.510	0.739	1.140

Hypothesis testing: To test H_0: $\delta = 0$, the statistic based on the MMLE is (ϕ in $u_{[.]}$ is replaced by $\hat{\phi}$)

$$T_3 = \sqrt{(2p/k) \sum_{i=1}^{n} \beta_i \,(u_{[i]} - \bar{u}_{[.]})^2}\ (\hat{\hat{\delta}}/\hat{\sigma}); \qquad (8.9.6)$$

$k = 2p - 3$, $p \geq 2$. The estimators $\hat{\hat{\delta}}$ and $\hat{\sigma}$ are given in (5.11.5) – (5.11.6). The coefficients β_i are replaced by β_i^* ($1 \leq i \leq n$) if C in (5.11.7) assumes a negative value as explained earlier. The statistic based on the LSE is G_3, given in (5.12.6).

Given in Table 8.16 are the values of the power of the T_3 and G_3 tests, for testing H_0: $\delta = 0$ against H_1: $\delta \neq 0$. The null distributions of T_3 and G_3 are referred to the Student t with $\nu = n - 3$ degrees of freedom. For $n > 30$, the null distributions are referred to the normal N(0, 1). The assumed population model is the LTS distribution (5.11.1) with $p = 3.5$. The alternatives are the ones considered in Table 8.15. We have given values only for $\phi = 0.5$. The differences between the values of the power of the T_3 and G_3 tests are essentially the same as for other values of ϕ ($-1 < \phi < 1$). It can be seen that the T_3 test is robust and has the power superiority.

Table 8.16: Power of (a) the T_3 test and (b) the G_3 test, for long-tailed symmetric distribuions; ϕ = 0.5.

	n = 50				n = 100			
	δ = 0.00	0.10	0.20	0.30	0.00	0.10	0.20	0.30
				True model				
(a)	0.056	0.18	0.49	0.80	0.052	0.30	0.78	0.98
(b)	0.056	0.15	0.42	0.76	0.049	0.24	0.71	0.96
				Model (1)				
(a)	0.062	0.23	0.63	0.89	0.048	0.37	0.89	0.99
(b)	0.056	0.17	0.51	0.80	0.046	0.29	0.75	0.96
				Model (3)				
(a)	0.062	0.18	0.48	0.80	0.047	0.27	0.75	0.97
(b)	0.058	0.16	0.43	0.75	0.056	0.27	0.75	0.96
				Model (6)				
(a)	0.052	0.23	0.62	0.88	0.040	0.41	0.92	1.00
(b)	0.056	0.18	0.50	0.79	0.051	0.31	0.74	0.96
				Model (8)				
(a)	0.055	0.22	0.63	0.90	0.055	0.38	0.89	0.99
(b)	0.048	0.19	0.50	0.79	0.049	0.29	0.75	0.95
				Model (9)				
(a)	0.052	0.16	0.51	0.83	0.053	0.31	0.81	0.99
(b)	0.049	0.16	0.48	0.80	0.052	0.24	0.74	0.96

Time series models: Tiku et al. (1999, 2000) study the robustness of the MMLE and the tests based on them for the time series AR(1) model (5.13.1). They show that the MMLE have definite advantages over the LSE for skew and long-tailed distributions. The test of H_0: ϕ = 0 based on the MMLE are powerful and robust to plausible alternatives to an assumed distribution. Similar results are true for symmetric short-tailed distributions (Akkaya and Tiku, 2002a). It may be noted that the LSE $\tilde{\sigma}$ is particulary inefficient for LTS distributions, and the LSE $\tilde{\phi}$ is particulary inefficient for STS distributions.

8.10 ROBUSTNESS IN EXPERIMENTAL DESIGN

As said earlier, experimental design is a very important area in statistics. It is, therefore, very important to have robust tests for treatment effects since the error distribution will not be known exactly. Şenoğlu and Tiku (2001) study the robustness properties of the F^* test (6.5.2) based on the MMLE and compare it with the corresponing F test (6.5.1) based on the LSE (normal-theory estimators). They proceed as follows.

Suppose that the distribution of the errors e_{ij} in (6.2.1) have the Generalized Logistic distribution GL(b, σ). In practice, b might be misspecified. Moreover, data often contain outliers. For illustration, suppose that b = 1 so that the underlying distribution is logistic. To represent a large number of plausible alternatives to the logistic, Şenoğlu and Tiku (2001) consider the following,

GL(b, σ): (1) b = 0.5, (2) b = 1.5, Single outlier: (3) (n – 1) observations come from GL(1, σ) and one observation (we do not know which) comes from GL(1, 4σ),

Mixture model: (4) 0.90GL(1, σ) + 0.10 GL(1, 4σ),

Contamination model: (5) 0.90GL(1, σ) + 0.10Uniform (– 1, 1).

Since the assumed model is GL(1, σ), the coefficients α_j and β_j $(1 \le j \le n)$ used in the computation of F^* are obtained from (6.4.4). The simulated values of the type I error and power of the F^* and F tests are given in Table 8.17, reproduced from Şenoğlu and Tiku (2001, p. 1346). It can be seen that the F^* test maintains higher power almost always besides having considerably smaller type I error for the outlier and mixture models (4) and (5) above.

The F^* test based on the MMLE for testing interaction effects has similar efficiency and robustness properties. See Şenoğlu and Tiku (2001).

Table 8.17: Values of the type I error and power for the F^* and F tests; n = 10 and c = 4 (four blocks).

	F*	F	F*	F	F*	F
	True model		Alternative model			
d	GL(1, σ)		(1)		(2)	
0.0	0.043	0.044	0.041	0.043	0.044	0.044
0.2	0.09	0.09	0.09	0.09	0.10	0.09
0.4	0.28	0.27	0.29	0.28	0.29	0.28
0.6	0.59	0.57	0.61	0.58	0.59	0.57
0.8	0.85	0.83	0.86	0.83	0.85	0.83
1.0	0.97	0.96	0.97	0.95	0.97	0.96
	Alternative model					
	(3)		(4)		(5)	
0.0	0.019	0.028	0.029	0.044	0.048	0.046
0.2	0.07	0.09	0.08	0.10	0.10	0.10
0.4	0.32	0.32	0.34	0.34	0.27	0.26
0.6	0.67	0.63	0.68	0.64	0.56	0.54
0.8	0.89	0.84	0.88	0.83	0.82	0.80
1.0	0.97	0.94	0.96	0.93	0.95	0.94

Assuming any other population model, e.g. Weibull W(p, σ) error distribution, the F^* test has efficiency and robustness properties similar to those in Table 8.17. For outlier and mixture models, the F^* test has a double advantage: it has not only smaller type I error but has also higher power than the F test. This is an interesting phenomenon indeed.

For testing a linear contrast, Şenoğlu and Tiku (2002) develop tests using both the MMLE and the LSE, the error distributions from block to block being not necessari1y identical. They show that the test based on the MMLE is remarkably powerful and robust.

Comment: There are a number of rank-based methods which provide robust estimators for long-tailed as well as short-tailed distributions. They are not, however, as sharp as the parametric robust methods (Şenoğlu and Tiku, 2001, p. 1343). Moreover, rank-based methods do not have the generality of the parametric robust methods; see, for example, Dunnett (1982), Şenoğlu and Tiku (2002), Akkaya and Tiku (2004).

SUMMARY

In this Chapter, we consider the very important issue of robustness. In most statistical applications, the objective is to obtain estimators and hypothesis testing procedures which have certain optimal properties with respect to an assumed distribution. In practice, however, an assumed distribution is hardly ever exactly correct. What is, in fact, more realistic is the premise that the underlying distribution is in reasonable proximity to the one assumed. The latter can be identified by resorting to Q-Q plots or goodness-of-fit tests (Chapter 9).

That brings the robustness issue in focus. An estimator is called robust if it is fully efficient (or nearly so) for an assumed distribution but maintains high efficiency for plausible alternatives. The latter include all the distributions which are in reasonable proximity to the one assumed. In this Chapter, we study the robustness of all the MMLE developed in previous chapters. We show that the MMLE are remarkably robust. This is due to the fact that the MMLE of μ and σ in an assumed location-scale distribution $(1/\sigma)f((y-\mu)/\sigma)$ are essentially of the form $\hat{\mu} = \sum_{i=1}^{n}\beta_i y_{(i)} / m$ $(m = \sum_{i=1}^{n}\beta_i)$ and $\hat{\sigma} = \sqrt{\left\{\sum_{i=1}^{n}\beta_i(y_{(i)}-\hat{\mu})^2/(n-1)\right\}}$. The coefficients β_i have half-umbrella ordering (i.e., they decrease in the direction of the long tail), umbrella ordering (i.e., they increase until the middle value and then decrease in a symmetric fashion), and inverted umbrella ordering (i.e., they decrease until the middle value and then increase in a symmetric fashion). Thus, the extreme order statistics $y_{(i)}$ and extreme deviations $(y_{(i)}-\hat{\mu})^2$ representing the tail(s) of skew and long-tailed symmetric distributions automatically receive small weights under half-umbrella and umbrella orderings, respectively, and the middle order statistics receive small weights under inverted umbrella ordering. This gives the MMLE the inherent robustness properties they have. The nature of the β_i $(1 \le i \le n)$ coefficients also makes them robust to a moderate number of outliers and inliers in a sample. The hypothesis testing procedures based on the MMLE also inherit these beautiful robustness properties. We also show that the Huber M-estimators and the tests based on them are inadequate for skew and short-tailed symmetric distributions. They have good efficiency and robustness properties only for long-tailed symmetric distributions. But then the MMLE $\hat{\mu}$ and $\hat{\sigma}$ and the tests based on them are equally efficient and robust for all alternative distributions which have finite mean and variance or in situations when the sample has moderate number of outliers. For very extreme and rare situations (e.g., the distribution has nonexistent mean or variance or the sample contains a large number of outliers), the estimators $\hat{\mu}_c$ and $\hat{\sigma}_c$ based on normal samples with about thirty percent smallest and largest observations censored have excellent robustness properties. In most situations, however, the underlying distribution has finite mean and variance and that makes MMLE very attractive realizing also that no observation is censored implicitly or explicitly in their application and that they are defined for all the three types of distributions: skew, short-tailed symmetric, and long-tailed symmetric. It may also be remembered that the Huber M-estimators of σ can have substantial downward bias even if the underlying distribution is long-tailed symmetric. There is no such problem with the MMLE $\hat{\sigma}$.

APPENDIX 8A

SIMULATED MEANS, VARIANCES AND EFFICIENCIES

Table A8.1: Means, variances and mean square errors of MMLE and M-estimators w24; σ =1.

	Var		Mean		MSE	
n	$\hat{\mu}$	$\hat{\mu}_w$	$\hat{\sigma}$	$\hat{\sigma}_w$	$\hat{\sigma}$	$\hat{\sigma}_w$
			True model: τ = 1			
10	0.090	0.091	1.062	0.857	0.106	0.092
20	0.043	0.044	1.039	0.901	0.047	0.045
40	0.021	0.022	1.019	0.917	0.020	0.024
60	0.014	0.015	1.012	0.921	0.013	0.017
			Model (1): τ = 1			
10	0.069	0.061	0.948	0.684	0.207	0.162
20	0.029	0.028	0.917	0.712	0.115	0.113
40	0.016	0.013	0.920	0.725	0.066	0.090
60	0.012	0.009	0.958	0.726	0.073	0.084
			Model (2): τ = 1			
10	0.082	0.079	1.030	0.794	0.167	0.113
20	0.038	0.038	0.986	0.822	0.059	0.066
40	0.018	0.018	0.969	0.841	0.024	0.042
60	0.012	0.012	0.963	0.842	0.023	0.036
			Model (3): τ = 1			
10	0.096	0.100	1.078	0.890	0.093	0.082
20	0.047	0.047	1.062	0.937	0.042	0.038
40	0.023	0.023	1.041	0.950	0.021	0.019
60	0.016	0.016	1.039	0.958	0.014	0.013
			Model (4): τ = 1			
10	0.105	0.110	1.098	0.928	0.079	0.067
20	0.053	0.053	1.085	0.974	0.040	0.030
40	0.025	0.025	1.082	0.998	0.023	0.014
60	0.018	0.018	1.078	1.002	0.017	0.010
			Model (5): τ = 1.140			
10	0.109	0.106	1.184	0.926	0.147	0.135
20	0.052	0.053	1.153	0.968	0.063	0.074
40	0.025	0.026	1.127	0.998	0.028	0.046
60	0.017	0.017	1.108	0.991	0.020	0.037
			Model (6): τ = 1.581			
10	0.149	0.117	1.493	0.967	0.463	0.488
20	0.065	0.056	1.429	1.011	0.199	0.380
40	0.029	0.028	1.351	1.035	0.113	0.325
60	0.027	0.019	1.524	1.034	0.104	0.318

			Model (7): τ = 1.581			
10	0.164	0.133	1.561	1.051	0.341	0.384
20	0.073	0.066	1.490	1.104	0.136	0.278
40	0.035	0.033	1.424	1.130	0.073	0.229
60	0.024	0.023	1.391	1.136	0.066	0.215
			Model (8): τ = 1.581			
10	0.160	0.124	1.488	0.995	0.573	0.491
20	0.068	0.060	1.421	1.020	0.243	0.382
40	0.030	0.029	1.342	1.037	0.145	0.328
60	0.026	0.018	1.493	1.038	0.139	0.316
			Model (9): t=0.953			
10	0.079	0.080	1.002	0.799	0.101	0.094
20	0.038	0.039	0.977	0.833	0.045	0.049
40	0.019	0.019	0.956	0.851	0.021	0.028
60	0.012	0.013	0.948	0.855	0.012	0.021

Table A8.2: Mean and variance of the MMLE and the relative efficiency of the LSE, for short-tailed symmetric distributions; σ = 1.

	Mean*		n(Variance)		RE	
n	$\hat{\sigma}$	s	$\hat{\mu}$	$\hat{\sigma}$	$\bar{y}$	s
			STS (r = 4, d = 0), τ = 1			
10	0.94	0.97	2.291	0.355	90	88
20	0.97	0.98	2.107	0.327	85	91
40	0.98	0.99	2.259	0.309	84	90
100	0.99	0.99	2.055	0.308	83	90
			Normal, τ = 0.623			
10	0.59	0.60	1.172	0.201	114	98
20	0.62	0.62	1.143	0.211	112	106
40	0.63	0.62	1.121	0.218	115	108
100	0.63	0.62	1.111	0.216	119	109
			Tukey (l=0.585), τ = 0.528			
10	0.49	0.52	0.541	0.068	76	77
20	0.57	0.52	0.510	0.061	72	75
40	0.51	0.53	0.490	0.055	67	72
100	0.51	0.53	0.484	0.051	66	72
			Tukey (l=1), τ = 0.360			
10	0.33	0.35	0.211	0.024	64	70
20	0.34	0.36	0.194	0.020	57	65
40	0.34	0.36	0.185	0.018	53	61
100	0.35	0.36	0.166	0.015	53	60

			Tukey (l=1.45) $\tau = 0.255$			
10	0.24	0.25	0.103	0.012	61	67
20	0.24	0.26	0.088	0.009	54	62
40	0.24	0.25	0.088	0.008	50	59
100	0.24	0.26	0.083	0.007	48	57
			Model (3), k = 1.5, $\tau = 0.166$			
10	0.14	0.15	0.047	0.006	83	82
20	0.14	0.15	0.045	0.005	79	78
40	0.15	0.15	0.045	0.005	79	77
100	0.15	0.15	0.045	0.004	77	77
			Model (3), k = 2.0, $\tau = 0.139$			
10	0.12	0.12	0.039	0.039	96	88
20	0.12	0.13	0.039	0.039	94	87
40	0.13	0.13	0.040	0.040	96	86
100	0.13	0.13	0.037	0.037	96	85
			Model (3), k = 3.0, $\tau = 0.105$			
10	0.09	0.10	0.028	0.005	114	95
20	0.10	0.10	0.029	0.005	115	100
40	0.10	0.10	0.029	0.005	115	100
100	0.10	0.10	0.028	0.005	118	100
			Model (4), k = 1.5, $\tau = 0.204$			
10	0.18	0.20	0.045	0.004	44	56
20	0.19	0.20	0.037	0.003	35	52
40	0.19	0.20	0.034	0.003	30	51
100	0.19	0.20	0.028	0.002	27	49
			Model (4), k = 2.0, $\tau = 0.220$			
10	0.19	0.22	0.043	0.003	35	47
20	0.20	0.22	0.033	0.002	25	44
40	0.20	0.22	0.025	0.002	20	44
100	0.20	0.22	0.022	0.002	17	44

* Since $\hat{\mu}$ and $\bar{y}$ are both unbiased estimators of μ, their means are not given.

Table A8.3: Variance of the MMLE and the relative efficiency of the LSE; θ_1=1 and σ = 1.

					Model			
	True model		(1)		(2)		(3)	
n	Var	RE	Var	RE	Var	RE	Var	RE
20	0.533	90	0.416	66	0.500	79	0.596	96
30	0.301	88	0.218	61	0.259	75	0.349	96
50	0.208	85	0.174	73	0.187	79	0.226	92

	Model (4)		Model (5)		Model (6)		Model (7)	
20	0.651	104	0.585	89	0.611	85	0.565	86
30	0.367	105	0.332	86	0.350	74	0.344	78
50	0.255	104	0.232	78	0.334	67	0.246	64
	Model (8)		Model (9)		Model (10)		Model (11)	
20	0.904	56	0.463	84	0.845	56	0.822	65
30	0.465	51	0.268	82	0.798	52	0.556	60
50	0.385	66	0.174	80	0.281	52	0.392	58

Table A8.4: Values of (1) Mean and (2) n(Variance) of MMLE and LSE; δ = 1.0, ϕ = 0.5. True model is Gamma G(k, σ); k = 3 and σ = 1.

Coeficient		n = 30 (1)	n = 30 (2)	n = 50 (1)	n = 50 (2)	n = 100 (1)	n = 100 (2)
		True model, τ = 1					
δ	MML	0.998	1.145	0.999	0.952	1.003	0.856
	LS	1.001	1.978	0.994	1.850	0.997	1.972
ϕ	MML	0.466	0.493	0.480	0.444	0.492	0.367
	LS	0.431	0.777	0.465	0.740	0.488	0.766
σ	MML	0.972	0.617	0.977	0.582	0.983	0.559
	LS	0.979	0.964	0.988	0.948	0.997	0.985
		Model (1), τ = 0.816					
δ	MML	1.000	0.361	1.000	0.272	1.003	0.229
	LS	0.997	0.993	0.998	0.913	1.007	0.909
ϕ	MML	0.473	0.316	0.487	0.254	0.493	0.227
	LS	0.430	0.809	0.462	0.701	0.481	0.695
σ	MML	0.619	0.346	0.620	0.330	0.614	0.304
	LS	0.691	0.674	0.694	0.699	0.706	0.751
		Model (2), τ = 1.080					
δ	MML	1.008	2.127	0.997	1.800	1.003	1.787
	LS	0.991	2.959	0.997	3.024	1.004	2.776
ϕ	MML	0.456	0.590	0.480	0.536	0.491	0.489
	LS	0.430	0.734	0.458	0.738	0.478	0.754
σ	MML	1.246	0.898	0.261	0.934	1.283	0.920
	LS	1.199	1.231	1.212	1.175	1.222	1.258
		Model (3^*), τ = 1.931					
δ	MML	1.001	1.488	1.007	1.602	0.993	2.324
	LS	0.991	7.198	1.006	5.955	0.993	5.231
ϕ	MML	0.542	0.191	0.583	0.179	0.616	0.270
	LS	0.599	0.535	0.721	0.423	0.808	0.320
σ	MML	1.556	3.751	1.522	3.364	1.526	3.533
	LS	2.100	15.414	2.039	14.120	1.910	12.049

			Model (4*), τ =1.931				
δ	MML	0.998	1.620	1.002	1.455	0.999	1.368
	LS	1.008	9.690	1.006	9.194	1.006	8.689
ϕ	MML	0.482	0.241	0.491	0.178	0.503	0.122
	LS	0.431	0.637	0.458	0.620	0.479	0.596
σ	MML	1.534	5.724	1.524	5.816	1.518	5.405
	LS	2.077	20.968	2.119	21.606	2.171	24.239
			Model (5), τ =1.083				
δ	MML	0.998	1.831	0.999	1.705	0.995	1.433
	LS	1.004	2.418	1.002	2.166	1.003	2.228
ϕ	MML	0.454	0.713	0.475	0.591	0.488	0.490
	LS	0.430	0.796	0.457	0.788	0.482	0.732
σ	MML	1.128	0.720	1.145	0.644	1.161	0.616
	LS	1.070	1.014	1.073	1.034	1.082	1.031

Table A8.5: Means and variances of the MMLE and the relative efficiency of the LSE; μ = 0, δ = 1 and ϕ = 0.5; σ = 1.

	Mean*		n(Variance)				Relative efficiency			
n	$\hat{\phi}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\delta}$	$\hat{\phi}$	$\hat{\sigma}$	E_1	E_2	E_3	E_4
			Population model (r = 2, d = 0)							
30	0.429	0.925	2.65	1.34	0.822	0.335	90	95	94	93
50	0.459	0.954	2.23	1.31	0.764	0.316	88	90	92	94
100	0.479	0.972	1.99	1.04	0.713	0.322	87	90	92	94
			Sample model (1), τ = 0.59							
30	0.434	0.545	0.81	0.56	0.675	0.076	77	80	82	82
50	0.466	0.562	0.66	0.28	0.583	0.074	73	77	78	79
100	0.482	0.574	0.55	0.32	0.575	0.069	70	73	75	78
			Sample model (2), τ =0.40							
30	0.438	0.369	0.32	0.22	0.687	0.027	65	70	73	71
50	0.468	0.379	0.25	0.13	0.536	0.025	62	65	69	70
100	0.480	0.386	0.24	0.13	0.529	0.025	62	64	64	68
			Sample model (3), τ = 0.29							
30	0.450	0.262	0.14	0.07	0.558	0.013	62	66	69	70
50	0.466	0.269	0.13	0.05	0.518	0.011	57	61	64	66
100	0.487	0.272	0.10	0.06	0.392	0.010	55	59	64	65
			Sample model (4), τ = 0.13							
30	0.481	0.122	0.03	0.02	0.173	0.004	79	81	98	79
50	0.476	0.125	0.03	0.02	0.369	0.003	75	78	84	76
100	0.490	0.128	0.03	0.02	0.381	0.003	71	75	82	73

* The bias in the MML and LS estimators of μ and δ are negligible. The means of the LSE of ϕ and σ are essentially the same as those of MMLE and are not, therefore, reported.

CHAPTER 9

Goodness-of-fit and Detection of Outliers

9.1 INTRODUCTION

The efficiency of a statistical procedure hinges on how accurately we can identify the underlying distribution. With limited information contained in a sample, despite our best efforts, it is not possible to identify the underlying distribution exactly. That impedes applications of classical procedures to real life problems. For example, the sample mean and variance have high efficiency only for symmetric distributions with kurtosis between 2.8 and 3.2 (Akkaya and Tiku, 2003). Robust statistical procedures have the advantage of maintaining high efficiency over a wider range of distributions. They have, therefore, a broader scope for real life applications. The question now is, given a random sample of size n, how do we identify a distribution which is in reasonable proximity to the true underlying distribution ? We also need to detect data anomalies (e.g., outliers, inliers, etc.) in a sample since their presence has an adverse effect on the efficiency of estimators, particularly the normal-theory estimators. An informal technique which can easily be implemented to achieve this goal is graph plotting and data exploratory techniques (Tukey, 1977; Mosteller and Tukey, 1977; Hoaglin et al, 1983; Hamilton, 1992), Q-Q plots being particularly useful. Formal techniques are goodness-of-fit tests and outlier and inlier detection procedures. We enunciate some of these procedures as follows.

9.2 Q-Q PLOTS

First consider a random sample $x_1, x_2, \ldots, x_n$ from a distribution of the type $(1/\sigma)f((x - \mu)/\sigma)$, μ and σ not known. The distribution of $z = (x - \mu)/\sigma$ is free of μ and σ and denoted by f(z). Let $F(z) = \int_{-\infty}^{z} f(z)\,dz$ be the cumulative distribution function. Assume a particular density f(z) and determine Q_i from the equation

$$F(Q_i) = i/(n + 1), \quad 1 \le i \le n; \tag{9.2.1}$$

Q_i are called population quantiles. Plot $x_{(i)}$ against Q_i $(1 \le i \le n)$. If it gives "close to a straight line" pattern, then f(z) qualifies to be a plausible model for the data. In fact, we should consider a few density functions f(z) and construct the corresponding Q-Q plots. The one that gives "closest to a straight line" pattern is the most plausible model for the sample; see, for example, Tiku and Vaughan (1997). A "close to a straight line" pattern will simply be called a pattern.

Example 9.1: Consider the 15 observations given in Example 2.9. Elveback et al. (1970) assume the underlying distribution to be logistic. Given in Fig. 1 is a Q-Q plot of the data. The plot has a pattern. We conclude that the logistic density provides a plausible model.

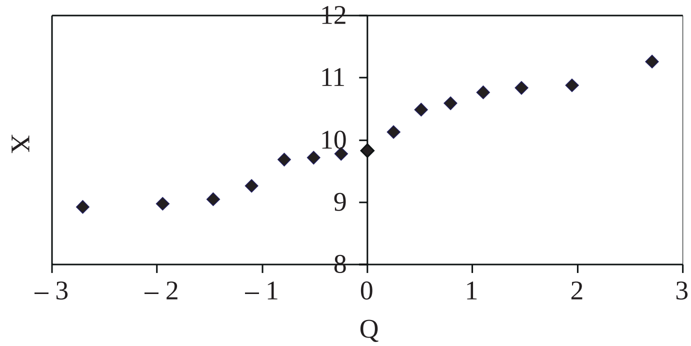

Figure 1 Q-Q Plot based on logistic density.

The BLUE calculated from the data are

$\mu^* = 10.010$ and $\sigma^* = 0.843$, with $V(\mu^*) = 0.063\,\sigma^2$ and $V(\sigma^*) = 0.048\,\sigma^2$.

The MMLE are

$\hat{\mu} = 10.011$ and $\hat{\sigma} = 0.841$, with $V(\hat{\mu}) = 0.063\,\sigma^2$ and $V(\hat{\sigma}) = 0.045\,\sigma^2$.

The usual assumption of normality gives

$\bar{x} = 10.012$ and $\sigma = 0.764$, with $V(\bar{x}) = 0.067\,\sigma^2$ and $V(s) \cong 0.053\,\sigma^2$.

The minimum variance bounds for estimating μ and σ are $0.062\,\sigma^2$ and $0.044\,\sigma^2$, respectively. The MMLE are clearly the most efficient, and much easier to compute than the BLUE.

Remark: The BLUE and MMLE above are adaptive in the terminology of Hogg (1982), Hogg et al. (1975) and Sprott (1978, 1982); see also Hogg (1967, 1972, 1974). An adaptive estimator utilizes the information in the sample about the shape of the underlying distribution. Hogg assumes that the underlying distribution is symmetric and uses the sample kurtosis to determine the shape of the underlying distribution. He then chooses an estimator accordingly. Sprott on the other hand assumes that the underlying distribution is of the type $(1/\sigma)f_\lambda((x-\mu)/\sigma)$, see Tiku et al. (1986, pp. 147-150). He works out the MLE $\hat{\mu}(\lambda)$ and $\hat{\sigma}(\lambda)$ as functions of λ. The MLE are obtained by replacing the shape parameter λ by a plausible estimate $\hat{\lambda}$. See also Frazer (1976).

Example 9.2: Consider the Darwin's data mentioned earlier. Arranged in ascending order of magnitude, the observations are

– 67 – 48 6 8 14 16 23 24 28 29 41 49 56 60 75

Here, no Q-Q plot has a pattern for any density f(z) belonging to the skew, STS or LTS families of distributions considered in previous chapters. However, the Q-Q plot based on the normal $N(\mu, \sigma^2)$ has an interesting property (Fig. 2): all the observations other than the two smallest and one largest are close to a straight line. This indicates that these observations are outliers, i.e., they are much different than the bulk of observations. Since the presence of outliers in the data adversely affects the efficiency of estimators, we give zero weights to these observations, i.e., we censor them. The MMLE based on the remaining twelve observations and their standard errors are given in Example 7.4. It can be seen that the MMLE $\hat{\mu}_c$ and $\hat{\sigma}_c$, besides being easy to compute, are considerably more efficient than the traditional estimators $\bar{x}$ and s, and some other adaptive estimators (Tiku et al., 1986, p.152). For example, Sprott

(1978, 1982) models the data by the Student t distribution with λ degrees of freedom and obtains a confidence interval for the location parameter μ; the shortest interval (11.560, 41. 749) is obtained by taking $\lambda = 2$. The confidence interval (12.696, 42.730) based on the MMLE calculated from the middle eleven observations is slightly shorter than that of Sprott (Tiku et al., 1986, p.152).

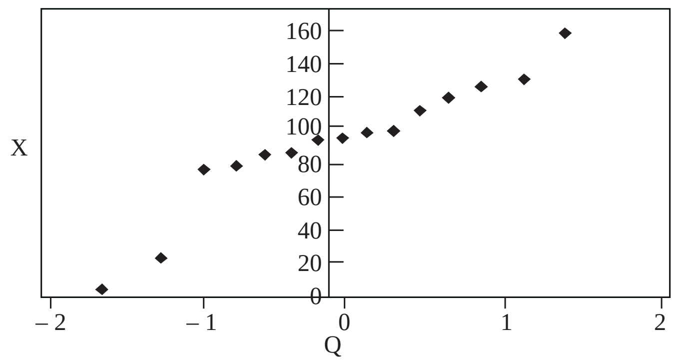

Fig. 2 Q-Q Plot based on normal density.

Example 9.3: The following ordered observations represent the survival times (the number of days/1000) of 43 patients suffering from granulocytic leukemia (Johnson and Johnson, 1979):

0.007	0.047	0.058	0.074	0.177	0.232	0.273	0.285	0.317	0.429	0.440
0.445	0.455	0.468	0.495	0.497	0.532	0.571	0.579	0.581	0.650	0.702
0.715	0.779	0.881	0.900	0.930	0.968	1.077	1.109	1.314	1.334	1.367
1.534	1.712	1.784	1.877	1.886	2.045	2.056	2.260	2.429	2.509	

The premise is that the survival times u come from the Weibull distribution (2.8.1) ($\theta = 0$). Now, $y = u^p$ is exponential and $y_{(i)} = u_{(i)}^p$ are the order statistics of a random sample of size n = 43 from the exponential $E(0, \delta)$, $\delta = \sigma^p$. A Q-Q plot is obtained by plotting $y_{(i)}$ against the quantiles of an exponential, $Q_i = -\ln(1 - q_i)$, $q_i = i/(n+1)$ $(1 \le i \le 43)$. A pattern emerges with p = 2.5 as can be seen from Fig. 3. The Weibull distribution with p = 2.5 is, therefore, a plausible model for the data.

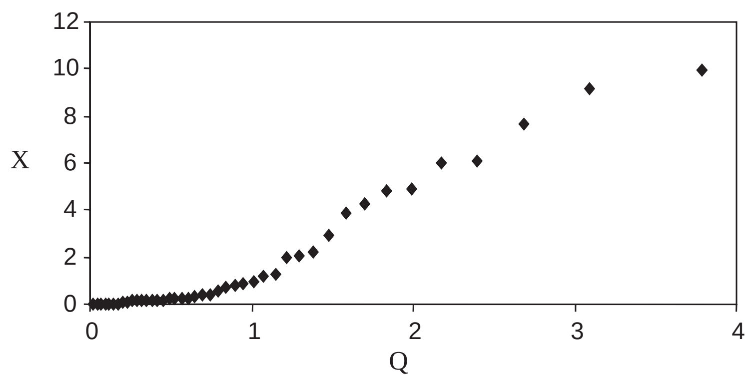

Fig. 3 Q-Q Plot based on Weibull density.

The MMLE of σ is obtained by using the techniques given in Chapter 2.

Example 9.4: Consider the data of Example 3.3. Here, the model is $y_i = \theta_0 + \theta_1 x_i + e_i$ $(1 \le i \le n)$; $V(e_i) = \sigma^2$. To have an idea about the error distribution, we use an easily computable estimator of θ_1, the LSE $\tilde{\theta}_1 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \,/\, \sum_{i=1}^{n}(x_i - \bar{x})^2$, and determine the order statistics $w_{(i)}$ of $w_i = y_i - \tilde{\theta}_1 x_i$ $(1 \le i \le 25)$. Since θ_0 is a location parameter and $\sigma > 0$ is a scale parameter, they have no role to play in determining $w_{(i)}$. The order statistics $w_{(i)}$ are plotted against $Q_i = t_{(i)}$ $(1 \le i \le 25)$, the latter obtained from (2.4.13). A pattern emerges by taking p = 5 (no other value of p results in an improvement). The corresponding MML and LS estimates of θ_1 are – 0.080 and – 0.079; their standard errors are 0.0087 and 0.0105, respectively. The MMLE $\hat{\theta}_1$ is clearly more precise.

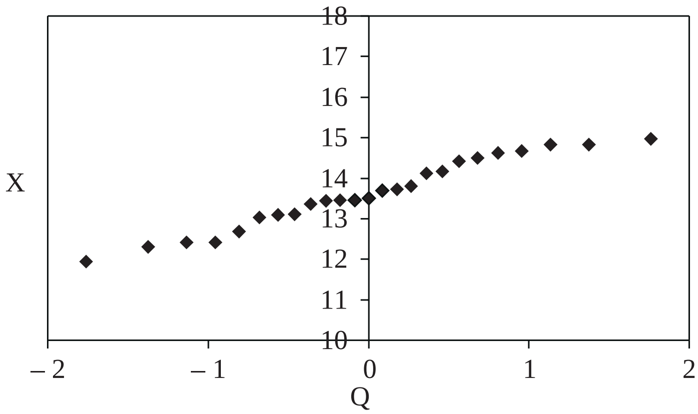

Fig. 4 Q-Q Plot of residuals based on LTS density.

Example 9.5: Kendall and Stuart (1968, p.407) give deviations from a simple 11-year moving average of marriage rate in England and Wales for the years 1843-1896. They are

– 6	1	12	10	– 6	– 8	– 6	3	4	7	11	3	– 8	– 2	–3	–7	3	4
– 5	–7	1	6	8	9	– 2	– 8	– 10	– 7	0	8	12	7	5	4	– 3	– 6
– 12	– 5	0	5	7	3	– 4	– 8	– 6	– 5	1	6	6	2	– 6	– 5	– 6	1

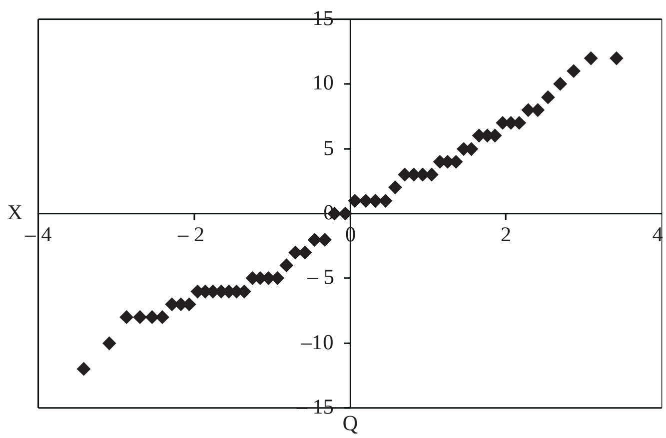

Fig. 5 Q-Q Plot based on STS density.

The deviations are presumed to have a short-tailed symmetric distribution. Given in Fig. 5 is the Q-Q plot obtained by plotting the ordered residuals against the quantiles $Q_i = t_{(i)}$ of the STS family (3.6.1) with r = 4 and d = 1.5; $t_{(i)}$ are obtained from (3.6.11). The plot has a pattern and, therefore, the density (3.6.1) with r = 4 and d = 1.5 is a plausible model.

None of the data in the examples above was found to be normal or in close proximity to it. This is not to say that normal data does not exist. Consider, for example, the following sample from Brownlee (1965, p.395). The data give the score of 14 subjects assigned randomly to a certain drug (Drug C):

35 31 55 43 44 28 33 13 39 58 18 17 41 25

The order statistics of this sample when plotted against the quantiles of a normal distribution N(0, 1) gives the following Q-Q plot:

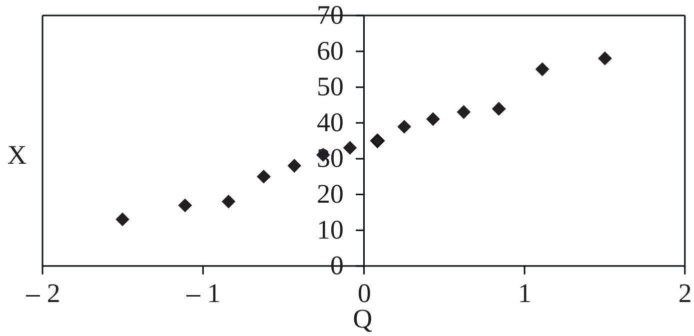

Fig. 6 Q-Q Plot based on normal density.

The data is beautifully modeled by a normal distribution $N(\mu, \sigma^2)$. Here, the estimates and their standard errors are

$$\bar{x} = 34.29 \quad \text{and} \quad s = 13.54 \text{ with } SE(\bar{x}) = \pm 3.62 \quad \text{and} \quad SE(s) \cong \pm 2.56.$$

The MMLE based on the sample, with r = [0.5 + 0.1(14)] = 1 smallest and largest observations censored, are

$$\hat{\mu}_c = 33.92 \text{ and } \hat{\sigma}_c = 12.83 \text{ with } SE(\hat{\mu}_c) = \pm 3.64 \text{ and } SE(\hat{\sigma}_c) = \pm 2.63.$$

The estimates and their standard errors are close to one another. This again illustrates that if a small number of extreme observations in a normal sample are censored (or small weights given to them) in the pursuit of robust estimators, that does not affect the efficiency too adversely, if the MMLE are used.

Remark: The difficulty with Q-Q plots is that they are very subjective. What might be a "straight line pattern" to one might not be so to another. The finding from a Q-Q plot, therefore, needs to be supported by a formal statistical test.

9.3 GOODNESS OF FIT TESTS

Realizing that Q-Q plots and other graphical techniques are highly subjective, we need formal tests to identify a plausible distribution for the data. We also need formal tests to identify outliers, inliers, and other data anomalies. As said earlier, estimation and hypothesis testing procedures which utilize the information obtained from Q-Q plots or goodness-of-fit tests are adaptive in the terminology of Hogg (1974, 1982), Hogg et al. (1984), and Sprott (1978, 1982).

Assume that the underlying distribution is of the type $(1/\sigma)f((x-\mu)/\sigma)$, μ and σ unknown. We have reason to believe that the functional form f is f_0 and that determines the null hypothesis

H_0: the underlying distribution is $(1/\sigma)f_0((x-\mu)/\sigma)$.

The alternative is

H_1: that f is not f_0 but some other density function.

Two situations are thought to be of primary interest: (a) some information about f is available, e.g., f is a symmetric density, and (b) no information about f is available. A procedure used to test H_0 against alternatives of type (a) is called a directional goodness-of-fit test. A procedure used to test H_0 against alternatives of type (b) is called an omnibus goodness-of-fit test. We first discuss some of the directional goodness-of-fit tests.

Remark: We recommend that a goodness-of-fit test be carried out at a significance level higher than 5 percent (say, 10 percent). The reason is that a robust procedure has almost the same efficiency when the density f is f_0 or close to it, but has higher efficiency when f is quite different from f_0. Thus, when H_0 is wrongly rejected with a high probability 0.10 (type I error) and subsequently a robust procedure adopted, that will result in only a slight loss in efficiency. On the other hand, a high value of type I error will yield high power (probability of correctly rejecting H_0) and that will lead to considerable gain in efficiency with the adoption of a robust procedure.

Directional test of Normality. The null hypothesis H_0 is that the underlying distribution is normal $N(\mu, \sigma^2)$, μ and σ both unknown. To test H_0 against skew alternatives, a well-known test is given by the sample skewness

$$\sqrt{b_1} = \sqrt{n}\,\textstyle\sum_{i=1}^{n}(x_i-\bar{x})^3 \,/\, \{\sum_{i=1}^{n}(x_i-\bar{x})^2\}^{3/2}; \tag{9.3.1}$$

$\sqrt{b_1}$ is location and scale invariant. Small and large values of $\sqrt{b_1}$ lead to the rejection of H_0 in favour of skew distributions. Large values indicate positive skewness and small values negative skewness.

The null distribution of $\sqrt{b_1}$ is symmetric but its functional form is not known. However, Pearson (1963) gives the exact variance and kurtosis of $\sqrt{b_1}$:

$$\mu_2(\sqrt{b_1}) = \frac{6(n-2)}{(n+1)(n+3)} \tag{9.3.2}$$

and

$$\beta_2(\sqrt{b_1}) = 3 + \frac{36(n-7)(n^2+2n-5)}{(n-2)(n+5)(n+7)(n+9)}. \tag{9.3.3}$$

D'Agostino and Pearson (1973) develop a Johnson S_u approximation to the null distribution of $\sqrt{b_1}$, namely, the distribution of

$$\delta \sinh^{-1}(\sqrt{b_1}/\lambda) \tag{9.3.4}$$

is referred to a normal distribution N(0, 1). They tabulate the values of δ and λ for $n \geq 8$.

An approximation in terms of the Student t is available (Tiku, 1966b). It regards

$$t = \sqrt{b_1}/h \tag{9.3.5}$$

as the Student t with ν degrees of freedom; ν and h are obtained by equating the second and fourth moments on both sides ($\beta_2 \geq 3$):

$$\nu = 4(\beta_2 - 1.5)/(\beta_2 - 3) \quad \text{and} \quad h = \sqrt{\mu_2(1-2/\nu)} \tag{9.3.6}$$

Thus, we have the approximation

$$P(\sqrt{b_1} \geq d_\alpha) \cong P(t \geq d_\alpha/h) \tag{9.3.7}$$

An IMSL subroutine in FORTRAN is available to evaluate (9.3.7) for both integer as well as noninteger values of ν of equation (9.3.6).

The values of the probability on the right hand side of (9.3.7) are given in Table 9.1. It can be seen that the t-approximation (9.3.5) is remarkably accurate.

Table 9.1: Values of the probability $P(\sqrt{b_1} \geq d_{0.95})$, $d_{0.95}$ being the exact 95 percent point of the null distribution.

n =	10	20	30	60	100
$d_{0.95}$	0.950	0.772	0.662	0.492	0.390
Prob	0.0502	0.0499	0.0498	0.0500	0.0500

To test normal $N(\mu, \sigma^2)$ against symmetric alternatives, a well-known test is given by the sample kurtosis

$$b_2 = n \sum_{i=1}^{n} (x_i - \bar{x})^4 / \{\sum_{i=1}^{n} (x_i - \bar{x})^2\}^2; \quad (9.3.8)$$

b_2 is location and scale invariant. Small and large values lead to the rejection of normal $N(\mu, \sigma^2)$.

It is very difficult to work out the null distribution of b_2. The distribution is, in fact, very skew and can not be approximated by a normal distribution even if n is as large as 1000. Pearson (1963) gives the first four moments of the null distribution (Thode, 2002):

$$\mu'_1(b_2) = \frac{3(n-1)}{(n+1)}$$

$$\mu_2(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2 (n+3)(n+5)} \quad (9.3.9)$$

$$\sqrt{\beta_1(b_2)} = \sqrt{\frac{216}{n}} \left[\frac{(n+3)(n+5)}{(n-3)(n-2)}\right]^{1/2} \frac{(n^2 - 5n + 2)}{(n+7)(n+9)}$$

and

$$\beta_2(b_2) = 3 + \frac{36(15n^6 - 36n^5 - 628n^4 + 982n^3 + 5777n^2 - 6402n + 900)}{n(n-3)(n-2)(n+7)(n+9)(n+11)(n+13)}.$$

D'Agostino and Pearson (1973) give charts to determine the percentage points for n = 20(5) 50, 100 and 200. They use Monte Carlo simulations and Pearson 4-moment curves (Johnson et al., 1963) to construct these charts. For other useful approximations, see D'Agostino (1970), D'Agostino and Tietjen (1971, 1973), Bowman and Shenton (1975, 1986) and Mulholland (1977). A recently published book (Thode, 2002) gives more detailed information about the null distributions of $\sqrt{b_1}$ and b_2 and gives tables of their percentage points. It might be noted, however, that D'Agostino and Tietjen (1973) approximation to the null distribution of $\sqrt{b_1}$ is exactly the same as (9.3.5) above.

A directional test of normality which is particularly powerful against symmetric short-tailed distributions is due to David et al. (1954), namely,

$$u = \sqrt{n}\ (w/S), \quad 0 < u < \infty; \quad (9.3.10)$$

$w = x_{(n)} - x_{(1)}$ is the sample range and $S = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}$. The statistic u is location and scale invariant. Small and large values of u lead to the rejection of normality.

It is known that under normality, u and S are independently distributed (David, 1981, p.111). Therefore,

$$E(u^k) = n^{k/2}E(W^k)/E(S^k) \quad (k = 1, 2,.....). \tag{9.3.11}$$

David, Hartley and Pearson (1954) give the percentage points of u, obtained by fitting Pearson curves to the first four moments of u.

The tests based on $\sqrt{b_1}$ and b_2 and u are, however, restricted to normal $N(\mu, \sigma^2)$. A directional test applicable to any location-scale distribution is due to Tiku (1974a, b; 1975b).

9.4 DIRECTIONAL TEST FOR ANY DISTRIBUTION

Under the null hypothesis H_0, let the distribution be $(1/\sigma)f_0((x-\mu)/\sigma)$. Directional test of H_0 is developed as follows:

The extreme observations in a random sample $x_1, x_2, ..., x_n$ primarily represent the tails of a distribution. They are, therefore, sensitive to changes in the tails. Censor r_1 smallest and r_2 largest observations. The remaining observations arranged in ascending order of magnitude constitute the censored sample (7.2.1). Since the MLE are generally intractable, let $\hat{\sigma}_c$ be the MMLE of σ (assuming that f_0 is the true density) so that $E(\hat{\sigma}_c) = \sigma$, at any rate for large n. Let $\hat{\sigma}$ be the estimator obtained by equating r_1 and r_2 to zero in the expression for $\hat{\sigma}_c$. The estimator $\hat{\sigma}$ is particularly sensitive to the tails but not $\hat{\sigma}_c$. Therefore, a directional goodness-of-fit statistic to test H_0 is given by (Tiku, 1974a)

$$T = \hat{\sigma}_c/\hat{\sigma}, \quad 0 < T < \infty; \tag{9.4.1}$$

T is location and scale invariant. To achieve high power, the following choices of r_1 and r_2 are made for testing H_0:

(i) against alternatives which have long tails on both sides, $r_1 = r_2 = [0.5+0.3n]$, (9.4.2)

(ii) against alternatives which have a long tail on the right hand side,

$$r_1 = 0 \quad \text{and} \quad r_2 = [0.5+0.6n].$$

If the distribution of X has a long tail on the left hand side, then the distribution of c-X (c being any constant) has a long tail on the right hand side. Therefore, we only need to consider the choices (i) and (ii).

Theorem 9.1: For fixed $q_1 = r_1/n$ and $q_2 = r_2/n$, the null distribution of T is asymptotically normal.

Proof: This immediately follows from the fact that $\hat{\sigma}$ converges to σ faster than $\hat{\sigma}_c$. Since the MMLE $\hat{\sigma}_c$ is asymptotically equivalent to the MLE, and the MLE are known to be asymptotically normal under some very general regularity conditions, the result follows.

In fact, the null distribution of T converges to normal very quickly. What we need is the mean and variance of T. For large n,

$$E(T) \cong E(\hat{\sigma}_c)/E(\hat{\sigma}) = 1 \quad \text{and} \quad V(T) \cong V(\hat{\sigma}_c) + V(\hat{\sigma}) - 2\,\mathrm{Cov}(\hat{\sigma}_c, \hat{\sigma}). \tag{9.4.3}$$

This follows from a very general result, namely, if $\hat{\theta}_1$ and $\hat{\theta}_2$ are both estimating the same parameter θ, then (Kendall and Stuart, 1969, p.234; Stuart and Ord, 1987, p.325),

$$V\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right) \cong \left\{\frac{E(\hat{\theta}_1)}{E(\hat{\theta}_2)}\right\}^2 \left[\frac{V(\hat{\theta}_1)}{\{E(\hat{\theta}_1)\}^2} + \frac{V(\hat{\theta}_2)}{\{E(\hat{\theta}_2)\}^2} - 2\frac{\mathrm{Cov}(\hat{\theta}_1, \hat{\theta}_2)}{E(\hat{\theta}_1)\,E(\hat{\theta}_2)}\right] \tag{9.4.4}$$

Since for large n, $\hat{\sigma}_c$ and $\hat{\sigma}$ are both unbiased estimators of σ, the results (9.4.3) follow.

Testing normal: Consider testing that f_0 is normal $N(\mu, \sigma^2)$. The MMLE is given in (7.3.11). Equating both r_1 and r_2 to zero, we get the traditional estimator $s = \sqrt{\{\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)\}}$. Now, $V(s) \cong \sigma^2/2n$, and $Cov(\hat{\sigma}_c, \hat{\sigma}) \approx \sigma^2/2n$ if $q_1 + q_2$ is small in which case

$$nV(t) \approx \left\{\frac{1}{2(1-q_1-q_2)-(q_2\alpha_2 t_2 - q_1\alpha_1 t_1)} - \frac{1}{2}\right\} \qquad (9.4.5)$$

from (7.3.16). For using T as a goodness-of-fit statistic, however, $q_1 + q_2$ is as large as 0.6. Based on part theory and part simulations, Tiku (1974a) gives the formula

$$nV(t) \approx \frac{n(A-1)}{A(n-1)}\left\{\frac{1}{2(1-q_1-q_2)-(q_2\alpha_2 t_2 - q_1\alpha_1 t_1)} - \frac{1}{2} - 0.532(q_1+q_2) + 1.1284(q_2\alpha_2 - q_1\alpha_1)\right\} \qquad (9.4.6)$$

the values of t_i, α_i and β_i (i = 1, 2) are given in (7.3.5) – (7.3.7).

The equation (9.4.6) gives remarkably accurate approximations as can be seen from the following values:

Table 9.2: Values of the variance of T: (a) $q_1 = 0$, $q_2 = [0.5+0.6n]/n$, and (b) $q_1 = q_2 = [0.5+0.3n]/n$.

		n = 20	24	30	40	50	60	80
(a)	Simul.	0.052	0.041	0.034	0.026	0.022	0.018	0.013
	Approx.	0.050	0.042	0.034	0.026	0.021	0.018	0.013
(b)	Simul.	0.057	0.046	0.040	0.030	0.025	0.020	0.015
	Approx.	0.058	0.047	0.040	0.030	0.025	0.020	0.015

The normal distribution with mean 1 and variance given in Table 9.2 gives remarkably accurate approximations for the percentage points of the null distribution of T. Small values of T lead to the rejection of normality. At 10 percent significance level, if the computed value of T is less than $1\text{-}1.645\sqrt{V}(T)$, $N(\mu, \sigma^2)$ is rejected as a plausible model for the data.

To compare the power of the $\sqrt{b_1}$ and T tests we consider, for illustration, the following alternatives to normal $N(\mu, \sigma^2)$.

Skew distributions: Chi-square with degrees of freedom (1) $\nu = 1$, (2) $\nu = 2$, (3) $\nu = 4$.

LTS distributions: The Student t with degrees of freedom (4) $\nu = 1$ (Cauchy), (5) $\nu = 2$, (6) $\nu = 4$. (9.4.7)

Short-tailed distributions: Tukey family $x = au^l - (1-u)^l$, u is uniform (0, 1), (7) a = 1, $l = 0.1$, (8) a = 1, $l = 1.5$, (9) a = 10, $l = 3.1$;

(7) and (8) are symmetric distributions with kurtosis μ_4/μ_2^2 equal to 3.21 and 1.75, respectively, and (7) has skewness $\mu_3/\mu_2^{3/2} = 0.97$ and kurtosis 2.8. The simulated values of the power are given in Table 9.3 based on [100,000/n] (integer value) Monte Carlo runs. Large values of

$\sqrt{b_1}$ lead to the rejection of normality in favour of positively skew distributions. Large values of $|\sqrt{b_1}|$ lead to the rejection of normality in favour of symmetric distributions. Small values of T lead to the rejection of normality; r_1 and r_2 are chosen as (i) and (ii) above. It can be seen that the T test is remarkably powerful. Both the tests are, however, ineffective against symmetric short-tailed distributions.

Table 9.3: Power of $\sqrt{b_1}$ and T tests for $N(\mu, \sigma^2)$, at 10% significance level.

	n = 10		n = 20		n = 50	
Alternative	$\sqrt{b_1}$	T	$\sqrt{b_1}$	T	$\sqrt{b_1}$	T
(1)	0.80	0.88	0.98	0.99	1.00	1.00
(2)	0.63	0.66	0.89	0.92	1.00	1.00
(3)	0.46	0.42	0.75	0.73	0.97	0.97
(4)	0.64	0.61	0.82	0.90	0.93	1.00
(5)	0.40	0.32	0.56	0.55	0.73	0.88
(6)	0.24	0.22	0.33	0.38	0.49	0.67
(7)	0.11	0.10	0.12	0.10	0.13	0.10
(8)	0.05	0.06	0.02	0.03	0.01	0.01
(9)	0.54	0.75	0.79	0.94	0.99	1.00

Example 9.6: The logarithm of the number of trees in orhards is presumed to be normal, at any rate close to it. If not, the distribution is presumed to be positively skew. For 10 randomly chosen orchards, Singh et al. (1982) give the following values of x = logarithm of the number of trees:

1.7918 2.3026 2.7726 3.2581 3.5264 3.8067 3.9703 4.0943 4.2905 4.5747

To use the statistic T, we choose $r_1 = 0$ and $r_2 = 6$. From the four smallest observations, we obtain

$$K = 2.869,\ A = 4,\ B = 1.847 \quad \text{and} \quad C = 2.165;$$

$$\hat{\sigma}_c = \frac{1.847 + (3.410 + 34.640)^{1/2}}{2\sqrt{12}} \quad \text{and} \quad \hat{\sigma} = \left[\frac{136.102 - (35.419)^2/10}{9}\right]^{1/2};$$

$$T = 1.063$$

Since this value is not smaller than the lower 10 percent point 0.476 of the null distribution of T, we have no reason to reject normality.

For the test of normality based on sample skewness,

$$\sqrt{b_1} = 0.158.$$

Since this value is not larger than the upper 10 percent point 0.950 of the null distribution of $\sqrt{b_1}$ (obtained from 9.3.5), the test agrees with the T test.

Example 9.7: Consider the Fisher data considered in Example 9.2. Here, the two smallest observations are too small and one largest observation is too large as compared to the bulk of observations. The alternative to normal $N(\mu, \sigma^2)$ is a distribution which, although not necessarily symmetric, has long tails on both sides. The directional T test is, therefore, appropriate with $r_1 = r_2 = [0.5+0.3(15)] = 5$. Now,

$$K = 23.116,\ A = 5,\ B = 50.726,\ C = 416.121;$$

$$\hat{\sigma}_c = \frac{50.726 + (2573.127 + 8322.428)^{1/2}}{2\sqrt{20}} = 17.342, \quad \hat{\sigma} = 37.744;$$

$$T = 0.459.$$

Since this value is smaller than the lower 10 percent point 0.514, we reject normal $N(\mu, \sigma^2)$ as a plausible model for the data.

In fact, the two smallest and one largest observations in this data are outliers (Section 9.16).

Comment: The T test is remarkably powerful for testing normal against skew and symmetric long-tailed distributions but it is ineffective against symmetric short-tailed distributions; see also Tiku (1974a, Table I). The u test is particularly powerful for testing normal against symmetric short-tailed alternatives. If short-tailed symmetric alternatives to normal are anticipated, the u test should be employed. Alternatively, the test based on the statistic U^* (Section 9.10) may be used; the test has remarkably high power.

The T and u tests beautifully complement one another. In practice, both may be computed and compared with the corresponding percentage points. This will give a better perspective of the nature of the underlying distribution.

Testing exponential: Under H_0, f_0 is the exponential $(1/\sigma)\exp\{-(x - \theta)/\sigma\}$, $\theta<x<\infty$; θ and σ not known. Here, the statistic T is

$$T_E = \hat{\sigma}_c/\hat{\sigma},\ 0 < T_E < \infty, \tag{9.4.8}$$

where
$$\hat{\sigma}_c = \frac{\sum_{i=1}^{n-r} x_{(i)} + r x_{(n-r)} - n x_{(1)}}{n - r - 1} \quad \text{and} \quad \hat{\sigma} = \frac{\sum_{i=1}^{n} x_{(i)} - n x_{(1)}}{n - 1} \tag{9.4.9}$$

Small and large values of T_E lead to the rejection of the exponential $E(\theta, \sigma)$.

Theorem 9.2: Under exponentiality the distribution of $y=\{(n - r - 1)/(n - 1)\}T_E$ is beta (a, b), namely,

$$f(y) = \frac{1}{\beta(a,b)}\, y^{a-1}(1 - y)^{b-1} \quad (0 < y < 1) \tag{9.4.10}$$

where $a = n - r - 1$ and $b = r$.

Proof: It is easy to show that

$$y = \sum_{i=1}^{n-r-1} D_i \Big/ \Big(\sum_{i=1}^{n-r-1} D_i + \sum_{i=n-r}^{n-1} D_i\Big), \quad D_i = (n - i)\{x_{(i+1)} - x_{(i)}\}.$$

Since $2D_i/\sigma$ $(1 \le i \le n - 1)$ are independently distributed as chi-squares, each with 2 degrees of freedom (Appendix 7A), the distribution of y is exactly the same as that of

$$w = \chi_1^2/(\chi_1^2 + \chi_2^2);$$

χ_1^2 and χ_2^2 are independent chi-square variates with $2(n - r - 1)$ and $2r$ degrees of freedom, respectively. It is easy to show that the distribution of y is beta (a, b).

Remark: For the beta(a, b),

$$E(y) = \frac{a}{a + b} \quad \text{and} \quad V(y) = \frac{ab}{(a + b)^2 (a + b + 1)}.$$

Hence, under exponentiality,

$$E(T_E) = 1 \quad \text{and} \quad V(T_E) = \frac{r}{n(n - r - 1)} \tag{9.4.11}$$

Choice of r: The T_E test has high power if r is large. Tiku et al. (1974) choose r = [0.5 + 0.5 n] since for this value of r the beta distribution (9.4.10) is also symmetric (almost). Consequently, the null distribution of

$$z = \sqrt{n(n-r-1)}(T_E - 1)/\sqrt{r} \tag{9.4.12}$$

tends to normal N(0, 1) very quickly (effectively, $n \geq 20$). For n = 20, for example, we have the following lower and upper 5 percent points: r = [0.5 + 0.5(20)] = 10:

	Lower	Upper
Exact	0.615	1.39
Approx.	0.612	1.39

Tiku (1974b), Dyer and Harbin (1981), Balakrishnan (1983) and Balakrishnan and Ambagaspitiya (1989) show that the T_E test (and its multisample version) has excellent power properties. Also, its power can be evaluated analytically as follows.

Lemma 9.1: The asymptotic power function of the T_E test is given by

$$P(z \leq d_1) + P(z \geq d_2) \tag{9.4.13}$$

where z is a standard normal variate and d_1 and d_2 are constants.

Proof: The power of the T_E test is given by

$$P(T_E \leq c_1 \mid H_1) + P(T_E \geq c_2 \mid H_1).$$

Now $\quad P(T_E \geq c_2) = P(\hat{\sigma}_c - c_2\hat{\sigma} \geq 0) = P(\sum_{i=1}^{n} l_{(i)} x_{(i)} \geq 0) = P(z \geq d_2)$

where $\sum_{i=1}^{n} l_{(i)} x_{(i)}$ is a linear function of order statistics, and

$$d_2 = -\sum_{i=1}^{n} l_i\ E(x_{(i)})/\sqrt{\{\sum_{i=1}^{n} l_i^2\ V(x_{(i)})\}}$$

is a constant; d_2 is free of σ. The result then follows from the fact that under some very general regularity conditions, a linear function of order statistics is asymptotically normal (David, 1981).

Testing E (0, σ)**:** In some situations, θ in the exponential E(θ, σ) is known (say, $\theta = 0$). Now

$$T_{E_0} = \hat{\sigma}_{c0} / \hat{\sigma}_0;\ \hat{\sigma}_{c0} = \{\sum_{i=1}^{n-r} x_{(i)} + r x_{(n-r)}\}/(n-r) \quad \text{and} \quad \hat{\sigma}_0 = \sum_{i=1}^{n} x_i/n; \tag{9.4.14}$$

r = [0.5+0.5n]. The null distribution of $\{(n-r)/n\}T_E$ is beta (a, b), a = n – r and b = r. For $n \geq 20$, the null distribution of

$$z_0 = \sqrt{(n-r)(n+1)}\ (T_{E_0} - 1)/(\sqrt{r})$$

is taken to be normal N(0, 1).

Example 9.8: Consider the data of Example 9.3. One wants to test whether $x = u^{2.5}$ is distributed as exponential E(0, σ), σ is not known. Here, r = [0.5 + 0.5(43)] = 22 and

$$z_0 = \sqrt{21(44)}\ (1.24 - 1)/\sqrt{22} = 1.56$$

Since the computed value of $|z_0|$ is less than 1.645, we do not reject the exponentiality of $x = u^{2.5}$. The Weibull (p = 2.5) is, therefore, a plausible model for the data.

Example 9.9: The following data represent the time to breakdown of a type of electrical insulating material subject to a constant-voltage stress (Nelson, 1970)

1.97	0.59	2.58	1.69	2.71	25.50	0.35	0.99	3.99	3.67
2.07	0.96	5.35	2.90	13.77					

The data clearly comes from a distribution with a long tail on the right hand side. Here, $r = [0.5+0.5(15)] = 8$. From the first $n - r = 7$ observations

0.35 0.59 0.96 0.99 1.69 1.97 2.07

we have

$$\sum_{i=1}^{n-r} x_{(i)} + r x_{(n-r)} = 25.18 \quad \text{and} \quad \sum_{i=1}^{n-r} x = 69.09,$$

$$T_{E_0} = \frac{25.18/7}{69.09/15} = 0.781.$$

The value of z_0 in (9.4.15) is

$$|z_0| = \left|\sqrt{7(16)}(0.781-1)/\sqrt{8}\right| = 0.82$$

which is less than 1.645. We do not reject the exponential at 10 percent significance level.

Numerous other directional tests of $E(\theta, \sigma)$ are available; see, for example, Epstein (1960a), Cox and Lewis (1966, Chap. 6), Wang and Chang (1977), Koul (1978) and Kanjo (1993). Many of the tests are designed to test the exponential $E(0, \sigma)$ against monotone increasing (IFR) or monotone decreasing (DFR) hazard functions (failure rates) $f(z)/[1 - F(z)]$, the failure rate being constant for the exponential. One of such tests is based on Gini statistic

$$G = \sum_{i=1}^{n-1} iD_i \Big/ (n-1)\sum_{i=1}^{n} D_i \quad (0 < G < 1); \tag{9.4.15}$$

$D_i = (n - i + 1)(x_{(i)} - x_{(i)-1})$ with $x_{(0)} = 0$ are the exponential spacings. Values of G close to zero are indicative of DFR and values close to 1 are indicative of IFR. Gail and Gastwirth (1978) show that the null distribution of

$$\sqrt{12(n-1)}\,(G - 0.5) \tag{9.4.16}$$

is approximately normal N(0, 1). They also show that the G test has good power properties against distributions with monotone hazard functions.

Testing uniform: To test the uniform $U(\theta_1, \theta_2)$ (θ_1 and θ_2 not known)

$$f(x) = 1/\theta,\ \theta_1 < x < \theta_2\ (\theta = \theta_2 - \theta_1), \tag{9.4.17}$$

against positively skew distributions, the directional T test is based on the statistic

$$T_u = \hat{\sigma}_c/\hat{\sigma},\ r = [0.5 + 0.5n] \tag{9.4.18}$$

$$\hat{\sigma}_c = (x_{(n-r)} - x_{(1)})/(n - r - 1) \quad \text{and} \quad \hat{\sigma} = (x_{(n)} - x_{(1)})/(n - 1).$$

Small and large values of T_u lead to the rejection of the uniform $U(\theta_1, \theta_2)$ in favour of positively skew distributions.

The null distribution of $y = \{(n - r - 1)/(n - 1)\}T_u$ is the beta distribution (9.4.10) with $a = n - r - 1$ and $b = r$ and converges to normal very quickly. This follows from the fact that $y = \sum_{i=1}^{n-r-1} u_i / \sum_{i=1}^{n-1} u_i$, where $u_i = (n + 1)\{x_{(i+1)} - x_{(i)}\}$ are the uniform spacings. It is well known that $u_i / \sum_{i=1}^{n-1} u_i$ are jointly distributed as $n - 1$ spacings generated by the order statistics of a random sample of size $n - 2$ from a Uniform (0, 1) distribution (Karlin, 1966). But then y is distributed as the $(n - r - 1)$th order statistic in a random sample of size $n - 2$ from the Uniform (0, 1). This distribution is the beta (a, b) with $a = n - r - 1$ and $b = r$.

To test the uniform $U(\theta_1, \theta_2)$ against symmetric alternatives, the directional T test is based on the statistic

$$T_u = \frac{(n-1)\{x_{(n-r)} - x_{(r+1)}\}}{(n-2r-1)\{x_{(n)} - x_{(1)}\}},\ r = [0.5 + 0.3\,n] \tag{9.4.19}$$

Small values of T_u lead to the rejection of $U(\theta_1, \theta_2)$ in favour of symmetric distributions. The null distribution of $y = \{(n-2r-1)/(n-1)\}T_u$ is the beta (a, b) distribution; $a = n - 2r - 1$ and $b = 2r$.

Testing uniform U(0, 1): Suppose that θ_1 and θ_2 in (9.4.17) are known, say $\theta_1 = 0$ and $\theta_2 = 1$. To test U(0, 1) against positively skew distributions, (9.4.18) reduces to

$$T_o = \{(n+1)/(n-r)\}\, x_{(n-r)},\ r = [0.5 + 0.5n]. \tag{9.4.20}$$

Small values of T_o lead to the rejection of U(0, 1). The null distribution of $x_{(n-r)}$ is the beta (a,b) distribution with $a = n - r$ and $b = r + 1$.

The T_o test is amazingly simple and powerful. For example, the values of the power of the T_o test are given below. Also given are the values of the Kalmogorov-Smirnov D test (introduced in the next section) which is known to be a very powerful test for the uniform U(0,1):

Power against $(1-u)^{k-1}$ $(0 < u < 1)$, at 10% significance level.

k	Skewness	Kurtosis		n = 10	20	40
1.5	0.12	2.05	T	0.30	0.46	0.67
			D	0.23	0.38	0.60
2	0.32	2.40	T	0.53	0.78	0.96
			D	0.53	0.78	0.97

For numerous real-life applications of the directional goodness-of-fit tests, see Chapter 11.

Multi-sample situation: Another interesting feature of the directional T test is that it admits a straightforward generalization to $k \ge 2$ independent samples. Consider, for example, k independent samples (Tiku, 1974a)

$$x_{i1}, x_{i2}, \ldots\ldots, x_{in_i} \quad (1 \le i \le k). \tag{9.4.21}$$

We want to test that these samples come from the populations $(1/\sigma_i) f_0((x-\mu_i)/\sigma_i)$, respectively. The functional form f_0 is specified but μ_i and σ_i are not known. The generalized statistic is

$$T^* = (1/k)\sum_{i=1}^{k} T_i \tag{9.4.22}$$

where T_i is the statistic calculated from the ith sample (9.4.21). Realize that T^* is location and scale invariant.

For large n_i $(1 \le i \le k)$, the null distribution of T^* is normal with mean 1 and variance $(1/k^2)\sum_{i=1}^{k} V(T_i)$.

9.5 OMNIBUS TESTS

We want to test the null hypothesis

$$H_0\text{: the underlying distribution is } f_0(x, \theta). \tag{9.5.1}$$

The function f_0 is specified and the parameters $\underset{\sim}{\theta}$ are known. Define

$$u = F(x) = \int_{-\infty}^{x} f_0(x, \theta)\, dx. \tag{9.5.2}$$

As said earlier, $u = F(x)$ is distributed as uniform Uniform (0, 1). Testing H_0 is, therefore, tantamount to testing that

$$u_{(i)} = F\{x_{(i)}\}, \quad 1 \le i \le n, \tag{9.5.3}$$

are the order statistics of a random sample of size n from U(0, 1). Realize that under H_0, $u_{(i)}$ is distributed as beta(i, n – i + 1). Consequently,

$$E\{u_{(i)}\} = \frac{i}{n+1} \quad (1 \le i \le n). \tag{9.5.4}$$

Therefore, a suitable function of the differences $u_{(i)} - i/n$ $(1 \le i \le n)$ is used to test H_0.

EDF Statistics: A very important omnibus goodness-of-fit statistic is that of Kolmogorov-Smirnov,

$$D = \max(D^+, D^-) \tag{9.5.5}$$

where $$D^+ = \max_{1\le i\le n}\left(\frac{i}{n} - u_{(i)}\right) \text{ and } D^- = \max_{1\le i\le n}\left(u_{(i)} - \frac{i-1}{n}\right). \tag{9.5.6}$$

Large values of D lead to the rejection of H_0; D^+ and D^{-1} can, in fact, be used as directional tests (Pearson and Hartley, 1972). The null distributions of D, D^+ and D^- are known for all n.

Other well-known omnibus goodness-of-fit statistics are the following:

Cramér-von Mises statistic $$W_1^2 = \sum_{i=1}^{n}\left(u_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n}, \tag{9.5.7}$$

Kuiper statistic $$V = D^+ + D^-, \tag{9.5.8}$$

Watson statistic $$U^2 = W_1^2 - n(\bar{u} - 1/2)^2, \quad \bar{u} = \sum_{i=1}^{n} u_i/n, \quad \text{and} \tag{9.5.9}$$

Anderson-Darling statistic

$$A = -\frac{1}{n}\left\{\sum_{i=1}^{n}(2i-1)[\ln u(i) + \ln(1 - u_{(n-i+1)})]\right\}. \tag{9.5.10}$$

The statistic A is generally more effective than other EDF statistics for detecting deviations in the tails (Pearson and Hartley, 1972, p.119). Large values of the statistics lead to the rejection of H_0. A simple table (Pearson and Hartley, 1972, Table 54) allows easy determination of the percentage points of the EDF statistics.

All the above EDF statistics are of limited use because of the requirement that the parameters θ be known. A natural thing to do, as suggested by Lilliefors (1967) and Srinivasan (1970), is to replace θ by a suitable estimate $\hat{\theta}$. Then, calculate the EDF statistics with u_i replaced by $\hat{u}_i$ $(1 \le i \le n)$:

$$\hat{u}_i = \int_{-\infty}^{x_i} f(x, \hat{\theta})\, dx. \tag{9.5.11}$$

The resulting statistics are called modified EDF statistics and denoted by $\hat{D}$, $\hat{W}_1^2$ and so on. However, the difficulty is that $\hat{u}_i$ $(1 \le i \le n)$ are no more uniformly distributed. Consequently, the distribution theory becomes complex: the null distributions may even depend on f_0. Based on part theory and part Monte Carlo simulations, Stephens (1974) constructed tables of the approximate percentage points of the modified EDF statistics for testing normal $N(\mu, \sigma^2)$ and exponential $E(0, \sigma)$. For $N(\mu, \sigma^2)$, the unknown parameters are replaced by the jointly sufficient statistics $\bar{x}$ and s^2. For the exponential $E(0, \sigma)$, σ is replaced by the sufficient statistic $\bar{x}$. The approximate percentage points of the modified EDF statistics are given in Pearson and Hartley (1972, Table 54), for testing $N(\mu, \sigma^2)$ and $E(0, \sigma)$.

For testing $N(\mu, \sigma^2)$, however, they are not on the whole as powerful as the Shapiro-Wilk W test given in the next section. For testing $E(0, \sigma)$, they are not as powerful as the Tiku Z test given in Section 9.8. See also Sürücü (2003) who gives a comprehensive power comparison of numerous tests.

9.6 SHAPIRO-WILK TEST

Let $z_{(i)} = (x_{(i)} - \mu)/\sigma$ $(1 \le i \le n)$ be the ith standardized ordered variate. Now,

$$E(x_{(i)}) = \mu + \sigma\mu_{i:n}, \quad \mu_{i:n} = E(z_{(i)}); \tag{9.6.1}$$

$\mu_{i:n}$ is the expected value of the ith standardized ordered variate which corresponds to the ith order statistic in a random sample of size n from $f_0(z)$, f_0 being the assumed density. The equation (9.6.1) is the same as (2.7.1) in Chapter 2. Let

$$\sigma^* = \sum_{i=1}^{n} a_{i:n}\, x_{(i)} \tag{9.6.2}$$

be the BLUE of σ. The Shapiro-Wilk test of $N(\mu, \sigma^2)$ is based on the statistic

$$W = \left(\sum_{i=1}^{n} a_{i:n}\, x_{(i)}\right)^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{9.6.3}$$

Small values of W lead to the rejection of normality (Shapiro and Wilk, 1965). Realize that W is location and scale invariant.

The coefficients $a_{i:n}$ are expressions in terms of the expected values and the variances and covariances of standardized variates $z_{(i)} = (x_{(i)} - \mu)/\sigma$. Their computation is involved. However, tables of the coefficients $a_{i:n}$ are available (Pearson and Hartley, 1972, Table 15); $a_{n-i+1} = -a_{i:n}$ because of symmetry.

The null distribution of W is intractable. Its percentage points have been obtained primarily by Monte Carlo simulations. Tables of the percentage points are available (Pearson and Hartley, 1972, Table 16).

Since the computation of the coefficients $a_{i:n}$ is formidable especially for large n, several modifications of W have been proposed (Mardia, 1980; Thode, 2002). For example, D'Agostino (1971) proposes the statistic

$$W^* = \sum_{i=1}^{n} \{i - (n+1)/2\} x_{(i)} \Big/ n^2 \left\{\sum_{i=1}^{n} (x_i - \bar{x})^2\right\}^{1/2}. \tag{9.6.4}$$

The numerator is essentially a simple linear estimator of σ due to Downton (1966). Small and large values of W* are indicative of non-normality (Locke and Spurrier, 1977).

The null distribution of W* is asymptotically normal with mean 0.282 and variance $(0.03)^2/n$, the convergence to normality being very slow. However, W is overall more powerful than W*. With tables of $a_{i:n}$ $(1 \le i \le n)$ and percentage points of W available for $n \le 50$, there is perhaps no incentive to use W* in preference to W unless n is greater than 50.

For numerous other less conventional tests of normality, see Mardia (1980) and Thode (2002).

Exponential E(θ, σ): The Shapiro-Wilk statistic for testing the exponential E(θ, σ) is

$$W_E = \left(1 - \frac{1}{n}\right)^{-1} (\bar{x} - x_{(1)})^2 \Big/ \sum_{i=1}^{n} (x_{(i)} - \bar{x})^2, \quad 0 < W < \infty. \tag{9.6.5}$$

Small and large values of W_E lead to the rejection of E(θ, σ). The null distribution of W_E is not known but Shapiro and Wilk (1972) provide simulated percentage points. The test, however, is not as powerful as some other tests (Table 9.4).

Shapiro-Wilk method can, in principle, be adopted for testing any distribution of the type $(1/\sigma)f((x - \mu)/\sigma)$. The computations are generally very involved and the distribution theory is complex. Moreover, other than for $N(\mu, \sigma^2)$, the tests based on Shapiro-Wilk statistics do not have power comparable to some other tests; see Section 9.8.

Transforms of spacings. Csörgö et al. (1973, 1975) use data transformations to give random variables with completely known distributions, on the same lines as Seshadri et al. (1969). For testing E(θ, σ), for example, they define the transforms

$$V_{(i)} = \sum_{j=1}^{i} D_j \Big/ \sum_{j=1}^{n-1} D_j \quad (1 \le i \le n-2) \qquad (9.6.6)$$

where D_j are the exponential spacings as in (9.8.1). Thus, testing that x_i $(1 \le i \le n)$ are from E(θ, σ) is tantamount to testing that $V_{(i)}$ $(1 \le i \le n-2)$ are jointly distributed as the order statistics of a random sample of size n – 2 from the Uniform (0,1). Any one of the EDF statistics D, V, W_1^2, U^2 and A can be used for that purpose. A convenient statistic to use is (Seshadri et al., 1969)

$$\chi^2 = -2 \sum_{j=1}^{n-2} \ln V_{(i)}. \qquad (9.6.7)$$

The null distribution of χ^2 is chi-square with 2(n – 2) degrees of freedom. The test is not, however, as powerful as some other tests (Table 9.4).

9.7 FILLIBEN AND SMITH-BAIN STATISTIC

Consider a location-scale distribution (1/σ)f((x – μ)/σ). The distribution of f(z), z = (x – μ)/σ, is free of μ and σ. To test H_0: f is f_0, we work out the population quantiles from the equation $F_0(Q_i) = i/(n+1)$, $1 \le i \le n$. The correlation statistic is (Filliben, 1975; Smith and Bain, 1976)

$$R = 1 - \hat{\rho}^2 \quad (0 < R < 1) \qquad (9.7.1)$$

where $\hat{\rho}$ is the ordinary product moment correlation coefficient between $x_{(i)}$ and $Q_i (1 \le i \le n)$. Values close to 1 lead to the rejection of H_0.

The computation of R is quite straightforward. Its null distribution is, however, not known for any f_0. Therefore, the percentage points of R have to be obtained by Monte Carlo simulation.

9.8 TIKU STATISTICS BASED ON SPACINGS

Consider the statistic T_E given in (9.4.8). For each r = 1, 2,..., n – 2, the estimator $\hat{\sigma}$ in the numerator is an unbiased estimator of σ with variance $\sigma^2/(n-r-1)$. A more powerful test of E(θ, σ) is obtained by combining these estimators (Tiku 1980b). A linear combination with weights inversely proportional to the variances, when divided by $\hat{\sigma}$ as in (9.4.8), yields the statistic

$$Z = 2\sum_{i=1}^{n-1} (n-1-i)D_i/(n-2)\sum_{i=1}^{n-1} D_i, \quad 0 < Z < \infty, \qquad (9.8.1)$$

$D_i = (n-i)\{x_{(i+1)} - x_{(i)}\}$ are the exponential spacings. Small and large values of Z lead to the rejection of E(θ, σ).

Theorem 9.3: Under exponentiality, the distribution of U = Z/2 is the same as the distribution of the mean of n – 2 iid (independently and identically distributed)

Uniform(0, 1) variates and is given by

$$f(u) = \frac{(n-2)^{n-2}}{(n-3)!} \quad \sum_{r=0}^{i} (-1)^r \binom{n-2}{r} \left(u - \frac{r}{n-2}\right)^{n-3} \qquad (9.8.2)$$

for $\frac{i}{n-2} \le u \le \frac{i+1}{n-2}$, i = 0, 1, 2, ..., n – 3.

Proof: It is easy to show that

$$\sum_{i=1}^{n-1}(n-1-i)D_i = \sum_{r=1}^{n-2}\sum_{i=1}^{n-r-1} D_i$$

and, therefore,

$$\frac{1}{2}(n-2)Z = \sum_{r=1}^{n-2}\sum_{i=1}^{n-r-1} u_i,\ u_i = D_i/\sum_{i=1}^{n-1} D_i.$$

Since D_i are iid exponential E(θ, σ), it is easy to show that u_i $(1 \le i \le n-1)$ are jointly distributed as n – 1 spacings generated by n – 2 order statistics of a random sample from Uniform (0,1); see, for example, Seshadri et al. (1969). But then $\sum_{i=1}^{n-r-1} u_i$ is the (n – r – 1)th order statistic in a random sample of size n – 2 from the Uniform (0,1). Thus, (n – 2)Z/2 is the sum of the n – 2 order statistics of a random sample of size n – 2. Since complete sums are invariant to ordering, i.e. $\sum_{i=1}^{n} x_{(i)} = \sum_{i=1}^{n} x_i$, the result follows (Kendall and Stuart 1969, p.258).

Since the mean and variance of a Uniform (0,1) variate is 1/2 and 1/12, respectively, it follows that

$$E(Z) = 1 \quad \text{and} \quad V(Z) = 1/3(n-2). \tag{9.8.3}$$

The cumulative distribution function of U=Z/2, obtained from (9.8.2), is

$$F(u_0) = P(U \le u_0) = \frac{(n-2)^{n-2}}{(n-2)!}\sum_{r=0}^{i}(-1)^r \binom{n-2}{r}\left(u_0 - \frac{r}{n-2}\right)^{n-2} \tag{9.8.4}$$

$$\text{for } \frac{i}{n-2} \le u_0 \le \frac{i+1}{n-2}, \quad i = 0, 1, 2, \ldots, n-3.$$

Remark: It is known that the distribution of the mean of iid uniform variates converges to normal very quickly with increasing n. The null distributon of

$$\sqrt{3(n-2)}(Z-1) \tag{9.8.5}$$

is, therefore, referred to normal N(0, 1) for all $n \ge 6$.

Known θ: In some situations, θ is known (say, θ = 0). To test E(0, σ), the statistic is

$$Z_0 = 2\sum_{i=1}^{n-1}(n-i)D_i/(n-1)\sum_{i=1}^{n} D_i \tag{9.8.6}$$

where $D_i = (n+1-i)\{x_{(i)} - x_{(i-1)}\}$ $(1 \le i \le n)$ are the exponential sample spacings; $x_{(0)} = 0$. Since $E(Z_0) = 1$ and $V(Z_0) = 1/3\,(n-1)$, the null distribution of

$$\sqrt{3(n-1)}\,(Z_0 - 1) \tag{9.8.7}$$

is referred to normal N(0,1) for all $n \ge 5$.

To have an idea about the accuracy of the normal approximation, we give below the exact and approximate upper 5 percent points of Z_0:

n =	5	7		11		21	
Exact	Approx.	Exact	Approx.	Exact	Approx.	Exact	Approx.
1.477	1.475	1.389	1.388	1.301	1.300	1.213	1.212

It can be seen that the normal approximation is remarkably accurate.

It is easy to show that the asymptotic power functions of the Z and Z_0 tests are exactly similar to that of T_E (Lemma 9.1).

For testing the exponential E(θ, σ), the power of the T_E and Z tests is determined by Tiku and Tamhankar (1980), Tiku (1980b), Dyer and Harbin (1981), Balakrishnan (1983) and Balakrishnan and Ambagaspitiya (1989) for numerous alternatives. They compare them with the corresponding values of the tests based on Shapiro-Wilk statistic W_E and Kalmogorov-

Smirnov and Cramér-von Mises EDF statistics (Seshadri et al., 1969; Srinivasan, 1970; Shapiro and Wilk, 1965, 1972; Durbin, 1975; Csörgö et al., 1975). What emerges from the comparisons is that Z is on the whole more powerful than other tests mentioned above. Consider, for example, testing $E(\theta, \sigma)$ against the ten alternatives considered in Csörgö et al. (1975), namely, the

Chi-square with degrees of freedom

$$(1)\ v = 1,\ (2)\ v = 3,\ (3)\ v = 4,\ (4)\ v = 6,\ (5)\ v = 8, \tag{9.8.8}$$

(6) log-normal ($\sigma = 1$), (7) half-normal, (8) Weibull (1/2),

(9) Weibull (1/2) and (10) beta (2,1).

The simulated values of the power are given in Table 9.4, reproduced from Tiku (1980b).

Table 9.4: Power of the tests at 10% significance level, n = 20.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Sum
Z	0.59	0.20	0.34	0.54	0.66	0.23	0.30	0.91	0.78	0.39	4.94
χ^2	0.65	0.21	0.36	0.55	0.65	0.16	0.28	0.95	0.71	0.23	4.79
W_E	0.39	0.20	0.33	0.52	0.64	0.27	0.34	0.83	0.79	0.43	4.74

It can be seen that the Z test is most powerful overall. See also Sürücü (2003), who has a comprehensive power study of the tests. His results are in agreement.

Example 9.10: Consider the data given in Example 9.8, n = 15. To test whether the data comes from the exponential $E(0, \sigma)$ we use the Z_0 statistic,

$$Z_0 = 2(372.22)/14(69.09) = 0.770.$$

The mean $E(Z_0) = 1$ and $V(Z_0) = 1/3(14) = 0.02381$. Since $|(0.770 - 1.0)|/\sqrt{0.02381} = 1.49$ is less than 1.645, we do not reject $E(0, \sigma)$ at 10 percent significance level.

9.9 EXTENSION TO NON-EXPONENTIAL DISTRIBUTIONS

The Z test can be extended to any location-scale distribution $(1/\sigma)f_0((x - \mu)/\sigma)$, μ and σ not known. Let

$$\mu_{i:n} = E\{z_{(i)}\},\ z_{(i)} = \{x_{(i)} - \mu\}/\sigma \quad (1 \le i \le n) \tag{9.9.1}$$

be the expected value of the ith standardized ordered variate. Tables of $\mu_{i:n}$ are available for numerous distributions, mostly for $n \le 20$. We will, however, use their approximate values $\mu_{i:n} \cong Q_i$ calculated from the equations

$$\int_{-\infty}^{Q_i} f_0(z)\,dz = \frac{i}{n+1}, \text{ i.e.,} \quad Q_i = F_0^{-1}\left(\frac{i}{n+1}\right), \quad 1 \le i \le n; \tag{9.9.2}$$

IMSL subroutines are available to calculate Q_i for numerous distributions. Using Q_i in place of $\mu_{i:n}$ has no substantial effect on the percentage points or the power of the tests. Thus, there is no need to sift through tables of $\mu_{i:n}$.

Generalized spacings: The generalized sample spacings are defined as

$$G_i = \{x_{(i+1)} - x_{(i)}\}/\{\mu_{1+i:n} - \mu_{i:n}\}, \quad 1 \le i \le n - 1. \tag{9.9.3}$$

Pyke (1965) has a very interesting asymptotic result: a set of sample spacings G_i are asymptotically distributed as iid exponential $E(0, \sigma)$. This is the motivation behind the following statistic Z*.

Normal N(μ, σ^2): To test H_0: f_0 is the normal density, the proposed statistic is (Tiku, 1980b)

$$Z^* = 2\sum_{i=1}^{n-1} (n-1-i)G_i/(n-2) \sum_{i=1}^{n-1} G_i \qquad (9.9.4)$$

where G_i are the spacings (9.9.3) with $\mu_{i:n}$ replaced by Q_i. Here,

$$(1/\sqrt{2\pi}) \int_{-\infty}^{Q_i} e^{-t^2/2}\, dt = \frac{i}{n+i}, \qquad 1 \le i \le n. \qquad (9.9.5)$$

Small and large values of Z^* lead to the rejection of H_0.

For large n, the percentage points (and the power) of Z* can be worked out in terms of a normal distribution (Lemma 9.1). Consider, for example, the determination of d_α such that

$$P(Z^* \ge d_\alpha \mid H_0) = \alpha. \qquad (9.9.6)$$

As in Lemma 9.1, the probability (9.9.6) is asymptotically equivalent to

$$P(Z_1 \ge -E(w)/\sqrt{V(w)} \mid H_0) = \alpha$$

where w is the difference between the numerator and d_α times the denominator in (9.9.4); Z_1 is a standard normal variate. The desired value of d_α is the positive root of the equation

$$z_\alpha{}^2 V(w) - \{E(w)\}^2 = 0 \qquad (9.9.7)$$

a quadratic in d_α; z_α is the 100(1 – α) percent point of a standard normal. The power of the Z* test can similarly be evaluated.

Using Pykes's result, however, it follows that the null distribution of Z* is for large n approximately normal. Interestingly, the normal approximation gives remarkably accurate results if the exact values of V(Z*) are used, E(Z*) ≅ 1. For large n, V(Z*) can be obtained from an equation similar to (9.4.3). We give in Table 9A.1 (Appendix 9A) the simulated values of V(Z*) which we use in a normal approximation; E(Z*) ≅ 1 for all n, e.g., the simulated values of E(Z*) are 1.002, 0.999 and 0.999 for n=7, 10 and 20, respectively. We also give the values of the probability

$$P\{\,|Z^* - 1| \ge 1.645\sqrt{V(Z^*)} \mid H_0\}. \qquad (9.9.8)$$

It can be seen that the normal approximation for the null distribution of Z* is remarkably accurate.

To have an idea about how powerful the Z* test is as compared to other tests, we consider the Shapiro-Wilk W test, the Filliben and Smith-Bain correlation R test and the EDF $\hat{A}$ test (the most powerful overall among all the EDF tests).

To represent skew and symmetric alternatives to normal, we consider the following densities; $\beta_1^* = \mu_3/\mu_2^{3/2}$ and $\beta_2^* = \mu_4/\mu_2^2$ are the coefficients of skewness and kurtosis:

(a) Long-tailed skew ($\beta_1^* \ne 0$, $\beta_2^* > 3$) : (1) Chi-square (ν = 1), (2) Chi-square (ν = 2), (3) Chi-square (ν = 4), (4) log-normal (σ = 1), (5) Weibull (1/2), (6) Weibull (2);

(b) Long-tailed symmetric ($\beta_1^* = 0$, $\beta_2^* > 3$) : (7) Cauchy, (8) Student t_2, (9) Student t_4, (10) Logistic, (11) $u^{0.1}(1-u)^{0.1}$, u is Uniform (0, 1);

(c) Short-tailed symmetric ($\beta_1^* = 0$, $\beta_2^* < 3$) : (12) Uniform (0, 1), (13) $u^{1.5}(1\text{-}u)^{1.5}$, u is Uniform (0, 1);

(d) Short-tailed skew ($\beta_1^* \ne 0$, $\beta_2^* < 3$) : (14) Beta (2, 1), (15) Johnson SB (0.533, 0.5).

(e) Dixon's location outlier model ($\beta_1^* \ne 0$, $\beta_2^* > 3$) : (n – r)N(μ, σ^2) and rN($\mu + c\sigma$, σ^2), (16) c = 2.0, (17) c = 3.0, (18) c = 4.0; r = [0.5+0.ln].

(f) Dixon's scale outlier model ($\beta_1^* = 0$, $\beta_2^* > 3$) : (n – r)N(μ, σ^2) and rN(μ, $(c\sigma)^2$), (19) c = 2.0, (20)c = 3.0, (21) c = 4.0; r = [0.5+0.ln] (integer value); alternatives of the type (f) have also been considered in Thode et al. (1983).

The values of the coefficients β_1* and β_2* for most of the distributions (a) – (f) are given in Tiku (1974 a, Table I).

The simulated values of the power are given in Table 9B.1 (Appendix), reproduced from Sürücü (2003); see also Tiku (1974a, Table I). What emerges from this power comparison is that no test is more powerful than others for all the alternatives (a) – (f) above. The Z^* test is overall most powerful for testing $N(\mu, \sigma^2)$ against skew alternatives. The R test is overall most powerful against long-tailed symmetric alternatives. The W test is, however, most powerful overall although it is ineffective against some, e.g., (11). Among the modified EDF tests, $\hat{A}$ is the most powerful overall but a little less powerful than W; see also Gan and Koehler (1990). The Z* test is ineffective against symmetric short-tailed alternatives. We present a U* test in the next section and show that the two, Z* and U*, beautifully complement one another and have a commonality, i.e., both are based on sample spacings and their null distributions are asymptotically normal (effectively, $n \geq 7$). In fact, we show that the Z* test has high power for testing skew distributions against both skew and symmetric alternatives, and for testing symmetric distributions against skew alternatives. The U* test has high power for testing symmetric dstributions against symmetric alternatives both short and long-tailed. Moreover, the null distributions of both Z* and U* are effectively normal. The only requirement for their application is the avaibility of their variances which can be obtained by simulation, if not from the equation (9.4.3). For testing normal, logistic, Student t_4 and extreme-value distributions, the variances of Z* are given in Table 9A.1; $E(Z^*) \cong 1$. Also given are the simulated values of the probability

$$P\{\,|Z^* - 1| \geq 1.645\sqrt{V(Z^*)}\;|\,H_0\}. \tag{9.9.9}$$

It can be seen that the normal approximations are remarkably accurate for $n \geq 7$, as said earlier.

Goodness-of-fit tests for an assumed distribution can also be based on the EDF statistics calculated from the generalized sample spacings G_i in (9.9.3); see Tiku (1980b, Section 5). Stephens (1986) has done that and he has some distributional results and power values. See also Tiku (1988, p.2383).

9.10 THE OMNIBUS U AND U* TESTS

To improve the statistic (9.4.19) for testing the Uniform (θ_1, θ_2), we realize that T_u is defined for every value of r = 1, 2,…, [n/2] (integer value). Combining these statistics as in Section 9.8, we obtain the improved statistic (Tiku, 1981a)

$$U = a \sum_{j=1}^{k-1} R_{n-2j-1}\,; \quad k = [n/2], \quad a = (n-k)/\{(k-1)(n-k-1)\}, \tag{9.10.1}$$

$$R_{n-2j-1} = \sum_{i=j+1}^{n-j-1} u_i \Big/ \sum_{i=1}^{n-1} u_i,$$

$u_i = (n+1)\{x_{(i+1)} - x_{(i)}\}$ $(1 \leq i \leq n-1)$ being the uniform sample spacings; [n/2] denotes the integer value of n/2. Large values of $|U|$ lead to the rejection of H_0: the underlying distribution is Uniform (θ_1, θ_2).

The statistic U is the ratio of two linear functions of order statistics. For large n, its mean and variance can be obtained from equations (9.4.3). Under H_0, Tiku (1981a) showed that for large n (effectively, $n \geq 10$)

$$E(U) \cong 1 \quad \text{and} \quad V(U) \cong 1/3(n-4), \tag{9.10.2}$$

and the null distribution of U converges to normal very quickly.

To compare the power of the U test with other tests, we first note that for the Uniform (θ_1, θ_2) the data transforms (9.6.6) are

$$V_{(i)} = \{x_{(i+1)} - x_{(i)}\}/\{x_{(n)} - x_{(1)}\}, \quad 1 \le i \le n-2. \tag{9.10.3}$$

The EDF statistics (9.5.7) – (9.5.10) and the chi-square statistic (9.6.7) are calculated from these transforms.

Given in Table 9.5 are the simulated values of the power of various tests for the Uniform (θ_1, θ_2) against a wide range of symmetric alternatives represented by the Tukey (1, k) family

$$u^k - (1-u)^{1-k}, \tag{9.10.4}$$

u being Uniform (0, 1). It can be seen that both U and R are more powerful than the EDF tests. The U test is not only more powerful than the R test but its null distribution is effectively normal and is, therefore, easy to implement. The percentage points of R have to be simulated.

Table 9.5: Values* of the power for testing Uniform (θ_1, θ_2), n = 20.

k =	1	0.25	0.1	– 0.25	– 0.5	– 1.0	– 2.0	Sum
U	0.10	0.42	0.59	0.88	0.95	0.99	1.00	4.93
R	0.10	0.25	0.44	0.82	0.93	0.96	1.00	4.50
D	0.10	0.23	0.40	0.75	0.87	0.97	1.00	4.32
W_1^2	0.10	0.22	0.39	0.76	0.89	0.97	1.00	4.33
A	0.10	0.21	0.38	0.74	0.88	0.97	1.00	4.28
χ^2	0.10	0.16	0.23	0.42	0.53	0.69	0.85	2.98

*W_1^2 is the Cramér-von Mises statistic (9.5.7).

To test any other symmetric distribution $(1/\sigma)f_0((x-\mu)/\sigma)$, μ and σ not known, against symmetric alternatives, the U* statistic is

$$U^* = a \sum_{j=1}^{k-1} R^*_{n-2j-1}, \tag{9.10.5}$$

where R^*_{n-2j-1} is the same as R_{n-2j-1} in (9.10.1) with u_i replaced by the generalized spacings (9.9.3),

$$u_i^* = (x_{(i+1)} - x_{(i)})/(Q_{i+1} - Q_i). \tag{9.10.6}$$

Consider testing a symmetric distribution $(1/\sigma)f_0((x-\mu)/\sigma)$ against symmetric alternatives. The simulated means and variances of U* are given in Table 9A.2. Also given are the values of the probability

$$P\{\,|\,U^* - E(U^*)\,| \ge 1.645\sqrt{V(U^*)}\ |\ H_0\}. \tag{9.10.7}$$

Since the convergence of E(U*) to 1 is rather slow, it is advisable to use its simulated value in (9.10.7). The normal approximation is amazingly accurate for all the three distributions considered: normal, logistic and Student t_4 (Tiku, 1981a; Sürücü , 2003).

The values of the power of the U* and R tests are given in Table 9.6 for numerous symmetric alternatives.

(1) Uniform, (2) Tukey (1, 0.5), (3) normal, (4) logistic, (5) Cauchy, (6) Student t_2, and (7) Student t_4.

It can be seen that the U* test is overall more powerful than the R test for testing symmetric distributions against symmetric alternatives, both short-and long-tailed. The modified EDF and Shapiro-Wilk statistics are not available for testing logistic and Student t_4 distributions due to the intractability of their percentage points and the formidable computations involved. Unlike the EDF statistics, the null distributions of the modified EDF statistics depend on the assumed density f_0 and that limits their usefulness. Moreover, they involve parameter estimation which can be cumbersome.

Table 9.6: Values of the power of U* and R tests, n = 20.

f_0		(1)	(2)	(3)	(4)	(5)	(6)	(7)	Sum
Normal	U*	0.43	0.22	0.10	0.17	0.90	0.56	0.27	2.65
	R	0.19	0.07	0.10	0.21	0.92	0.61	0.36	2.46
Logistic	U*	0.52	0.31	0.12	0.10	0.83	0.41	0.16	2.45
	R	0.20	0.05	0.04	0.10	0.86	0.52	0.22	1.99
Student t_4	U*	0.58	0.37	0.15	0.08	0.72	0.26	0.10	2.26
	R	0.14	0.03	0.01	0.03	0.72	0.33	0.10	1.36

9.11 EXTREME VALUE DISTRIBUTION

To illustrate further that for testing an assumed skew distribution the test based on Z* has high power, consider the situation when f_0 is the extreme-value distribution EV(δ, η) given in (2.6.1), δ and η unknown. Since EV(δ, η) is of the form $(1/\sigma)f_0((x-\mu)/\sigma)$, the Z* statistic is appropriate. The simulated values of its variance are given in Appendix 9A (Table 9A.1); E(Z*) ≅ 1. Also given are the values of the probability

$$P\{|Z^* - 1| \geq 1.645\sqrt{V(Z^*)} \mid H_0\}.$$

The normal approximation is remarkably accurate.

To test the EV(δ, η) distribution the modified EDF statistics can, of course, be used with δ and η replaced by their maximum likelihood estimates. The estimates are obtained as solutions of the equations

$$\bar{x} = \hat{\eta} + \sum_{i=1}^{n} x_i \exp(-x_i/\hat{\eta})/\sum_{i=1}^{n} \exp(-y_i/\hat{\eta}) \tag{9.11.1}$$

and

$$\hat{\delta} = -\hat{\eta}\ \ln\{\textstyle\sum_{i=1}^{n} \exp(-y_i/\hat{\eta})/n\}.$$

The first equation is solved iteratively and then $\hat{\delta}$ is obtained; the computations are, however, formidable (Thomas et al., 1969). Stephens (1977) gives the approximate percentage points of the EDF statistics.

Two other statistics have been used for testing EV(δ, η): the correlation statistic R and the Mann-Scheuer-Fertig statistic

$$S = \sum_{i=1}^{n-1} G_i/\sum_{i=1}^{k} G_i,\ k = \left[\frac{n}{2}\right] \text{ (integer value)}, \tag{9.11.2}$$

where G_i $(1 \leq i \leq n-1)$ are the sample spacings (9.9.3).

Littell et al. (1979) give the simulated values of the power of the tests based on the EDF statistics $\hat{D}$, $\hat{W}_1^2$ and $\hat{A}$ and the tests based on R and S for six distributions. We revise their list to the distributions given in Table 9.7, the first four being the same as in Littell et al (1979). They also consider log χ^2 (1) and log χ^2 (2) as in Tiku and Singh (1981, p. 913). We did not include these two distributions since the power of all the tests is only slightly bigger than the type I error. The updated values of the power are given in Table 9.7. It can be seen that the Z* test is overall at least as powerful as the $\hat{A}$ test and more powerful than the R and S tests.

The R and $\hat{A}$ tests are considerably more powerful than Z* for the Cauchy. On the other hand, the Z* test is considerably more powerful than $\hat{A}$ and R for the normal. See also Sürücü (2003).

Table 9.7: Values of the power, for 5% significance level.

n	$\hat{A}$	R	S	Z*	$\hat{A}$	R	S	Z*
		Normal				Logistic		
10	0.08	0.09	0.09	0.10	0.11	0.13	0.13	0.14
25	0.23	0.12	0.28	0.32	0.34	0.21	0.21	0.40
40	0.39	0.17	0.44	0.54	0.54	0.28	0.28	0.61
		Double exponential				Cauchy		
10	0.19	0.20	0.13	0.21	0.59	0.60	0.40	0.53
25	0.52	0.35	0.38	0.48	0.94	0.91	0.64	0.74
40	0.75	0.46	0.55	0.68	0.99	0.98	0.71	0.82
		Chi-square (ν = 1)					Chi-square (ν = 2)	
10	0.55	0.31	0.45	0.52	0.22	0.13	0.18	0.21
25	0.95	0.60	0.87	0.94	0.52	0.23	0.44	0.56
40	1.00	0.77	0.97	0.99	0.75	0.29	0.62	0.78
		Chi-square (ν = 4)					Half-Cauchy	
10	0.08	0.06	0.07	0.08	0.67	0.56	0.61	0.67
25	0.14	0.08	0.11	0.14	0.96	0.90	0.95	0.97
40	0.19	0.08	0.16	0.22	1.00	0.98	0.99	1.00

9.12 MULTISAMPLE SITUATIONS

The statistics Z* and U* above immediately generalize to multisample situations. To test that k independent samples

$$x_{i1}, x_{i2}, \ldots, x_{in_i} \tag{9.12.1}$$

come from the populations $(1/\sigma_i)f_0((x - \mu_i)/\sigma_i)$, i = 1, 2,…, k ($\mu_i$ and σ_i not known), the generalized Z* statistic is

$$Z^{**} = (1/k) \sum_{i=1}^{k} Z_i{}^* \tag{9.12.2}$$

where Z_i^* is the Z* statistic calculated from the ith sample. Small and large values of Z** lead to the conclusion that at least one of the samples (9.12.1) does not come from the assumed distribution f_0.

The null distribution of Z** is for large n_i $(1 \le i \le k)$ normal with mean 1 and variance

$$Z^{**} = (1/k^2) \sum_{i=1}^{k} V(Z_i{}^*). \tag{9.12.3}$$

The generalized U** statistic is similar to Z** and both tests have the same power properties as the single sample U* and Z* tests; see, for example, Dyer and Harbin (1981) and Balakrishnan (1983) who investigate the power properties of the multisample Z and T tests, the multisample Shapiro-Wilk test and the Kolmogorov-Smirnov test (when parameters are estimated), for testing the exponential distribution $E(\theta_i, \sigma_i)$, i = 1, 2, …, k. They find Z** considerably more powerful overall than these prominent competitors.

The multisample Shapiro-Wilk test, and similarly the Kolmogorov-Smirnov test, is executed by calculating the values of W_i from each sample and defining (Dyer and Harbin, 1981)

$$\chi^2 = -\sum_{i=1}^{k} \ln P_i \tag{9.12.4}$$

where $P_i = \text{Prob}\,(W \le W_i)$. The null distribution of χ^2 is chi-square with 2k degrees of freedom. The values of P_i, however, have to be obtained from the available tables of the percentage points by using tedious interpolation formulas and the computations are formidable (Dyer and Harbin, 1981).

While large values of $|Z^{**} - 1|$ (or any such multisample statistic) indicate that at least one sample does not come from the corresponding assumed distribution, values of $|Z^{**} - 1|$ which are not significantly large do not necessarily imply that all the samples come from the corresponding distributions. It is, therefore, always advisable to follow up a multisample test by testing each sample individually.

Example 9.11: The following data give the scores of 56 subjects assigned randomly to four drugs A, B, C and D (Brownlee, 1965, p.395),

A:	24	30	39	41	27	46	56	25	18	25	31	52	38	45
B:	18	32	33	35	19	28	41	37	33	39	38	41	38	36
C.	35	31	55	43	44	28	33	13	39	58	18	17	41	25
D:	28	40	34	27	47	30	24	28	21	28	39	26	46	42

Is it reasonable to assume that the samples come from normal populations ?

Here, the fourteen quantiles Q_i of the normal N(0, 1) are

– 1.498, – 1.112, – 0.842, – 0.622, – 0.432, – 0.253, – 0.083,
0.083, 0.253, 0.432, 0.622, 0.842, 1.112, 1.498

The values of the spacings

$$G_i = \{x_{(i+1)} - x_{(i)}\}/(Q_{i+1} - Q_i), \quad 1 \le i \le 13,$$

can now be calculated from each sample and the statistics Z_i (i = 1, 2, 3, 4) calculated. They are

0.901, 1.348, 0.962 and 0.855

respectively. Thus,

$$Z^{**} = \frac{1}{4}\,(4.066) = 1.0165.$$

The null distribution of Z^{**} is normal with mean 1 and variance 0.0234/4 = 0.00586 (Table 9A.1). Since $|1.0165 - 1|/\sqrt{0.00586} = 0.215$ is less than 1.645, we do not reject normality at 10 percent significance level.

Since the Z^{**} test is not much effective against symmetric short-tailed alternatives, we also perform the U^{**} test which is particulary effective against symmetric alternatives both short-and long-tailed:

$$U^{**} = (1/k) \sum_{i=1}^{k} U_i^{*}, \tag{9.12.5}$$

U_1^* is the statistic (9.10.4) calculated from the i^{th} sample. The values of U_i^* (i = 1, 2, 3, 4) are 1.117, 0.892, 1.003 and 1.150, respectively;

$$U^{**} = \frac{1}{4}(4.222) = 1.0555$$

with mean 0.963 and variance 0.00738. Since $|1.0555 - 0.963|/\sqrt{0.00738} = 0.65$ which is less than 1.645, we do not reject normality.

9.13 SAMPLE ENTROPY STATISTICS

In engineering and management sciences, goodness-of-fit tests based on the Kullback-Lieber information are often used. For testing the normal N(μ, σ^2), the statistic is (Vasicek, 1976)

$$K_{mn} = \frac{n}{2ms}\left\{\prod_{i=1}^{n}(x_{(i+m)} - x_{(i-m)})^{1/n}\right\} \tag{9.13.1}$$

where $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/n$ and m is a positive integer less than n/2; $x_{(i)}$ is the i^{th} order statistic and $x_{(i)} = x_{(1)}$ if i < 1 and $x_{(i)} = x_{(n)}$ if i > n. Small values of K_{mn} lead to the rejection of normality. It may be noted that K_{mn} is location and scale invariant and estimation of μ and σ is not, therefore, necessary.

The null distribution of K_{mn} is not known but its simulated percentage points are given in Vasicek (1976). Choosing m is problematic since for a given n, there is no single m which gives maximum power for every alternative. One then calculates the power of K_{mn} over numerous alternatives and uses that value of m which gives maximum power overall. This is done for each n. In fact, Vasicek (1976) gives the values of m so obtained for n = 10, 20 and 50. However, Şenoğlu and Sürücü (2004) have shown that the K_{mn} test is not overall as powerful as the Shapiro-Wilk W test: K_{mn} test is in general more powerful than the W test for short-tailed (μ_4/μ_2^2 less than 3) alternatives but less powerful for long-tailed alternatives. On the whole, however, the W test is more powerful. There is, therefore, no overwhelming reason to use the K_{mn} test in preferance to the W test.

Location and scale invariant K_{mn} statistics for testing the exponential E(θ, σ) and the uniform U(θ_1, θ_2) are also available (Şenoğlu and Sürücü, 2004). For testing the exponential E(θ, σ), for example,

$$K_{mn} = H_{mn} + \ln(\bar{x} - x_{(1)}) + 1 \tag{9.13.2}$$

where $$H_{mn} = \frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{n}{2m}(x_{(i+m)} - x_{(i-m)})\right). \tag{9.13.3}$$

The K_{mn} test is, however, considerably less powerful than the Tiku Z test(equation 9.8.1). The null distribution of K_{mn} is also unknown: it is asymptotically normal but the convergence to normality is very slow indeed (Zhoo and Zhang, 2001). Realizing also the difficulty in choosing m, there is no advantage in using the sample entropy statistics K_{mn}.

For testing the Uniform(θ_1, θ_2), Şenoğlu and Sürücü (2004) compare the power of K_{mn} with the correlation statistic R. They find R more powerful overall. Realize that the statistic U above provides a more powerful test than R (see Table 9.6). See also Sürücü (2003).

9.14 CENSORED SAMPLES

The Z and U (and Z* and U*) statistics immediately generalize to censored samples. Consider, for example, the censored sample

$$x_{(r_1+1)} \le x_{(r_1+2)} \le .... \le x_{(n-r_2)} \tag{9.14.1}$$

of size $m = n - r_1 - r_2$. To test that it comes from the exponential E(θ, σ), θ and σ not known, the Z statistic is

$$Z_c = 2\sum_{i=r_1+1}^{n-r_2-1}(n - r_2 - 1 - i)\,D_i/(m-2)\sum_{i=r_1+1}^{n-r_2-1}D_i \tag{9.14.2}$$

where D_i are the exponential sample spacings given in (9.8.1). The null distribution of (1/2) Z_c is the same as the distribution of m – 2 iid Uniform (0,1) variates. For large m(≥ 6), therefore, the null distribution of Z_c is normal with

$$E(Z_c) \cong 1 \text{ and } V(Z_c) \cong 1/3(m-2). \qquad (9.14.3)$$

Small and large values of Z_c lead to the rejection of H_0. The test is remarkably powerful (Tiku, 1980b, p.266). For n = 20, for example, we have the simulated values of the power of the tests based on Z_c and the correlation statistic $R = 1 - \hat{\rho}^2$ reproduced from Tiku (1980b), $\hat{\rho}$ being the product moment correlation coefficient between $Q_i = -\ln\{1 - i/(n+1)\}$ and $x_{(i)}$, $r_1 + 1 \le i \le n - r_2$. The percentage points of R are simulated.

Table 9.8: Values of the power for testing E(θ, σ), at 10% significance level.

	$r_1 = r_2 = 0$		$r_1 = 0, r_2 = 4$		$r_1 = 2, r_2 = 2$	
Alternative	Z	R	Z	R	Z	R
Chi-Square, v = 1	0.59	0.26	0.49	0.31	0.37	0.23
= 2	0.10	0.10	0.10	0.10	0.10	0.10
= 3	0.20	0.08	0.16	0.11	0.13	0.10
=4	0.34	0.10	0.29	0.16	0.17	0.10
Beta (2,1)	0.99	0.98	0.87	0.77	0.83	0.73
Weibull, c = 1/2	0.91	0.61	0.77	0.53	0.69	0.46
= 2	0.78	0.25	0.56	0.32	0.39	0.22
Lognormal (σ = 2.4)	0.99	0.87	0.86	0.71	0.89	0.72
Half-Cauchy	0.75	0.71	0.24	0.23	0.41	0.41
Half-normal	0.30	0.10	0.19	0.14	0.10	0.13
Cauchy	0.90	0.90	0.97	0.95	0.68	0.60
Normal	0.95	0.63	0.84	0.69	0.61	0.39

Tiku also evaluates the power of Z_c * for testing Gamma, Normal and Logistic (Tiku, 1980b) and of U_c * for testing Uniform, Normal, Logistic and Student t_4 distributions (Tiku, 1981a). He shows that, like Z* and U*, Z_c * and U_c * tests beautifully complement one another: Z_c * is powerful for testing skew distributions against both skew and symmetric alternatives, and for testing symmetric distributions against skew alternatives; U_c * is powerful for testing symmetric distributions against symmetric alternatives. The Z_c * and U_c * statistics have another very desirable feature: their null distributions are asymptotically normal and convergence to normality is very quick. The only requirements for their application are their means which are almost 1 and their variances which can be obtained by Monte Carlo simulation, if not from the formula (9.4.3).

Other goodness-of-fit tests (e.g., modified EDF, Shapiro-Wilk type, etc.) based on censored samples are also available; see Pettitt and Stephens (1976), Verrill and Johnson (1987), O'Reilly and Stephens (1988) and Thode (2002, Chapter 8). Their power as compared to the U and Z (and U* and Z*) statistics is essentially the same as for complete samples (Sürücü, 2003).

Sürücü (2002) developes goodness-of-fit tests for bivariate (and multivariate) distributions. He performs univariate tests on the marginal and the conditional distributions and combines the probabilities.

9.15 OUTLIER DETECTION

Observations in a sample which are too small or too large as compared to the bulk of observations are called outliers. Since their presence adversely affects the efficiency of most statistical procedures, it is very important to detect them and take remedial action. Two authoritative books (Barnett and Lewis, 1994; Hawkins, 1980) on the subject are available. Our aim here is limited to presenting outlier tests which have a commonality with goodness-of-fit tests, the two being closely related to one another. If outliers are found, they may be censored (if grossly anomalous) or small weights given to them for efficient statistical estimation and hypothesis testing (Chapter 11).

First, it is important to realize that whether an observation is an outlier or not is governed by the distribution one assumes. An observation in a sample might be an outlier if normal $N(\mu, \sigma^2)$ is assumed but might not be so if, for example, the Student t_4 is assumed. Since most practitioners of statistics are comfortable with using some simple distributions, e.g., normal, exponential and uniform, we will particularly focus on outlier detection with reference to these distributions. Our approach is, however, very general and can be used for any location-scale distribution $(1/\sigma)f((x - \mu)/\sigma)$; see, for example, Mann (1982).

Several models for outliers have been proposed; see Barnett and Lewis (1994) for a full discussion of these models. The most popular model perhaps is the one given by Dixon (1950) as follows.

For a single outlier to the right of a sample, the model stipulates that in a random sample of size n, n – 1 observations

$$x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n \text{ come from normal } N(\mu, \sigma^2)$$

and the i^{th} observation (i unknown)

$$x_i \text{ comes from } N(\mu + \lambda\sigma, \sigma^2),\ \lambda > 0; \qquad (9.15.1)$$

μ, σ and λ are not known. This is called a location-shift outlier model.

A scale-shift outlier model stipulates that

$$x_i \text{ comes from } N(\mu, \lambda\sigma^2),\ \lambda > 1. \qquad (9.15.2)$$

A location and scale shift outlier model stipulates that x_i comes from $N(\mu + \delta\sigma, \lambda\sigma)$, $\delta > 0$, $\lambda > 1$.

The models can, of course, be generalized to multiple outliers to the right or left of a sample and to non-normal distributions.

The models (9.15.1) – (9.15.2) have an important feature: they preserve the independence of x_i $(1 \le i \le n)$. However, they might not clearly separate an outlier from the bulk of the data (Tiku, 1977).

Another location-shift outlier model that has received attention is due to Tiku (1975c); see, for example, Mann (1982). For a single outlier to the right of the sample, this model stipulates that

$$x_{(i)}\ (1 \le i \le n - 1) \text{ are the first } n - 1 \text{ order statistics of a}$$

random sample of size n from f (9.15.3)

and $x_{(n)}$ is the largest order statistic of this sample plus $\delta\sigma$, $\delta > 0$;

f is any specified density of the type $(1/\sigma)f((x - \mu)/\sigma)$, μ and σ unkown. Barnett and Lewis (1994) call this model the label-slippage model. This model clearly separates an outlier from the bulk of observations in a sample. Since the observations are now ordered, they are no more independently distributed. This makes mathematics a little difficult but not insurmountable as seen in Chapters 7 and 8.

Both the models (9.15.1) and (9.15.3) can, of course, be readily generalized to multiple outliers to the right or left. Scale-shift outlier models can similarly be defined (Tiku, 1975c; 1977; Mann, 1982).

Remark: To evaluate the efficiencies of the classical estimator $\bar{x}$, the Huber M-estimator $\hat{\mu}_w$, the Tukey estimator $\hat{\mu}_T$ and the MMLE $\hat{\mu}_c$ (Chapter 7), we give below their variances under the outlier model similar to (9.15.3) which preserves symmetry, i.e., $c\sigma$ is substracted from $x_{(1)}$ and added to $x_{(n)}$; n = 20. Because of symmetry, all these estimators are unbiased. We also give the simulated means and variances of the estimators of σ; f is assumed to be a normal density; σ = 1 without loss of generality:

	Variance				Mean			
	$\bar{x}$	$\hat{\mu}_w$	$\hat{\mu}_T$	$\hat{\mu}_c$	s	$\hat{\sigma}_w$	$\hat{\sigma}_T$	$\hat{\sigma}_c$
0	0.050	0.106	0.053	0.052	0.99	0.98	0.91	0.98
1	0.074	0.109	0.053	0.052	1.22	1.05	0.91	0.98
2	0.107	0.114	0.051	0.051	1.48	1.02	0.91	0.98
3	0.155	0.106	0.053	0.052	1.76	0.95	0.91	0.98
4	0.206	0.106	0.051	0.051	2.06	0.89	0.91	0.98

	Variance			
c	s	$\hat{\sigma}_w$	$\hat{\sigma}_T$	$\hat{\sigma}_c$
0	0.026	0.060	0.031	0.036
1	0.036	0.082	0.031	0.036
2	0.052	0.120	0.032	0.037
3	0.069	0.121	0.032	0.037
4	0.091	0.094	0.032	0.037

The classical estimators $\bar{x}$ and s^2 are highly inefficient as expected. The lack of independence resulting from the outlier model (9.15.3) is detrimental to the M-estimators $\hat{\mu}_w$ and $\hat{\sigma}_w$ but not to $\hat{\mu}_T$, $\hat{\mu}_c$, $\hat{\sigma}_T$ and $\hat{\sigma}_c$. As an estimator of σ, $\hat{\sigma}_c$ has the least bias. For c = 0 (normal samples), $\bar{x}$ and s are most efficient as expected.

9.16 TESTING FOR OUTLIERS

For testing a single outlier to the right of a random sample $x_1, x_2, \ldots, x_n$ from $N(\mu, \sigma^2)$, the most popular statistic is that of Pearson and Chandra Sekar (1936), namely,

$$D = (x_{(n)} - \bar{x})/s \tag{9.16.1}$$

where $x_{(n)}$ is the largest order statistic and $\bar{x}$ and s^2 are the sample mean and variance, respectively. Large values of D lead to the conclusion that $x_{(n)}$ is, in fact, an outlier. For reasons given earlier, we recommend that a test for outliers be performed at 10 percent significance level.

Writing

$$D_1 = \sum_{i=1}^{n-1}(x_{(i)} - \bar{x}_{(n)})^2 \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2, \quad \bar{x}_{(n)} = \sum_{i=1}^{n-1} x_{(i)}/(n-1) \tag{9.16.2}$$

it is easy to show that

$$D_1 = 1 - nD^2/(n-1).$$

Thus, the statistics D and D_1 are equivalent.

For testing that there is a single outlier to the left of a sample, the statistic $(\bar{x} - x_{(1)})/s$ is used. Its null distribution is exactly the same as that of D. This follows from symmetry.

The generalization of D_1 for testing that r outliers exist to the right of the sample is due to Grubbs (1950, 1969) for r = 2 and Tietjen and Moore (1972) for general r with $r \le [n/2]$. Their statistic is

$$L_r = \sum_{i=1}^{n-r} (x_{(i)} - \bar{x}_{(n-r)})^2 / \sum_{i=1}^{n-1} (x_i - \bar{x})^2, \quad \bar{x}_{n-r} = \sum_{i=1}^{n-r} x_{(i)}/(n-r) \qquad (9.16.3)$$

Small values of L_r lead to the conclusion that the r largest observations $x_{(i)}$ $(n - r + 1 \le i \le n)$ are outliers.

Assuming normality, Grubbs (1950, 1969) gives the null distribution of L_1 and L_2. Tietjen and Moore (1972) give the simulated percentage points of the null distribution of L_r for r = 1(1)10.

For testing that r outliers exist to the left of the sample, the statistic is

$$L_r^* = \sum_{i=r+1}^{n} (x_{(i)} - \bar{x}_{n-r})^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2, \quad \bar{x}_{n-r} = \frac{1}{n-r} \sum_{r+1}^{n} x_{(i)}. \qquad (9.16.4)$$

The null distribution of L_r^* is exactly the same as that of L_r. This follows from symmetry. See also Tietjen (1986).

Example 9.10: The following data arranged in order represent the percent elongation at break of certain material (Grubbs, 1969):

2.02 2.22 3.04 3.23 3.59 3.73 3.94 4.05 4.11 4.13

A Q-Q plot based on the normal $N(\mu, \sigma^2)$ indicates that the two smallest observations might be outliers. To test this formally, we calculate L*:

$$n = 10,\ \bar{x}_8 = \sum_{i=3}^{10} x_{(i)}/8 = 3.728, \quad \sum_{i=3}^{10} (x_{(i)} - \bar{x}_{(8)})^2 = 1.197$$

$$\sum_{i=1}^{10} (x_{(i)} - \bar{x})^2 = 5.351.$$

Thus,

$$L_2^* = 1.197/5.351 = 0.224.$$

Under normality, the 10 percent point of the null distribution of L_2^* is 0.287 (Tietjen and Moore, 1972, Table I-d). Since the calculated value is smaller, we conclude that the two smallest observations are in fact outliers.

To verify whether the observations above other than the two smallest can be regarded as a censored sample from normal $N(\mu, \sigma^2)$, μ and σ unknown, we utilize the statistic Z_c^* obtained by replacing D_i in (9.14.2) by

$$G_i = (x_{(i+1)} - x_{(i)})/(Q_{(i+1)} - Q_{(i)});$$

Q_i $(1 \le i \le 10)$ are the quantiles of the standard normal N(0, 1).

The statistic

$$Z_c^* = 2 \sum_{i=3}^{8} (9 - i) G_i / 6 \sum_{i=3}^{9} G_i = 1.37$$

Now, for a complete sample of size n = 10 the upper 5 percent point of Z* is $1 + 1.645\sqrt{(0.034)} = 1.30$ (Table 9A.1). Since the variance of Z_c^* based on a censored sample of size n – r is larger than the variance of Z* based on a complete sample of size n (Tiku 1980b), the upper 5 percent point of Z_c^* is surely greater than 1.30. The calculated value of Z_c^* being so close to 1.30, the normal model for the eight largest observations is quite reasonable, at 10 percent significance level since the test is two sided.

Let us now compare the classical estimator $\bar{x}$ with the MMLE $\hat{\mu}_c$ (based on the censored sample consisting of the eight largest observations):

$$\bar{x} = 3.41 \text{ with standard error } \pm s/\sqrt{n} = \pm 0.24$$

$$\hat{\mu}_c = 3.64 \text{ with standard error } \pm \sqrt{\{V(\hat{\mu}_c)\}}_{\sigma=\hat{\sigma}_c} = \pm 0.15$$

(equation 7.3.16); this is also equal to $\hat{\sigma}_c/m = \pm 0.15$

since the covariance between $\hat{\mu}_c$ and $\hat{\sigma}_c$ is small.

The standard error of the MMLE is considerably smaller, as expected.

Another way to detect multiple outliers, say to the right, under Dixon's outlier model is to proceed sequentially. First, test whether there is a single outlier to the right of the sample by using, say, the Pearson-Chandra Sekar statistic (9.16.1). If not, we conclude that no outliers are present in the sample. Otherwise, we put aside $x_{(n)}$ and regard the remaining observations as a random sample of size n – 1 (under the Dixon outlier model, of course). We now test whether there is a single outlier to the right of the reduced sample, adjusting the percentage point to the new sample size n – 1. We continue the process till no outlier is found. However, this sequential procedure might turn out to be ineffective due to the masking effect, i.e., the phenomenon of some observations being close to one another than they are close to the bulk of observations (Pearson and Chandra Sekar, 1936). To illustrate this, consider the ten observations considered in Example 9.10. Let us proceed sequentially and, first, test whether there is one outlier to the left of the sample. Here,

$$D_1 = \sum_{i=2}^{10} (x_{(i)} - \bar{x}_9)^2 / \sum_{i=1}^{10} (x_i - \bar{x})^2$$
$$= 3.217/5.351 = 0.601.$$

The 10 percent point of D_1 is 0.490 (Tietjen and Moore, 1972, Table I-d). Since the computed value is not less than 0.490, we conclude that the smallest observation 2.02 is not an outlier. We have, however, established that under normal $N(\mu, \sigma^2)$, the two smallest observations are outliers. The sequential procedure does not even get started here due to the masking effect.

To detect up to k outliers (k known in advance) under Dixon's outlier model, Rosner (1975) proposes a different sequential procedure for k = 2: at the first stage, a suitable statistic for testing a single outlier is calculated. The largest (smallest) observation is censored whether or not the statistic yields a significant value. At the second stage, the statistic is calculated from the reduced sample of size n – 1. Using Monte Carlo simulations, Rosner finds the Pearson-Chandra Sekar type statistic $\max|x_i - \bar{x}|/s$ slightly the superior one among several statistics designed to detect up to two outliers. Extension of Rosner's procedure to detect up to k outliers is due to Rosner (1977) for k = 4 and Hawkins (1979) for general k. Prescott (1979) and Jain (1981) also study such sequential procedures.

9.17 TIKU APPROACH TO OUTLIER DETECTION

Let $x_1, x_2, \ldots, x_n$ be a random sample from a distribution of the type $(1/\sigma)f((x - \mu)/\sigma)$, the functional form f is specified but μ and σ are unkown. To test that r_1 smallest and r_2 largest observations are outliers, a slightly ammended version of the Tiku (1975c) statistic is ($r_1 \geq 0$, $r_2 \geq 0$)

$$T(r_1, r_2) = \hat{\sigma}_c/\hat{\sigma} \quad (r_1 + r_2 \geq 1); \tag{9.17.1}$$

$\hat{\sigma}_c$ is the MMLE calculated from the censored sample (Chapter 7)

$$x_{(r_1+1)} \le x_{(r_1+2)} \le \ldots \le x_{(n-r_2)}, \tag{9.17.2}$$

r_1 and r_2 are pre-determined, and $\hat{\sigma}$ is $\hat{\sigma}_c$ with $r_1 = r_2 = 0$. We require that both $\hat{\sigma}$ and $\hat{\sigma}_c$ be unbiased, at any rate for large n.

The presence of outliers affects $\hat{\sigma}$ but not $\hat{\sigma}_c$. Thus, $T(r_1, r_2)$ becomes a meaningful staistic for detecting outliers.

Normal distribution: For normal $N(\mu, \sigma^2)$, $\hat{\sigma}_c$ is given in (7.3.11) with $y_{(i)}$ replaced by $x_{(i)}$ $(r_1 + 1 \le i \le n - r_2)$, and $\hat{\sigma} = \sqrt{\{\sum_{i=1}^{n}(x_i - \bar{x}^2)/(n-1)\}}$. Small values of T indicate that the smallest r_1 and the largest r_2 observations are outliers.

It is difficult to work out the exact null distribution of $T(r_1, r_2)$ but Tiku (1975c) shows that the distribution is closely approximated by a variant of a Beta distribution. The 100γ percentage point of the null distribution of $T(r_1, r_2)$ is approximately

$$T_\gamma = \left[\frac{n-1}{n-r_1-r_2-1}u_\gamma + \frac{1}{5n}\left(1+\frac{1}{n-2r_2+1}\right)\right]\left(\frac{n(A-1)}{A(n-1)}\right)^{1/2}; \tag{9.17.3}$$

u_γ is the 100 γ percentage point of the beta distribution (9.4.10) with $a = n - r_1 - r_2 - 1$ and $b = r_1 + r_2$.

Exponential distribution: For the exponential distribution $E(\theta, \sigma)$, the statistic $T(r_1, r_2)$ is

$$T_E(r_1, r_2) = \frac{(n-1)\{\sum_{i=r_1+1}^{n-r_2} x_{(i)} + r_2 x_{(n-r_2)} - (n-r_1)x_{(r_1+1)}\}}{(n-r_1-r_2-1)\sum_{i=1}^{n}(x_{(i)} - x_{(1)})}. \tag{9.17.4}$$

Small and large values of T_E indicate outliers.

Theorem 9.4: Under exponentiality, the null distribution of

$$y = \{(n-r_1-r_2-1)/(n-1)\}T_E(r_1, r_2)$$

is the beta $(n - r_1 - r_2 - 1, r_1 + r_2)$.

Proof: Follows exactly along the same lines as Theorem 9.2.

Uniform distribution: For the Uniform (θ_1, θ_2), the statistic is

$$T_U(r_1, r_2) = (n-1)\{x_{(n-r_2)} - x_{(r_1+1)}\}/(n-r_1-r_2-1)(x_{(n)} - x_{(1)}). \tag{9.17.5}$$

Small and large values indicate that r_1 smallest and r_2 largest observations are outliers.

Theorem 9.5: Under uniformity, the distribution of

$$y = \{(n-r_1-r_2-1)/(n-1)\}\, T_U(r_1, r_2)$$

is beta $(n - r_1 - r_2 - 1, r_1 + r_2)$.

Proof: It is easy to show that

$$y = \sum_{i=r_1+1}^{n-r_2-1} u_i \Big/ \sum_{i=1}^{n-1} u_i, \quad u_i = (n+1)\{x_{(i+1)} - x_{(i)}\},$$

where u_i $(1 \le i \le n-1)$ are the uniform sample spacings. It is known that $u_i/\sum_{i=1}^{n-1} u_i$ $(1 \le i \le n-1)$ are jointly distributed as n-1 spacings generated by the order statistics of a random sample of size $n - 2$ from the Uniform (0, 1); see, for example, Karlin (1966). But then y is distributed as the $(n - r_1 - r_2 - 1)$th order statistic of a random sample of size $n - 2$ from the uniform (0, 1), and the result follows.

Example 9.13: The following are the 15 coded observations (arranged in increasing order) on the vertical semidiameters on Venus made in 1846 by Lt. Herdon (Grubbs, 1950):

– 1.40	– 0.44	– 0.30	– 0.24	– 0.22	– 0.13	– 0.05	0.06	0.10
0.18	0.20	0.39	0.48	0.63	1.01			

The data are supposed to come from a normal distribution but the two extreme observations are suspect. To test whether they are outliers, we use the statistic T(1, 1) given in (9.17.1):

$$n = 15,\ \alpha_1 = \alpha_2 = 0.637,\ \beta_1 = \beta_2 = 0.880,\ A = 13,\ \hat{\sigma} = 0.551,$$

$$\hat{\sigma}_c = \{0.686 + \sqrt{(0.470 + 90.630)}\}/2\sqrt{156} = 0.410;$$

$$T(1, 1) = 0.744.$$

The approximate lower 10 percent point is

$$\left[\frac{14}{12}(0.732) + \frac{1.0714}{75}\right]\left(\frac{180}{182}\right)^{1/2} = 0.865$$

Since the computed value is smaller than 10 percent point, we conclude that the observations –1.40 and 1.01 are, in fact, outliers. This agrees with the conclusion reached by Tietjen and Moore (1972) who use the L_2 statistic.

To compare the sample mean and the MMLE (based on the middle thirteen observations), we have

$$\bar{x} = 0.018 \text{ with standard error } \pm s/\sqrt{n} = \pm 0.142;$$

$$\hat{\mu}_c = (0.66 + 0.1655)/14.722 = 0.056 \text{ with standard error } \pm \hat{\sigma}_c/\sqrt{m} = \pm 0.107.$$

Clearly, the sample mean $\bar{x}$ ceases to be an efficient estimator if the sample contains outliers as said earlier.

To illustrate the usefulness of the statistic $T(r_1, r_2)$ for testing multiple outliers, some to the left and some to the right, we have the following example.

Example 9.14: Consider the Darwins data mentioned earlier in Example 9.2. Assuming normality of the data, one would like to test whether the two smallest and the two largest observations are outliers. Here, $r_1 = 2$, $r_2 = 2$ and n = 15 and

$$\hat{\sigma}_c = 25.498,\ \hat{\sigma} = 37.723$$

$$T(2, 2) = 25.498/37.723 = 0.676$$

The lower 10 percent point (approximate) of T(2, 2) is

$$\left[\frac{14}{10}(0.556) + \frac{1.0833}{75}\right]\left(\frac{150}{154}\right)^{1/2} = 0.79.$$

The computed value being less than the 10 percent point, we conclude that the two smallest and the two largest observations are outliers.

Example 9.15: The following ordered observations (representing time-to-failure of certain electronic components) are similar to the data given in Kambo and Awad (1985, Example 1):

0.020	0.100	0.142	0.156	0.157	0.158	0.160	0.292	0.315	0.633
0.784	0.820	0.903	0.959	1.001	1.016	1.206	1.329	2.988	3.098

Assuming exponential $E(\theta, \sigma)$, we want to test whether the two largest observations are outliers.

Here

$$n = 20,\ r_1 = 0,\ r_2 = 2,\ \{(n - r_1 - r_2 - 1)/(n - 1)\}T_E = 0.62.$$

The lower 10 percent point is 0.80. Since the computed value is smaller than the tabulated value, the indication is that the two largest observations are outliers.

9.18 POWER OF THE OUTLIER TESTS

To evaluate the performance of the outlier tests, several measures have been proposed; see, for example, Ferguson (1961) and Hawkins (1980, Chapter 2). The most useful one perhaps is the power. For a prespecified significance level, the power of an outlier test V is the probability

$$\text{Power} = P(V \le d \mid H_1). \tag{9.18.1}$$

Under H_1, the sample contains a specified number r of outliers. We presume that small values of V lead to the rejection of H_0 which leads to the conclusion that outliers are present in the sample. The probability

$$\gamma = P(V \le d \mid H_0) \tag{9.18.2}$$

is the type I error.

It is difficult to derive the power functions of the L_r and T (r_1, r_2) tests. Using Monte Carlo methods, Tiku (1977) determined the values of the power of L_r and T(0, r) for testing r outliers to the right of a normal sample. To do that, one generates n observations from a normal $N(\mu, \sigma^2)$. For r = 1 (single outlier to the right), one adds $\xi\sigma$ ($\xi > 0$) to an arbitrary x_i observation under Dixon's model and adds $\delta\sigma$ ($\delta > 0$) to the largest order statistic $x_{(n)}$ under Tiku's model. For r = 2, one adds $\xi_1\sigma$ ($\xi_1 > 0$) and $\xi_2\sigma$ ($\xi_2 > 0$) to two randomly chosen observations x_i and x_j under Dixon's model and adds $\delta_1\sigma$ ($\delta_1 > 0$) to $x_{(n-1)}$ and $\delta_2\sigma$ ($\delta_2 > 0$) to $x_{(n)}$ under Tiku's model. In particular, $\xi_1 = \xi_2 = \xi$ and $\delta_1 = \delta_2 = \delta$. The empirical power of the T(0, r) and L_r tests are given by a_1/N and a_2/N where N(= 10,000) is the number of samples generated and a_1 is the number of samples which yield a significantly small value of T(0, r) and, similarly, a_2/N. The values of the power are given in Table 9.9, reproduced from Tiku (1977).

Table 9.9: Values of power of the outlier tests L_r and T(0, r), significance level is 0.05; L = L_r and T = T(0, r).

		Tiku's model					Dixon's model			
		r = 1		r = 2			r = 1		r = 2	
n	$\delta\sigma$	T	L	T	L	$\xi\sigma$	T	L	T	L
10	0	0.05	0.05	0.05	0.05	0	0.05	0.05	0.05	0.05
	1	0.35	0.35	0.37	0.37	2	0.16	0.16	0.13	0.14
	2	0.88	0.85	085	0.82	4	0.66	0.67	0.59	0.61
	3	1.00	0.99	0.99	0.98	8	1.00	1.00	1.00	1.00
20	0	0.05	0.05	0.05	0.05	0	0.05	0.05	0.05	0.05
	1	0.50	0.49	0.67	0.61	3	0.39	0.42	0.44	0.50
	2	1.00	0.97	1.00	0.99	5	0.91	0.93	0.96	0.98
40	0	0.05	0.05	0.05	0.05	0	0.05	0.05	0.05	0.05
	1	0.68	0.56	0.89	0.74	3	0.39	0.42	0.51	0.57
	2	1.00	0.99	1.00	1.00	5	0.93	0.95	0.98	0.99

It can be seen that under Tiku's outlier model, T can be substantially more powerful than L. Under Dixon's model, T can be a little less powerful than L. This was to be expected since L is known to be most powerful under Dixon's model at any rate for r = 1 (David, 1981; Hawkins, 1977).

Tiku (1975c) shows that $T_E(r_1, r_2)$ and $T_U(r_1, r_2)$ also have high power. In an interesting paper, Mann (1982) studies T and variants of it for testing outliers in samples from a Weibull distribution. She works with Tiku's outlier model and shows that her tests have the desirable optimality properties. She tabulates percentage points of the tests and applies them to detect outliers in two sets of biomedical data.

Choice of r_1 and r_2: To use the statistic $T(r_1, r_2)$, and similar other statistics, one has to know in advance the number of possible outliers r_1 and r_2. In practice, one might have no idea about r_1 and r_2. To attain high power, it is very important that r_1 and r_2 be never underspecified. Tiku (1977) recommends the use of $T(r_1^*, r_2^*)$ with $r_1^* \geq r_1$ and $r_2^* \geq r_2$; r_1^* and r_2^* can be determined by a visual inspection of the data or from a Q-Q plot. We have the following examples.

Example 9.16: A set of eight mass spectrometer measurements on a particular isotope of uranium arranged in ascending order of magnitude are (Tietjen and Moore, 1972, Example 3)

199.31	199.53	200.19	200.82	201.92	201.95	202.18	245.57

The largest order statistic appears to be an outlier. From visual inspection, there is no need to take $r_2^* \geq 2$. We take $r_2^* = 2$, i.e., we censor the two largest observations.

The MMLE calculated from the remaining six observations is

$$\hat{\sigma}_c = \frac{1.6351 + 14.7197}{2\sqrt{30}} = 1.493 \quad \text{and} \quad \hat{\sigma} = 15.849;$$

$$T(0, 2) = 0.094$$

The calculated value is even smaller than the lower 1 percent point (equation 9.17.3)

$$\left[\frac{7}{5}(0.294) + \frac{1.2}{40}\right]\left(\frac{40}{42}\right)^{1/2} = 0.431.$$

We conclude that the two largest observations are outliers.

Let us compare the MMLE with the classical estimator $\bar{x}$:

$$\bar{x} = 206.434 \text{ with standard error } \pm \frac{s}{\sqrt{n}} = \pm \frac{15.849}{\sqrt{8}} = \pm 5.603.$$

Censoring only the largest observation 245.57, we obtain the MMLE (Chapter 7)

$$\hat{\mu}_c = 201.113 \text{ with standard error } \pm \frac{\hat{\sigma}_c}{\sqrt{m}} = \pm \frac{1.419}{\sqrt{7.828}} = \pm 0.507$$

Censoring the two largest observations, we obtain

$$\hat{\mu}_c = 201.197 \text{ with standard error } \pm \frac{1.493}{\sqrt{7.536}} = \pm 0.544.$$

The sample mean is highly inefficient as expected. The two MMLE and their standard errors are close to one another. The first standard error being a little smaller than the second indicates that the second largest observation 202.18 is perhaps not an outlier. Censoring this "good" observation does not, however, diminish the efficiency of the MMLE in any substantial way. On the other hand, if the outlier 245.57 is not censored (or its influence depleted some other way), the MMLE is the same as the sample mean and is highly inefficient. It is indeed important to take remedial action if the sample contains outliers.

Example 9.17: The following data, arranged in order of magnitude, represent the values of 31 contrasts in a 2^5 factorial expriment (Daniel, 1959) rescaled by Dempster and Rosner (1971):

– 3.143	– 2.666	– 1.305	– 0.8980	– 0.8138	– 0.7577	– 0.7437	– 0.4771
– 0.3087	– 0.2526	– 0.0982	– 0.0842	– 0.0561	0.0000	0.0281	0.1263
0.1684	0.1964	0.2245	0.2947	0.3929	0.4069	0.4209	0.4350
0.4630	0.5472	0.6595	0.7437	1.0800	2.147		

The data is assumed to come from normal $N(\mu, \sigma^2)$. From a visual inspection it is apparent that not more than three smallest and not more than two largest observations are outliers. To verify this formally, we use the statistic $T(r_1^*, r_2^*)$ with $r_1^* = 3$ and $r_2^* = 2$. Here,

$$\hat{\sigma}_c = \frac{2.806 + \sqrt{7.876 + 934.304}}{2\sqrt{650}} = 0.657 \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{30.001}{30}} = 1.0000;$$

$T(3, 2) = 0.657$

The lower 10 percent point of $T(3, 2)$ is

$$\left[\frac{30}{25}(0.736) + \frac{1.0357}{155}\right]\left(\frac{775}{780}\right)^{1/2} = 0.887$$

The calculated value being smaller than the percentage point, we conclude that there are multiple outliers in the data, three to the left and two to the right.

We now compare the sample mean with the MMLE as follows,

Classical: $\bar{x} = -0.132$ with standard error $\pm \dfrac{s}{\sqrt{n}} = 0.180$,

MMLE (with three smallest and two largest observations censored):

$$\hat{\mu}_c = -0.054 \text{ with standard error } \pm \frac{\hat{\sigma}_c}{\sqrt{m}} = \pm 0.118.$$

The sample mean is only 66 percent efficient as compared to the MMLE.

Since a symmetric censored sample has some analytical advantages, let us censor the third largest observation 0.7437 also. From the middle 25 observations, we calculate the MMLE

$$\hat{\mu}_c = -0.051 \text{ with standard error } \pm \frac{\hat{\sigma}_c}{\sqrt{m}} = \pm 0.122.$$

The MMLE above and their standard errors are essentially the same. Censoring a small number of good observations has very little adverse effect on the MMLE as long as the sample is cleared of all outliers.

Estimating r_1 and r_2: It has been suggested that a test of outliers be caried out in two stages. At the first stage, r_1 and r_2 are estimated from the data. At the second stage, a test of outliers is applied with $r_1 = \tilde{r}_1$ and $r_2 = \tilde{r}_2$. The following proposals have been made.

Tietjen and Moore (1972): Calculate the gaps to the right of the sample median, i.e., $L_i = x_{(i+1)} - x_{(i)}$ ($[n/2] + 1 \le i \le n - 1$). If L_j is the largest gap, then $\tilde{r}_2 = n - j$ and similarly for $\tilde{r}_1$.

Tiku (1975c): Calculate the sample spacings $G_i = \{x_{(i+1)} - x_{(i)}\}/(\mu_{i+1:n} - \mu_{i:n})$ to the right of the median. If G_j is the largest spacing, then $\tilde{r}_2 = n - j$, and similarly for $\tilde{r}_1$.

Jain and Pingel (1981): Calculate

$$RDM_i = G_i/\{(x_{(i)} - \text{median})/\mu_{i:n}\}$$

to the right of the sample median. If RDM_j is the largest RDM_i, then $\tilde{r}_2 = n - j$ and similarly $\tilde{r}_1$.

The distribution theory of such two stage tests is very complex. Using Monte Carlo simulations, Jain and Pingel (1981, 1982) find their proposal most satisfactory under normality. This is due to the fact that the RDM_i estimates are mostly larger than, if not equal to, the actual number of outliers in a sample. Other proposals, particularly that of Tietjen and Moore above, produce estimates which are often smaller than the actual number of outliers in a sample; any subsequent test can, therefore, be completely ineffective.

We recommend that any procedure to estimate r_1 and r_2 be supplemented by the computation of the MMLE $\hat{\mu}$ and its standard error and/or a visual inspection of G_i. That will provide a verification whether $\tilde{r}_1$ and $\tilde{r}_2$ are chosen such that there is virtually no possibility that outliers still exist in the censored sample (after censoring $\tilde{r}_1$ smallest and $\tilde{r}_2$ largest observations). Consider, for example, the data in Example 9.14. Here, the gaps L_i $(1 \le i \le 30)$ are

$$0.477\ 1.361\ 0.407, \ldots, 0.084, 0.336, 1.067$$

and give $\tilde{r}_1 = 2$ and $\tilde{r}_2 = 1$.

Under normality, the spacings G_i $(1 \le i \le 30)$ are ($\mu_{i:n}$ are given in Pearson and Hartley, 1972, Table 9)

$$1.123, 5.466, 2.204, 0.558, \ldots, 0.744, 0.455, 1.350, 2.511$$

and give $\tilde{r}_1 = 2$ and $\tilde{r}_2 = 1$.

Under normality, the values of RDM_i $(1 \le i \le 30)$ are

$$0.68, 5.67, 2.85, \ldots, 0.87, 2.61, 3.92$$

and give $\tilde{r}_1 = 2$ and $\tilde{r}_2 = 1$. The three methods are in agreement.

Assuming normality and $k \le 4$ (number of outliers in the sample), Rosner (1977) uses his one-at-a-time sequential procedure to detect outliers. He concludes that there are two outliers to the left and one outlier to the right, i.e., $\tilde{r}_1 = 2$ and $\tilde{r}_2 = 1$. See also Jain et al. (1982).

If, however, one examines the spacings G_i above carefully, realizing that under normality $E(G_i) = \sigma$ $(1 \le i \le 30)$, the most appropriate choice is $r_1 = 3$ and $r_2 = 2$. The corresponding statistic T(3, 2) is also significantly small. Let us compare the MMLE of μ:

with $r_1 = 2$ and $r_2 = 1$, $\hat{\mu}_c = -0.073$ with standard error ± 0.135, and

with $r_1 = 3$ and $r_2 = 2$, $\hat{\mu}_c = -0.054$ with standard error ± 0.118.

The spacings G_i and the standard errors indicate that $r_1 = 3$ and $r_2 = 2$ is the most appropriate choice. Alternatively, one might try to find a distribution which adequately models all the 31 observations (using Q-Q plots, for example), but we could not find one. The optimal strategy, therefore, is to censor the outliers and calculate the MMLE from the remaining observations (censored sample).

9.19 TESTING THE SAMPLE FOR OUTLIERS

In certain situations, one simply wants to find out if a sample contains outliers irrespective of how many (Ferguson, 1961). If a statistic employed for this purpose turns out to be significant, the sample is rejected. Assuming normality, the use of the sample skewness $\sqrt{b_1}$ has been recommended if outliers occur only on one side of the sample. The use of the sample kurtosis has been recommended if outliers occur on both sides of the sample (Ferguson, 1961). The null distributions of $\sqrt{b_1}$ and b_2 are given in Section 9.3. Tiku (1977) recommends the use of T(r_1*, r_2*) with r_1* and r_2* clearly greater than,if not equal to, the actual number of outliers in the sample. He simulates values of the power (from N=10,000 monte Carlo runs) and shows that T(0, r_2*) is more powerful than $\sqrt{b_1}$ even if r_2* is chosen to be twice (even three times) the actual number of outliers r_2 to the right, and similarly for T(r_1*, r_2*) as compared to b_2. We give values of the power in Table 9.10 under Dixon's location-shift model, reproduced from Tiku (1977). It may be noted that under Dixons location-shift model, $\sqrt{b_1}$ is known to be locally most powerful (Ferguson, 1961; Hawkins, 1977). From a practical point of view, however, a locally most powerful test is not of much value since the power is too low to be of any practical importance.

Example 9.18: The following ten observations represent the breaking strength of 0.104- inch hard-drawn copper wire (Grubbs, 1969, Example 1):

568 570 570 570 572 572 572 578 584 596

Does the sample contain outliers?

If at all, the outliers clearly occur to the right of the sample. For the T(0, r*) test, it suffices to take $r^* = 2$. Now,

$$\hat{\sigma}_c = 4.684, \quad \hat{\sigma} = 8.702; \quad T(0, 2) = 4.684/8.702 = 0.538.$$

Table 9.10: Values of the power of $\sqrt{b_1}$ and T(0, r*) tests for rejecting the sample, r being the actual number of outliers to the right.

		r = 1			r = 2			
			T				T	
N	$\xi\sigma$	$\sqrt{b_1}$	r* = 2	r* = 3	$\sqrt{b_1}$	r* = 3	r* = 4	r* = 6
10	0	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	1	0.06	0.06	0.05	0.05	0.06	0.06	0.06
	3	0.26	0.30	0.23	0.13	0.29	0.25	0.16
	5	0.75	0.76	0.62	0.39	0.77	0.66	0.37
20	0	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	1	0.05	0.06	0.06	0.06	0.06	0.06	0.06
	3	0.29	0.37	0.34	0.35	0.46	0.46	0.42
	5	0.83	0.90	0.85	0.92	0.96	0.96	0.92

The lower 10 percent point of the null distribution of T(0, 2) is

$$\left[\frac{9}{7}(0.594) + \frac{1.1429}{50}\right]\left(\frac{70}{72}\right)^{1/2} = 0.776.$$

Since the computed value is less than the percentage point, we conclude that the sample contains outliers, of course under the normality assumption.

For the test based on sample skewness, we have

$$\sqrt{b_1} = \sqrt{\frac{10(9702.336)}{26702.469}} = 1.149$$

Large values of $\sqrt{b_1}$ indicate outliers to the right of the sample irrespective of how many. The upper 10 percent point of $\sqrt{b_1}$ is 0.950. We conclude that the sample contains outliers.

To compare the sample mean $\bar{x}$ with the MMLE (with two largest observations censored) we have

$\bar{x}$ = 575.20 with standard error $\pm s/\sqrt{n}$ = ± 3.19.

MMLE $\hat{\mu}_c$ = 573.306 with standard error $\pm \hat{\sigma}_c/\sqrt{m}$ = ± 4.684/$\sqrt{9.5828}$ = ± 1.513.

The superiority of the MMLE is obvious.

In conclusion it must be stated that the statistic $T(r_1, r_2)$ is versatile: (i) it can be used for testing outliers in samples from any distribution of the type $(1/\sigma)f((x - \mu)/\sigma)$, (ii) it can be used for testing a specified number of outliers on either side of the sample, (iii) it can be used to test whether the sample contains outliers, irrespective of how many. Besides, variants of $T(r_1, r_2)$ can be used for distributions which are not necessarily location-scale type (Mann, 1982). The exact null distributions of the statistics are known for certain populations: for the uniform $U(\theta_1, \theta_2)$ and the exponential $E(\theta, \sigma)$, the null distributions of $y = \{(n - r_1 - r_2 - 1)/(n - 1)\}\, T(r_1, r_2)$ are exactly beta $(n - r_1 - r_2 - 1,\ r_1 + r_2)$, and is approximately beta for the normal $N(\mu, \sigma^2)$. See also Mann (1982).

The statistic readily generalizes to multisample situations as in Section 9.12. The statistic $T(r_1, r_2)$ also generalizes to multivariate populations and has good optimality properties; see Tiku and Singh (1981a), and Sürücü (2002). The statistic for testing outliers in a sample $(x_{i1}, x_{i2}, \ldots, x_{ik})$ $(1 \le i \le n)$ from a k-variate normal population is

$$G = \prod_{j=1}^{k} (\hat{\sigma}_{cj}/\hat{\sigma}_j) \tag{9.19.1}$$

where $\hat{\sigma}_{cj}/\hat{\sigma}_j$ is exactly the same as in (9.17.1) and is calculated from the marginal sample $(x_{1j}, x_{2j}, \ldots, x_{nj})$. The statistic G has excellent power properties (Sürücü, 2002).

Inliers: A natural statistic for testing inliers in a normal sample is

$$T = \hat{\sigma}_c/s \tag{9.19.2}$$

where $\hat{\sigma}_c$ is the MMLE given in (7.6.11) and s is the sample standard deviation. The distribution and the power properties of the test statistic have to be studied. Two inlier models have been proposed by Tiku et al. (2001) and Akkaya and Tiku (2003). Their relevance to real life data has also to be investigated. Since $\hat{\sigma}_c$ in (9.19.2) is based on an ordered sample with the middle observations censored, T should provide an efficient test statistic for detecting inliers. This is under investigation at the present time.

SUMMARY

In this chapter, we discuss goodness-of-fit tests and outlier detection procedures. We present some of the most prominent test statistics. We show that the Shapiro-Wilk statistic W is overall the most powerful for testing the normal $N(\mu, \sigma^2)$, followed closely by the Anderson-Darling statistic $\hat{A}$. We show that the Tiku statistics Z and Z* (and U and U*) based on sample spacings are overall very powerful for testing other distributions, e.g., exponential, logistic, Student t, extreme value and uniform. We show that the tests based on the modified EDF statistics do not have the power superiority over the tests based on the Shapiro-Wilk and Tiku statistics. Moreover, the modified EDF statistics are difficult to work with since they involve parameter estimation. We also discuss procedures for detecting outliers. Several real life examples are given to illustrate the usefulness of goodness-of-fit tests and outlier detection procedures. They play a pivotal role in choosing an appropriate robust procedure (Chapter 11).

APPENDIX 9A

Table 9A.1: Simulated* values of the variances of Z* and the probability $P\{|Z^* - 1| \geq 1.645\sqrt{V(Z^*)}\}$.

N	Normal n × Variance	Logistic n × Variance	Extreme-Value n × Variance
7	0.396	0.445	0.422
8	0.387	0.427	0.398
10	0.354	0.413	0.390
15	0.310	0.380	0.340
20	0.298	0.373	0.317
25	0.277	0.359	0.313
30	0.276	0.353	0.305
35	0.275	0.351	0.308
40	0.274	0.347	0.300
45	0.262	0.344	0.294
50	0.266	0.341	0.289
>50	0.77n/3(n – 2)	0.98n/3(n – 2)	0.83n/3(n – 2)

Probability

n = 7	10	15	20	25	30	35	40	45	50
Normal									
0.101	0.100	0.101	0.099	0.097	0.098	0.098	0.101	0.100	0.100
Logistic									
0.102	0.102	0.104	0.098	0.104	0.102	0.097	0.100	0.101	0.098
Extreme value									
0.104	0.095	0.095	0.100	0.101	0.096	0.096	0.095	0.098	0.098

* For intermediate values of n, the variances may be obtained by linear interpolation.

Table 9A.2: Simulated* values of the mean and variance of U* and the probability $P\{|U^* - E(U^*)| \geq 1.645\sqrt{V(U^*)}\}$.

	Normal		Logistic		Student t_4	
n	Mean	n × Variance	Mean	n × Variance	Mean	n × Variance
7	0.957	0.601	0.950	0.615	0.930	0.640
8	0.958	0.569	0.948	0.593	0.931	0.646
10	0.965	0.491	0.952	0.503	0.929	0.578
15	0.962	0.394	0.952	0.426	0.926	0.510
20	0.966	0.363	0.954	0.398	0.932	0.494
25	0.969	0.353	0.958	0.377	0.934	0.466

30	0.970	0.336	0.960	0.368	0.940	0.477
35	0.972	0.316	0.962	0.362	0.943	0.466
40	0.974	0.326	0.965	0.355	0.944	0.468
45	0.976	0.308	0.966	0.348	0.947	0.470
50	0.976	0.306	0.967	0.348	0.949	0.472
>50	0.975	0.84n/3(n – 4)	0.970	0.96n/3(n – 4)	0.950	1.303n/3(n – 4)

Probability

n = 7	10	15	20	25	30	35	40	45	50
Normal									
0.0.98	0.100	0.105	0.097	0.099	0.102	0.098	0.102	0.099	0.100
Logistic									
0.103	0.100	0.099	0.100	0.100	0.100	0.102	0.100	0.099	0.101
Student t_4									
0.098	0.097	0.100	0.100	0.100	0.100	0.097	0.094	0.098	0.098

* For intermediate values of n, the values may be obtained by linear interpolation.

APPENDIX 9B

Table 9B.1: Simulated values of the power for testing N(μ, σ^2), μ and σ unknown.

	n = 20				n = 50			
Alternative	Z*	W	R	$\widehat{A}$	Z*	W	R	$\widehat{A}$
Normal	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
(1)	0.99	0.99	0.98	1.00	1.00	1.00	1.00	1.00
(2)	0.90	0.90	0.85	0.86	1.00	1.00	1.00	1.00
(3)	0.68	0.65	0.58	0.59	0.98	0.98	0.95	0.94
(4)	0.96	0.96	0.94	0.94	1.00	1.00	1.00	1.00
(5)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
(6)	0.27	0.25	0.22	0.23	0.59	0.58	0.44	0.44
(7)	0.72	0.90	0.93	0.91	0.82	1.00	1.00	1.00
(8)	0.48	0.60	0.67	0.62	0.60	0.85	0.94	0.89
(9)	0.28	0.31	0.38	0.31	0.35	0.45	0.66	0.50
(10)	0.17	0.19	0.23	0.18	0.19	0.20	0.36	0.24
(11)	0.11	0.11	0.12	0.12	0.12	0.10	0.15	0.12
(12)	0.07	0.36	0.10	0.30	0.06	0.95	0.44	0.72
(13)	0.08	0.43	0.14	0.36	0.07	0.98	0.56	0.80
(14)	0.37	0.47	0.29	0.40	0.77	0.95	0.73	0.83
(15)	0.60	0.84	0.63	0.78	0.92	1.00	0.99	1.00
(16)	0.19	0.17	0.20	0.16	0.28	0.21	0.28	0.22
(17)	0.43	0.40	0.46	0.39	0.72	0.66	0.73	0.66
(18)	0.75	0.73	0.79	0.71	0.97	0.97	0.98	0.97
(19)	0.19	0.19	0.23	0.18	0.23	0.23	0.37	0.24
(20)	0.36	0.40	0.45	0.37	0.43	0.56	0.73	0.56
(21)	0.50	0.58	0.62	0.55	0.56	0.79	0.89	0.79

CHAPTER 10

Estimation in Sample Survey

10.1 INTRODUCTION

Sample survey is another important area from applications point of view. In most applications, the objective is to estimate the mean of a finite population. Let Π_N denote a finite population consisting of N measurements (elements) $Y_1, Y_2,\ldots, Y_N$. The mean

$$\overline{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i \tag{10.1.1}$$

is unknown. Given a random sample $y_1, y_2,\ldots, y_n$ chosen without replacement from Π_N, the objective is to estimate $\overline{Y}_N$ and to find the standard error of the estimate.

There are essentially two approaches in sample survey. One suggested approach is to consider the finite population Π_N as having been selected at random from a hypothetical population, called super-population. A sample $y_1, y_2,\ldots, y_n$ being a random sample from Π_N is also a random sample from the assumed super-population. A second approach does not assume a super-population. Instead, an empirical verification entails repeatedly sampling the same objective finite population Π_N. One may refer to Cochran (1946) and Godambe (1955) for a discussion of various aspects of these two approaches.

Irrespective of the two approaches, the sample mean

$$\bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$$

is used as an estimator of $\overline{Y}_N$. We show that in the context of sample survey as well, $\bar{y}_n$ has essentially the same properties as those in the classical framework (Chapters 2 and 8). That is, $\bar{y}_n$ is efficient only if the underlying population is normal or near-normal. For other populations, the MMLE developed in Chapter 2 are considerably more efficient than the sample mean.

10.2. SUPER-POPULATION MODEL

Consider the situation when the super-population has a location-scale distribution of the type $(1/\sigma)f((y-\mu)/\sigma)$ with mean μ and variance σ^2 both unknown. Let Π_N denote a finite population consisting of N elements $Y_1, Y_2,\ldots, Y_N$ chosen randomly from f. Let $y_1, y_2,\ldots, y_n$ be n elements chosen randomly without replacement from Π_N. In the framework of a super-population model, Y_i $(1 \le i \le N)$ and y_i $(1 \le i \le n)$ are both random samples from f. Writing

$$\overline{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i \quad \text{and} \quad \overline{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (10.2.1)$$

it is clear that $E(\overline{y}_n - \overline{Y}_N) = 0$ since $E(\overline{y}_n) = \mu$ and $E(\overline{Y}_N) = \mu$. We now show that the MSE (mean square error) is

$$E(\overline{y}_n - \overline{Y}_N)^2 = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right). \qquad (10.2.2)$$

Let $Y_1, Y_2, \ldots, Y_{N-n}$ denote the $N - n$ elements in Π_N other than $y_1, y_2, \ldots, y_n$. Writing

$$\overline{Y}_{N-n} = \frac{1}{N-n}\sum_{i=1}^{N-n} Y_i$$

it is easy to show that

$$\overline{Y}_N = \frac{n}{N}\overline{y}_n + \left(1 - \frac{n}{N}\right)\overline{Y}_{N-n}. \qquad (10.2.3)$$

Since $\overline{y}_n$ and $\overline{Y}_{N-n}$ are unconditionally independent of one another (Fuller, 1970),

$$E(\overline{y}_n - \overline{Y}_N)^2 = \left(1 - \frac{n}{N}\right)^2 E\{(\overline{y}_n - \mu) - (\overline{Y}_{N-n} - \mu)\}^2$$

$$= \left(1 - \frac{n}{N}\right)^2 \left(\frac{\sigma^2}{n} + \frac{\sigma^2}{N-n}\right) = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right);$$

$(1 - n/N)$ is called finite population correction. An unbiased estimator of σ^2 is the sample variance

$$s_n^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2/(n-1).$$

An estimate of $\overline{Y}_N$ is, therefore, $\overline{y}_n$ with standard error

$$\pm\sqrt{\frac{s_n^2}{n}\left(1 - \frac{n}{N}\right)}. \qquad (10.2.4)$$

The result we have established above is that under any super-population model $(1/\sigma)f((y - \mu)/\sigma)$, the sample mean $\overline{y}_n$ is an unbiased estimator of the mean of the finite population Π_N with standard error given in (10.2.4). Other than for normal or near-normal populations, however, $\overline{y}_n$ is an inefficient estimator. To establish that, we proceed as follows.

10.3 SYMMETRIC FAMILY

Suppose that the underlying super-population is one of the members of the family $f(y; p)$ given in (2.2.9). The MMLE $\hat{\mu}_n$ and $\hat{\sigma}_n$ given in (2.4.8) are calculated from the order statistics

$$y_{(1)} \le y_{(2)} \le \ldots. \le y_{(n)} \qquad (10.3.1)$$

of the random sample $y_1, y_2, \ldots, y_n$. Since $E(\hat{\mu}_n) = \mu$ (because of symmetry)

$$E(\hat{\mu}_n - \overline{Y}_n) = 0.$$

Proposition: As an estimator of $\overline{Y}_N$, the MMLE $\hat{\mu}_n$ is more efficient than $\overline{y}_n$ for all n. To establish this, we simply have to show that

$$E(\overline{y}_n - \overline{Y}_N)^2 \ge E(\hat{\mu}_n - \overline{Y}_N)^2. \qquad (10.3.2)$$

Realizing that $\hat{\mu}_n$ and $\overline{Y}_{N-n}$ are independent of one another, we obtain from (10.2.3),

$$E(\hat{\mu}_n - \overline{Y}_N)^2 = V(\hat{\mu}_n) + \frac{n}{N}\frac{\sigma^2}{n} - 2\frac{n}{N}\text{Cov}(\hat{\mu}_n, \overline{y}_n). \tag{10.3.3}$$

To show that $\hat{\mu}_n$ is more efficient than $\overline{y}_n$, in view of (10.2.2), we simply have to show that

$$\frac{\sigma^2}{n} \geq V(\hat{\mu}_n) + 2\frac{n}{N}\left\{\frac{\sigma^2}{n} - \text{Cov}(\hat{\mu}_n, \overline{y}_n)\right\}. \tag{10.3.4}$$

The expression on the right hand side will be called the bound $B\sigma^2$. Since $\hat{\mu}_n$ and $\overline{y}_n$ are both linear functions of order statistics, it is not difficult to find the exact values of the variance of $\hat{\mu}_n$ and the covariance of $\hat{\mu}_n$ and $\overline{y}_n$ (Chapter 2).

The values of 1/n and B are given in Table 10.1, for the sample fractions n/N = 0.1 and 0.2. In most surveys the sample fraction is less than or equal to 0.1 but, for a broader coverage, we include n/N = 0.2. It can be seen that $\hat{\mu}_n$ is considerably more efficient than the sample mean $\overline{y}_n$. For $p = \infty$, the super-population f(y; p) reduces to the normal $N(\mu, \sigma^2)$ in which case $\hat{\mu}_n$ reduces to the sample mean $\overline{y}_n$.

Although we have given values only for $n \leq 20$, but B is smaller than 1/n for all sample sizes n. For the symmetric family f(y; p), therefore, $\hat{\mu}_n$ is more efficient than the sample mean $\overline{y}_n$.

Table 10.1: Exact values of the bound B.

		n = 6		n = 10		n = 20	
p	n/N	1/n	B	1/n	B	1/n	B
2	0.1	0.1667	0.1112	0.1000	0.0648	0.0500	0.0314
	0.2		0.1254		0.0746		0.0365
3.5	0.1	0.1667	0.1547	0.1000	0.0909	0.0500	0.0448
	0.2		0.1569		0.0928		0.0460
5	0.1	0.1667	0.1619	0.1000	0.0962	0.0500	0.0477
	0.2		0.1627		0.0969		0.0482
∞	0.1	0.1667	0.1667	0.1000	0.1000	0.0500	0.0500
	0.2		0.1667		0.1000		0.0500

To have a precise idea about how efficient $\hat{\mu}_n$ is as compared to $\overline{y}_n$, we give below the exact values of the relative efficiency of $\overline{y}_n$,

$$RE = 100E(\hat{\mu}_n - \overline{Y}_N)^2/E(\overline{y}_n - \overline{Y}_N)^2. \tag{10.3.5}$$

The values are calculated from the exact values of $V(\hat{\mu}_n)$ and $\text{Cov}(\hat{\mu}_n, \overline{y}_n)$ given in Table 10A.1. Notice that the relative efficiency of $\overline{y}_n$ decreases as n increases.

Relative efficiency of the sample mean.

	p = 2		p = 3.5		p = 5		p = ∞	
n/N	n = 10	20	10	20	10	20	10	20
0.1	54.8	53.0	80.9	79.7	86.1	85.4	100	100
0.2	54.6	53.2	72.8	72.1	76.9	76.4	100	100

Mean square error: From the values given in Table 10A.1 (Appendix 10A), it is seen that $\mathrm{Cov}(\hat{\mu}_n, \bar{y}_n)$ is close to $V(\hat{\mu}_n)$. For $p = \infty$, in fact, the two are equal because (David, 1981)

$$(1/\sigma^2)\sum_{j=1}^{n} \mathrm{Cov}(y_{(i)}, y_{(j)}) = 1 \quad (1 \le i \le n).$$

Since for large n (equation 2.3.19)

$$V(\hat{\mu}_n) \cong (p+1)(p-3/2)\sigma^2/np(p-1/2) \quad (p \ge 2), \tag{10.3.6}$$

it follows that

$$E(\hat{\mu}_n - \bar{Y}_N)^2 \cong M\sigma^2$$
$$= \left[\frac{(p+1)(p-3/2)}{np(p-1/2)}\left(1-\frac{n}{N}\right)+\frac{3}{2}\,\frac{1}{np(p-1/2)}\,\frac{n}{N}\right]\sigma^2. \tag{10.3.7}$$

If the sample fraction n/N is very small, say $n/N \le 0.10$, the second term may be ignored in which case

$$E(\hat{\mu}_n - \bar{Y}_N)^2 \cong M_1\sigma^2$$
$$= \frac{(p+1)(p-3/2)}{np(p-1/2)}\left(1-\frac{n}{N}\right)\sigma^2. \tag{10.3.8}$$

We give in Table 10.2 the exact values of $(n/\sigma^2)\mathrm{MSE}(\hat{\mu}_n)$, and the values calculated from (10.3.7) and (10.3.8). It can be seen that (10.3.7) gives accurate approximations for all values of the sample fraction n/N. Even (10.3.8) gives reasonably accurate approximations for all $p > 2$.

A $100(1-\gamma)$ percent confidence interval for $\bar{Y}_N$ is

$$\hat{\mu}_n \pm t_{\gamma/2}(\nu)\sqrt{M\hat{\sigma}_n^{\,2}} \tag{10.3.9}$$

where $\hat{\sigma}_n^{\,2}$ is the MMLE of σ (equation 2.4.8) calculated from (10.3.1); $t_\gamma(\nu)$ is the $100(1-\gamma)\%$ point of the Student t distribution with $\nu = n - 1$ degrees of freedom. For $p > 2$, M may be replaced by M_1.

The confidence interval (10.3.9) is considerably shorter than the classical interval

$$\bar{y}_n \pm t_{\gamma/2}(\nu)(s_n/\sqrt{n}). \tag{10.3.10}$$

For $p = \infty$, (10.3.9) reduces to (10.3.10).

10.4 FINITE POPULATION MODEL

For a finite population Π_N the elements Y_i $(1 \le i \le N)$ are fixed. A sample consists of n elements y_i $(1 \le i \le n)$ chosen at random without replacement from Π_N. The MMLE $\hat{\mu}_n$ and $\hat{\sigma}_n$ of μ and σ, respectively, are calculated from the order statistics $y_{(i)}$ $(1 \le i \le n)$ of the sample and are exactly the same as those for the symmetric super-population model considered in Section 10.2.

Table 10.2: Exact and approximate values of the MSE.

p	n/N	Exact		Approximate	
		n = 10	n = 20	nM	nM_1
2	0.05	0.550	0.529	0.500	0.475
	0.10	0.548	0.530	0.500	0.450
	0.20	0.546	0.532	0.500	0.400
3.5	0.05	0.849	0.835	0.821	0.814
	0.10	0.809	0.797	0.785	0.771
	0.20	0.728	0.721	0.760	0.747
5	0.05	0.907	0.899	0.890	0.887
	0.10	0.861	0.854	0.847	0.840
	0.20	0.769	0.764	0.760	0.747

Denote by $Y_{(i)}$ $(1 \le i \le N)$ the order statistics of Π_N. The sampling distribution of $y_{(i)}$ is (Brownlee, 1965; Wilks, 1962)

$$p\{y_{(i)} = Y_{(t)}\} = \binom{t-1}{i-1}\binom{N-t}{n-i}\Big/\binom{N}{n} \qquad (10.4.1)$$

where t = i, i+1, …,N – n + i.

The joint sampling distribution of $y_{(i)}$ and $y_{(j)}$ is $(i \le j)$

$$p\{y_{(i)} = Y_{(t_1)}, y_{(j)} = Y_{(t_2)}\} = \binom{t_1-1}{i-1}\binom{t_2-t_1-1}{j-i-1}\binom{N-t_2}{n-j}\Big/\binom{N}{n} \qquad (10.4.2)$$

$$= P_{N,n,i,j}(t_1, t_2)$$

$t_1 = i, i+1, \ldots, N-n+i$ and $t_2 = j, j+1, \ldots, N-n+j$. Thus

$$E(y_{(i)}) = \sum_{t=i}^{N-n+i} Y_{(t)}\, p_{N,n,i}(t) \qquad (10.4.3)$$

$$E(y_{(i)}^2) = \sum_{t=i}^{N-n+i} Y_{(t)}^2\, p_{N,n,i}(t) \qquad (10.4.4)$$

and

$$E(y_{(i)}\, y_{(j)}) = \sum_{t_1=i}^{N-n+i} \sum_{t_2=j}^{N-n+j} Y_{(t_1)}, Y_{(t_2)}\, p_{N,n,i\,j}(t_1, t_2). \qquad (10.4.5)$$

The expressions (10.4.3) – (10.4.5) are computed from the order statistics $Y_{(i)}$ $(1 \le i \le N)$. As a check on these computations, if $Y_{(i)}$ $(1 \le i \le N)$ are replaced by the integers 1, 2,…, N, respectively, then

$$E(y_{(i)}) = i(N+1)/(n+1), \quad 1 \le i \le n, \qquad (10.4.6)$$

$$V(y_{(i)}) = i(N-n)(N+1)(n-i+1)/(n+1)^2(n+2) \qquad (10.4.7)$$

and

$$\text{Cov.}(y_{(i)}, y_{(j)}) = i(n-j+1)(N-n)(N+1)/(n+1)^2(n+2). \qquad (10.4.8)$$

A further check on these values is that

$$V(\bar{y}_n) = w'm_{i:n} = \bar{Y}_n \text{ and } V(\bar{y}_n) = w'Vw = (S^2/n)(1-n/N), \quad i < j, \qquad (10.4.9)$$

$$S^2 = \sum_{i=1}^{N} (Y_{(i)} - \bar{Y}_N)^2/(N-1);$$

$$\mu_{i:n} = E(y_{(i)}),\ \sigma_{ij:n} = \text{Cov}(y_{(i)}, y_{(j)}),\ V = (\sigma_{ij:n})\ (i, j = 1, 2, \ldots, n),$$

and

$$nw' = (1, 1, \ldots, 1).$$

Efficiency: Tiku and Vellaisamy (1997) evaluate the relative efficiency (10.3.5) of $\bar{y}_n$ under a finite population model. They generate N values from f(y; p) which constitute the finite population Π_N. From the order statistics $Y_{(i)}$ $(1 \le i \le N)$ and equations (10.4.3) – (10.4.5), they calculate the exact values of $V(\hat{\mu}_n)$ and Cov $(\hat{\mu}_n, \bar{y}_n)$ and use them in (10.3.4) to obtain the following results:

Relative efficiency of the sample mean, n = 20.

	n/N = 0.1					n/N = 0.2			
p =	1.5	2.5	5	∞	p =	1.5	2.5	5	∞
	27.8	81.3	95.2	100		48.1	92.6	97.1	100

It can be seen that the sample mean $\bar{y}_n$ is inefficient under a finite population model as well. We do not give details but the relative efficiencies decrease as n increases other than for $p = \infty$.

Comment: The MMLE are robust to plausible deviations from an assumed value of p in f(y; p). To illustrate this, consider the following ten values:

8.4490	9.0782	9.4161	9.6709	9.8933	10.1067
10.3291	10.5839	10.9218	11.5551		

The values are, in fact, the expected values of the order statistics of a random sample of size n = 10 from the distribution f(y; p) with $\mu = 10$, $\sigma = 1$ and p = 3.5. Here,

$$\hat{\mu}_n = 10 \quad \text{and} \quad \hat{\sigma}_n = 1, \quad \text{and} \quad \bar{y}_n = 10 \quad \text{and} \quad s_n = 0.91.$$

The classical estimator of σ has a downward bias.

We now work with different values of p and have the following values.

For p = 3,

$$\hat{\mu}_n = 10 \quad \text{and} \quad \hat{\sigma}_n = 1.03$$

For p = 4,

$$\hat{\mu}_n = 10 \quad \text{and} \quad \hat{\sigma}_n = 0.98$$

It is seen that the MMLE are remarkably robust numerically, as expected. This is due to the fact that the coefficients β_i $(1 \le i \le n)$ we use to calculate the estimates have umbrella ordering (Chapter 8).

10.5 SAMPLE SIZE DETERMINATION

Often in practice, one wants to pre-determine the sample size n such that the mean square error of a subsequent estimator μ does not exceed a given limit D. The estimators $\bar{y}_n$ and $\hat{\mu}_n$ are typically suited to solve such sample-size determination problems as follows.

Equating the MSE of $\bar{y}_n$ to D, we get

$$n = \sigma^2/(D + \sigma^2/N) \tag{10.5.1}$$

n is chosen to be the integer just greater than the value on the right hand side.

Equating the mean square error of $\hat{\mu}_n$ to D, we obtain $(p \ge 2)$

$$n = \frac{(p+1)(p-3/2)}{p(p-1/2)}\left\{\sigma^2 \Big/ \left(D + \frac{\sigma^2}{N}\,\frac{2p^2 - p - 6}{2p(p-1/2)}\right)\right\}; \tag{10.5.2}$$

n is the integer just greater than the right hand side. The bound on the right hand side reduces to that in (10.5.1) if $p = \infty$ (normal distribution). It is considerably smaller for $2 \le p < \infty$.

Ignoring terms of order O (N^{-1}), the ratio of (10.5.2) to (10.5.1) is

$$(p + 1)(p - 3/2)/p(p - 1/2) \tag{10.5.3}$$

which is less than 1 for $2 \le p < \infty$. For all finite values of p, therefore, a smaller sample size n is needed to attain a pre-determined mean square error if the MMLE $\hat{\mu}_n$ is used.

In practice, σ in (10.5.1) and (10.5.2) is replaced by an external estimate, e.g., the one obtained from a pilot survey.

10.6 STRATIFIED SAMPLING

Suppose that a population is divided into L non-overlapping strata and there are N_h elements in the h th strata. Let $(1 \le h \le L)$

$$y_{h1}, y_{h2}, \ldots, y_{hn_h} \tag{10.6.1}$$

be a simple random sample of size n_h from the h th strata. The order statistics of (10.6.1) are used to calculate the MMLE $\hat{\mu}_h$ and $\hat{\sigma}_h$ $(1 \le h \le L)$. The estimators depend on the shape parameter p in the family (2.2.9). Let p_h be the value of p in the h th strata. In particular, p_h could be all equal.

The MMLE of the population mean

$$\overline{Y}_N = \frac{1}{N}\sum_{h=1}^{L} N_h \overline{Y}_N \left(N = \sum_{h=1}^{L} N_h\right) \tag{10.6.2}$$

is

$$\hat{\mu}_{st} = \frac{1}{N}\sum_{h=1}^{L} N_h \hat{\mu}_h . \tag{10.6.3}$$

The standard error of this estimator is

$$\pm\left(\frac{1}{N^2}\sum_{h=1}^{L} N_h^2 M_h \hat{\sigma}_h^2\right)^{1/2} \tag{10.6.4}$$

where
$$M_h = \frac{(p_h+1)(p_h-3/2)}{n_h p_h (p_h - 1/2)}\left(1 - \frac{n_h}{N_h}\right) + \frac{3}{2}\,\frac{1}{n_h p_h (p_h - 1/2)}\,\frac{n_h}{N_h} \tag{10.6.5}$$

for $p_h > 2$ $(1 \le h \le L)$, the last term may be ignored.

For $p_h = \infty (1 \le h \le L)$, $\hat{\mu}_{st}$ reduces to the classical estimator

$$\bar{y}_{st} = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{y}_h \tag{10.6.6}$$

and (10.6.5) reduces to the standard error of $\bar{y}_{st}$, i.e.,

$$\pm\left[\frac{1}{N^2}\sum_{h=1}^{L} N_h^2 \frac{s_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right)\right]^{1/2} \tag{10.6.7}$$

where $\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ and $s_h{}^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(y_{hi} - \bar{y}_h)^2$.

Cost function: The simplest cost function is of the form (Cochran, 1977, p.96)

$$C^* = C_0 + \sum_{h=1}^{L} c_h n_h . \tag{10.6.8}$$

Assume for simplicitly that $p_h = p$ $(1 \le h \le L)$, i.e., the underlying distributions in the L strata are identical.

In the first place consider the situation when $p = \infty$ (normal distribution). Let D be a pre-determined limit for the MSE of $\bar{y}_{st}$. Equating the MSE of $\bar{y}_{st}$ to D we obtain the equation

$$D + \sum_{h=1}^{L} W_h{}^2 \frac{\sigma_h^2}{N_h} = \sum_{h=1}^{L} W_h{}^2 \frac{\sigma^2}{n_h}, \quad W_h = \frac{n_h}{N_h}. \tag{10.6.9}$$

To find n_h which minimizes the variance subject to the allocated cost C*, we minimize

$$\sum_{h=1}^{L} W_h^{\,2}\,\frac{\sigma^2}{n_h} + \eta\sum_{h=1}^{L} c_h n_h$$

where η is a Lagrange multiplier. Differentiating with respect to n_h we obtain

$$\sqrt{\eta}\,n_h = W_h\sigma_h/\sqrt{c_h}\,. \tag{10.6.10}$$

Summing both sides over h = 1, 2, ..., L, we obtain

$$\sqrt{\eta} = \frac{1}{n}\,(\Sigma W_h\sigma_h/\sqrt{c_h}\,).$$

This gives the value of n_h,

$$n_h = n\,\frac{W_h\sigma_h}{\sqrt{c_h}} \Big/ \sum_{h=1}^{L}\frac{W_h\sigma_h}{\sqrt{c_h}}\,. \tag{10.6.11}$$

Substituting (11.6.11) in (11.6.9) gives the value of n to attain a pre-determined MSE subject to the allocated cost C*,

$$n = \frac{(\Sigma_h W_h\sigma_h\sqrt{c_h}\,)\Sigma_h(W_h\sigma_h/\sqrt{c_h}\,)}{D + (1/N)\Sigma_h W_h\sigma_h^2}\,. \tag{10.6.12}$$

This result is due to Cochran (1977).

Proceeding exactly along the same lines and ignoring the second term in (10.6.5) for simplicity we obtain the following value of n,

$$n = \frac{(\Sigma_h W_h\sigma_h\sqrt{c_h}\,)\Sigma_h(W_h\sigma_h/\sqrt{c_h}\,)}{\{p(p-1/2)/(p+1)(p-3/2)\}\,D + (1/N)\Sigma_h W_h\sigma_h^2}\,. \tag{10.6.13}$$

Since p(p – 1/2)/(p + 1)(p – 3/2) is greater than 1, (10.6.13) is always smaller than (10.6.12). Thus, a smaller sample size n is needed to satisfy the cost constraint (10.6.8) if we use the MMLE $\hat{\mu}_{st}$ in place of the traditional estimator $\bar{y}_{st}$. For p = ∞, of course, (10.6.13) reduces to (10.6.12).

The results above are true for the symmetric family (2.2.9) which has the interesting property that the location parameter (mode) μ is also the mean of the distribution. We now show that the MMLE are enormously more efficient than the traditional estimators for skew distributions as well. We will illustrate it in the framework of a super-population model. The results for a finite population model are similar.

10.7 SKEW DISTRIBUTIONS IN SAMPLE SURVEY

For illustration assume that the underlying super-population is the Generalized Logistic given in (2.5.1). Here, we will denote it by GL(y; b). The mean of the distribution is $\mu + \{\psi(b) - \psi(1)\}\sigma$, μ being the mode of the distribution. In practice, one might be interested in estimating the mode μ.

Case I (σ known): In the context of the super-population GL(y; b), the mode of a finite population Π_N is

$$\bar{Y}_{N.} = \bar{Y}_N - \{\psi(b) - \psi(1)\}\sigma \tag{10.7.1}$$

which, of course, is not known.

Let $\bar{y}_n$ be the mean of a simple random sample $y_1, y_2, \ldots, y_n$ from Π_N. The classical estimator of $\bar{Y}_{N.}$ is

$$\bar{y}_{n.} = \bar{y}_n - \{\psi(b) - \psi(1)\}\sigma \tag{10.7.2}$$

with variance

$$V(\bar{y}_{n.}) = V(\bar{y}_n) = \{\psi'(b) + \psi'(1)\}(\sigma^2/n). \qquad (10.7.3)$$

The values of the psi-function $\psi(b)$ and its derivative $\psi'(b)$ are given in Chapter 2 (Appendix 2D).

For the super-population GL(y; b), $E(\bar{y}_{n.})$ and $E(\bar{Y}_{N.})$ are both equal to μ. Hence, $E(\bar{y}_{n.} - \bar{Y}_{N.}) = 0$. Here, the MSE

$$E(\bar{y}_{n.} - \bar{Y}_{N.})^2 = \left(1 - \frac{n}{N}\right)^2 \left(\frac{\sigma^2}{n} + \frac{\sigma^2}{N-n}\right) (\psi'(b) + \psi'(1)\} \qquad ((10.7.4)$$

$$= \left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n} (\psi'(b) + \psi'(1)\}.$$

The MMLE: The MMLE of μ is computed from the order statistics $y_{(i)}$ $(1 \le i \le n)$,

$$\hat{\mu}_{n.} = \hat{\mu}_n - \frac{1}{m}\left(\sum_{i=1}^{n} \alpha_i - \frac{n}{b+1}\right)\sigma \qquad (10.7.5)$$

where $\hat{\mu}_n = (1/m)\sum_{i=1}^{n} \beta_i y_{(i)}$ and $m = \sum_{i=1}^{n} \beta_i$. (10.7.6)

The coefficients α_i and β_i are given in (2.5.5). Note that β_i are all positive.

Since $\hat{\mu}_{n.}$ and $\bar{Y}_{N-n.} = \bar{Y}_{N-n} - \{\psi(b) - \psi(1)\}\ \sigma$ are independent of one another and Cov (X + c, Y + d) = Cov(X, Y), c and d being constants, the MSE

$$E(\hat{\mu}_{n.} - \bar{Y}_{N.})^2 = E\{(\hat{\mu}_{n.} - \mu) - (\bar{Y}_{N.} - \mu)\}^2 \qquad (10.7.7)$$

$$= V(\hat{\mu}_n) - 2\frac{n}{N}\,\text{Cov}\,(\hat{\mu}_n, \bar{y}_n) + \frac{\sigma^2}{N}\ \{\psi'(6) + \psi'(1)\}$$

which is similar to the expression in (10.3.4). Therefore,

$$E(\bar{y}_{n.} - \bar{Y}_{N.})^2 \ge E(\hat{\mu}_{n.} - \bar{Y}_{N.})^2$$

if $$\{\psi'(b) + \psi'(1)\}\ \frac{\sigma^2}{n} \le V(\hat{\mu}_n) + 2\frac{n}{N}\ \{\psi'(b) + \psi'(1)\}\ \frac{\sigma^2}{n} - \text{Cov}\,(\hat{\mu}_n, \bar{y}_n). \qquad (10.7.8)$$

We now show that (11.7.8) is true for all values of b in GL(y; b). Consequently, $\hat{\mu}_{n.}$ is always more efficient than the classical estimator $\bar{y}_{n.}$, in fact, considerably more efficient.

Large n: When n is large, the variance of $\hat{\mu}_{n.}$ is given by (Chapter 2)

$$1/\{-E(d^2 \ln L/d\mu^2)\} = (b + 2)\sigma^2/nb, \qquad (10.7.9)$$

L being the likelihood function of the random sample y_i $(1 \le i \le n)$. See also Table 10.4 which gives the exact values of $V(\hat{\mu}_{n.})$ and the values calculated from (10.7.9).

Since the order statistics $y_{(i)}$ and $y_{(j)}$ are positively correlated, and the coefficients β_i are all positive, $\text{Cov}(\hat{\mu}_n, \bar{y}_n)$ is always positive. Consequently,

$$\rho = \text{Cov}(\hat{\mu}_n, \bar{y}_n)/\sqrt{V(\hat{\mu}_n)\,V(\bar{y}_n)}$$

is always positive and $0 < \rho < 1$. Since for the family GL(y; b), $V(\hat{\mu}_n) \le V(\bar{y}_n)$ as established in Chapter 2, Cov $(\hat{\mu}_n, \bar{y}_n)$ is equal to or greater than $\rho V(\hat{\mu}_n)$ for all positive values of ρ. This implies that Cov $(\hat{\mu}_n, \bar{y}_n) \ge V(\hat{\mu}_n)$. A similar argument leads to the result that Cov $(\hat{\mu}_n, \bar{y}_n)$ $\le V(\bar{y}_n)$; see also Table 10.4 which gives the exact values of the variances and covariances. Noting that the variance

$$V(\bar{y}_n) = \{\psi'(b) + \psi'(1)\}\ (\sigma^2/n),$$

(10.7.8) holds if

$$\{\psi'(b) + \psi'(1)\} \geq B^* \tag{10.7.10}$$

where the bound B* is

$$B^* = \frac{b+2}{b} + 2\frac{n}{N}\left[\{\psi'(b) + \psi'(1)\} - \frac{b+2}{b}\right]. \tag{10.7.11}$$

The values of $\psi'(b) + \psi'(1)$ are given in Table 10.3. Also given are the values of B* for the sample fractions n/N=0.01, 0.10 and 0.20. It can be seen that (10.7.10) is always true. Hence, $\hat{\mu}_{n.}$ is more efficient than $\bar{y}_{n.}$ for all values of b.

Table 10.3: Exact values of the bound B*.

n/N	b	$\psi'(b) + \psi'(1)$	B*	b	$\psi'(b) + \psi'(1)$	B*
0.01	0.2	27.911	11.338	0.5	6.5797	5.0316
0.10			14.382			5.3159
0.20			17.764			5.6319
0.01	1	3.2898	3.0058	2	2.2898	2.0058
0.10			3.0580			2.0580
0.20			3.1159			2.1159
0.01	4	1.9287	1.5086	8	1.7780	1.2606
0.10			1.5857			1.3556
0.20			1.6715			1.4612

Small n: Since $E\{y_{(i)}\} = \mu + \sigma t_{(i)}$ $(1 \leq i \leq n)$, where $t_{(i)} = E\{z_{(i)}\}$ and $z_{(i)} = \{y_{(i)} - \mu\}/\sigma$,

$$E(\hat{\mu}_{n.}) = \mu - \frac{1}{m}\left[\sum_{i=1}^{n}(1+e^{t_{(i)}})^{-1} - \frac{n}{b+1}\right]\sigma, \tag{10.7.12}$$

so that

$$(\text{Bias})^2 = \frac{1}{m^2}\left[\sum_{i=1}^{n}(1+e^{t_{(i)}})^{-1} - \frac{n}{b+1}\right]^2 \sigma^2. \tag{10.7.13}$$

Let $\Omega = (v_{ij})$ $(i, j = 1, 2, \ldots, n)$

be the variance-covariance matrix of the standardized ordered variates $z_{(i)}$ $(1 \leq i \leq n)$. Then,

$$(1/\sigma^2)V(\hat{\mu}_n) = (\beta'\,\Omega\,\beta)/m^2$$

and (10.7.14)

$$(1/\sigma^2)\,\text{Cov}\,(\hat{\mu}_n, \bar{y}_n) = (\beta'\,\Omega\,1)/mn;$$

$\beta' = (\beta_1, \beta_2, \ldots, \beta_n)$ and $1' = (1, 1, \ldots, 1)$ are $n \times 1$ vectors. For $n \leq 15$ and b = 1.0 (0.5)5, 6,7 and 8, the elements v_{ij} of Ω are tabulated in Balakhrishnan and Leung (1988) as said earlier in Chapter 2. Using these tables, we calculate the values of $(\text{Bias})^2$, $V(\hat{\mu}_n)$, $\text{Cov}\,(\hat{\mu}_n, \bar{y}_n)$ and the bound B*. They are given in Table 10.4 for b = 2 and 8 for illustration; $\sigma = 1$ without any loss of generality. The values of $\psi'(b) + \psi'(1)$ are given in Table 10.3:

Table 10.4: Exact values of $(\text{Bias})^2$, $V(\hat{\mu}_n)$ and $\text{Cov}(\hat{\mu}_n, \bar{y}_n)$ and the bound B*.

	b = 2			b = 8		
	n = 6	10	15	n = 6	10	15
$(\text{Bias})^2$	0.0046	0.0016	0.0006	0.0176	0.0067	0.0030
$V(\hat{\mu}_n)$	0.3444	0.2041	0.1354	0.2209	0.1292	0.0850
(b + 2)/nb	0.3333	0.2000	0.1333	0.2083	0.1250	0.0833
$\text{Cov}(\hat{\mu}_n, \bar{y}_n)$	0.3450	0.2044	0.1355	0.2237	0.1311	0.0854
B*: n/N						
0.01	2.0708	2.0459	2.0362	1.3509	1.3020	1.2849
0.10	2.1104	2.0902	2.0825	1.4293	1.3858	1.3741
0.20	2.1543	2.1393	2.1339	1.5165	1.4789	1.4732

Comparing the values of B* with $\psi'(b) + \psi'(1)$ it is seen that $\hat{\mu}_n$ is more efficient than the classical estimator $\bar{y}_n$ for all n. Realize that the bias in $\hat{\mu}_n$ is negligibly small and the asymptotic formula (10.7.9) gives accurate approximations even for small sample sizes as small as n = 10.

Relative efficiency: Consider the difference

$$(1/\sigma^2)\{\text{Cov}(\hat{\mu}_n, \bar{y}_n) - V(\hat{\mu}_n)\} = \{l'\,\Omega\,(1 - l)\}/n^2 \tag{10.7.15}$$

where

$$l' = (n/m)(\beta_1, \beta_2, \ldots, \beta_n) \quad \text{and} \quad l' = (1, 1, \ldots\ldots, 1) \tag{10.7.16}$$

are n × 1 vectors of constant coefficients. The elements of l' are finite and positive since $m/n \cong b/(b+1)(b+2)$ (equation 2D.5) and β_i are all positive fractions (Chapter 2). Also, the elements ν_{ij} of Ω are all finite and positive with

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \nu_{ij} = n\{\psi'(b) + \psi'(1)\}.$$

Therefore, (10.7.15) is of order 1/n and tends to zero as n becomes large. In fact, (10.7.15) tends to zero very quickly since some of the elements in the 1 × n vector 1 – l are negative while others are positive. Consequently, the sum of the negative and positive terms counterbalance each other reducing (10.7.15) to zero (almost). For b = 2, for example, the exact values of (10.7.15) are 0.0006, 0.0003 and 0.0001 for n = 6, 10 and 15, respectively. The values are similarly almost zero for b ≠ 2. Thus, $\text{Cov}(\hat{\mu}_n, \bar{y}_n)$ can be replaced by $V(\hat{\mu}_n)$ to obtain the simpler formula

$$E(\hat{\mu}_{n.} - \bar{Y}_{N.})^2 \cong \left(1 - \frac{n}{N}\right) V(\hat{\mu}_n) + \frac{n}{N}\left[\frac{\sigma^2}{n}\{\psi'(b) + \psi'(1)\} - V(\hat{\mu}_n)\right] \tag{10.7.17}$$

the second term on the right hand side being of order $O(N^{-1})$.

Given below are the values of the relative efficiency of $\bar{y}_n$ defined in (10.3.5). It can be seen that the MMLE $\hat{\mu}_{n.}$ is enormously more efficient than the classical estimator $\bar{y}_n$. It may be remembered that for b = 1, GL(y; b) is the logistic distribution which is in close proximity to a normal distribution. Even here the MMLE is more efficient.

Relative efficiency (percent) of the sample mean

	b =	0.2	0.5	1	2	4	8
n/N =	0.01	40	76	91	88	78	70
	0.10	46	79	93	88	82	74
	0.20	55	82	93	91	86	78

Comment: The variance $V(\hat{\mu}_n)$ is essentially equal to $(b+2)\,\sigma^2/nb$ as in (10.7.9). The difference $(1/N)\{\psi'(b) + \psi'(1) - (b + 2)/b\}\,\sigma^2$ is small for all sample fractions $n/N \le 0.2$, since N is large. For n = 10 and n/N = 0.1, for example, N = 100 in which case the difference is equal to $0.003\ \sigma^2$ and $0.005\ \sigma^2$ for b = 2 and 8, respectively. Therefore, it suffices to take the MSE to order O(1/N) as

$$E(\hat{\mu}_{n.} - \overline{Y}_{N.})^2 \cong \frac{b+2}{nb}\left(1-\frac{n}{N}\right)\sigma^2. \qquad (10.7.18)$$

It is interesting to note that (10.7.18) is exactly similar to (10.3.8).

Sample-size determination: The sample size determination problems are resolved along the same lines as before. To determine n to attain a pre-determined mean square error D, for example, we have the following results.

Equating the MSE of $\bar{y}_n$ to D, we get

$$n = \sigma^2 \Big/ \left[D\{\psi'(b) + \psi'(1)\}^{-1} + \frac{\sigma^2}{N}\right].$$

Equating the MSE of $\hat{\mu}_{n.}$ to D, we get

$$n = \sigma^2 \Big/ \left[D\frac{b}{b+2} + \frac{\sigma^2}{N}\right]. \qquad (10.7.19)$$

Since $\psi'(b) + \psi'(1)$ is greater than $(b+2)/b$, a smaller sample size n is needed to attain a pre-determined MSE if the MMLE $\hat{\mu}_{n.}$ is used.

Remark: The results extend to stratified sampling, exactly along the same lines as in Section 10.6.

Case II (unknown scale): If σ is not known, it is replaced by the MMLE $\hat{\sigma}_n$ calculated from the order statistics $y_{(i)}$ $(1 \le i \le n)$;

$$\hat{\sigma}_n = \{B + \sqrt{(B^2 + 4nC)}\}/2\sqrt{\{n(n-1)\}} \qquad (10.7.20)$$

where

$$B = (b+1)\sum_{i=1}^{n}\left(\frac{1}{b+1} - \alpha_i\right)(y_{(i)} - \hat{\mu}_n) \quad \text{and} \quad C = (b+1)\sum_{i=1}^{n}\beta_i(y_{(i)} - \hat{\mu}_n)^2.$$

The MMLE of μ is, therefore,

$$\hat{\mu}_{n..} = \hat{\mu}_n - a\hat{\sigma}_n, \quad a = (1/m)\left\{\sum_{i=1}^{n}\alpha_i - \frac{n}{b+1}\right\}. \qquad (10.7.21)$$

The mean square error $E(\hat{\mu}_{n..} - \overline{Y}_{N.})^2$ is estimated by

$$\frac{b+2}{nb}\hat{\sigma}_n^2\left(1-\frac{n}{N}\right). \qquad (10.7.22)$$

The classical estimator of μ is

$$\bar{y}_{n..} = \bar{y}_n - \{\psi(b) - \psi(1)\}s_n, \quad s_n{}^2 = \sum_{i=1}^{n}(y_i - \bar{y}_n)^2/(n-1). \qquad (10.7.23)$$

The mean square error $E(\bar{y}_{n..} - \bar{Y}_{N.})^2$ is estimated by

$$\frac{\psi'(b) + \psi'(1)}{n} s_n^2 \left(1 - \frac{n}{N}\right). \qquad (10.7.24)$$

For large n, $\hat{\sigma}_n$ and s_n are both unbiased estimators and will be close to σ. But (10.17.22) is smaller than (10.17.24) since (b+2)/b is smaller than $\psi'(b) + \psi'(1)$ for all values of b.

Revised MSE: More accurate values of the MSE

$$E(\bar{y}_{n..} - \bar{Y}_{N.})^2 \quad \text{and} \quad E(\hat{\mu}_{n..} - \bar{Y}_{N.})^2$$

can, of course, be obtained by incorporating the variances and the covariance of the estimators of μ and σ. It is, however, difficult to evaluate the covariance of $\bar{y}_n$ and s_n for non-normal distributions and we will not pursue it here. In fact, the variances and the covariance are expressions in terms of the Fisher κ — statistics (see, for example, Kendall and Stuart, 1969). For the MMLE, however, the variances and the covariance are given by I^{-1}, where

$$I = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

Here,

$$I_{11} = \frac{nb}{(b+2)\sigma^2}, \quad I_{12} = I_{21} = \frac{nb}{(b+2)\sigma^2}\{\psi(b+1) - \psi(2)\} \qquad (10.7.25)$$

and

$$I_{22} = \frac{n}{\sigma^2} + \frac{nb}{(b+2)\sigma^2}[\psi'(b+1) + \psi'(2) + \{\psi(b+1) - \psi(2)\}^2].$$

Example 10.1: Take b = 2. The Fisher information matrix is

$$I = \frac{n}{\sigma^2}\begin{bmatrix} 0.50 & 0.25 \\ 0.25 & 1.6449 \end{bmatrix}$$

which gives the variance-covariance matrix

$$I^{-1} = \frac{\sigma^2}{n}\begin{bmatrix} 2.1643 & -0.3289 \\ -0.3289 & 0.6579 \end{bmatrix}.$$

Notice that the variance of $\hat{\sigma}_n$ is much smaller than that of $\hat{\mu}_n$, in agreement with the normal-theory result that $V(\bar{y}_n) = \sigma^2/n$ and $V(s_n) \cong \sigma^2/2n$.

It follows from the results given in Chapter 2 (Appendix A2.4) that for large n,

$$a \cong \psi(b+1) - \psi(2).$$

For b = 2, $a \cong 0.5$. Thus, the revised value of the mean square error $E(\hat{\mu}_{n..} - \bar{Y}_{N.})^2$ is 2.493 $(1 - n/N)\,\hat{\sigma}_n^2/n$ as against $2(1 - n/N)\,\hat{\sigma}_n^2/n$ given in (10.7.18). The revised value is marginally bigger than that obtained from (10.7.18) as expected.

10.8 ESTIMATING THE MEAN

Rather than estimating the mode of the finite population Π_N, one may like to estimate its mean. The mean of the super-population GL(y; b) is

$$\theta = \mu + \{\psi(b) - \psi(1)\}\sigma. \qquad (10.8.1)$$

In the first place, we note that values of b less than 0.5 or greater than 8 are not of interest for the parameter θ to be meaningful. The reason is that for such values of b, a proportion substantially different from ½ of the population values exceed the mean θ. In fact, the probability P(Y < θ) is given by

$$P(Y < \theta) = b \int_{-\infty}^{c} \{e^{-z} / (1 + e^{-z})^{b+1}\} \, dz \tag{10.8.2}$$

$$= (1 + e^{-c})^{-b}, \quad c = \psi(b) - \psi(1).$$

Given below are the values of this probability

b =	0.2	0.5	1	2	4	8
prob:	0.389	0.447	0.500	0.534	0.553	0.562

It is clear that only the values of b in the range $0.5 \le b \le 8$ are of interest. Irrespective of the value of b, however, the classical estimator of the mean $\overline{Y}_N$ of Π_N is the sample mean $\bar{y}_n$ with MSE

$$E(\bar{y}_n - \overline{Y}_N)^2 = \left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n} \{\psi'(b) + \psi'(1)\}. \tag{10.8.3}$$

The MMLE of θ is

$$\hat{\theta}_n = \hat{\mu}_n - d\hat{\sigma}_n \tag{10.8.4}$$

where $\hat{\mu}_n$ and $\hat{\sigma}_n$ are as usual the MMLE of μ and σ calculated from the order statistics $y_{(i)}$ $(1 \le i \le n)$, and

$$d = (1/m)\{\textstyle\sum_{i=1}^{n} \alpha_i - n/(b+1)\} - \psi(b) + \psi(1) \cong \psi(b+1) - \psi(b) - 1 \tag{10.8.5}$$

since $\psi(2) - \psi(1) = 1$.

For small n/N (≤ 0.2) to order $O(N^{-1})$,

$$E(\hat{\theta}_n - \overline{Y}_N)^2 \cong \left(1 - \frac{n}{N}\right)\{V(\hat{\mu}_n) - 2d\,\mathrm{Cov}(\hat{\mu}_n, \hat{\sigma}_n) + d^2 V(\hat{\sigma}_n)\} \tag{10.8.6}$$

where the variances and the covariance are given by I^{-1}. The elements of I are given in (10.7.25). Calculations show that (10.8.6) is smaller than (10.8.3) for all b in the range 0.5 to 8.

Example 10.2: Take b = 2 in which case d = – 0.5. From the elements of I^{-1}, we obtain

$$E(\hat{\theta}_n - \overline{Y}_N)^2 \cong \left(1 - \frac{n}{N}\right)(2.1969 - 0.1968 + 0.0794)\frac{\sigma^2}{n} = \left(1 - \frac{n}{N}\right)\frac{2.0795\sigma^2}{n}.$$

From (10.8.3), we obtain

$$E(\bar{y}_n - \overline{Y}_N)^2 = \left(1 - \frac{n}{N}\right)\frac{2.2898\sigma^2}{n};$$

$\bar{y}_n$ is clearly less efficient than $\hat{\theta}_n$.

Robustness: The estimators above are remarkably robust to plausible deviations from an assumed distribution, to outliers in a sample and to mixtures and contaminations. This is due to umbrella or half-umbrella ordering of the coefficients β_i given to the order statistics $y_{(i)}$ $(1 \le i \le n)$, as explained in Chapter 8.

Remark: Tiku (1983) develops an estimator $\hat{\mu}_A$ of the mean of a finite population based on a censored normal sample of size A = n – 2r. He also develops Fuller (1970) and Rao (1975) type estimator

$$\mu_n{}^* = \frac{n}{N}\bar{y}_n + \left(1 - \frac{n}{N}\right)\hat{\mu}_A \tag{10.8.7}$$

He shows that both $\hat{\mu}_A$ and $\hat{\mu}_n*$ have good robustness properties for long-tailed symmetric distributions and in situations when the sample contains outliers. It may be noted that outliers in sample survey data can particularly occur due to measurement errors.

SUMMARY

In this chapter, we discuss estimation of the mean of a finite population in the context of sample survey. The traditional estimator is the sample mean. We study the MMLE and show that they are enormously more efficient than the sample mean for non-normal populations. We discuss sample-size determination problems and show that by using the MMLE, smaller sample sizes are needed to attain certain optimality properties. We extend the results to stratified sampling.

APPENDIX 10A

Table 10A.1: Exact values* of the variances and covariances; (1)$(1/\sigma^2)V(\hat{\mu}_n)$ and (2) $(1/\sigma^2)\mathrm{Cov}(\hat{\mu}_n, \hat{\sigma}_n)$.

n	(1)	(2)	(1)	(2)	(1)	(2)
	p = 1.5		p = 2.0		p = 2.5	
6	0.18782	0.23378	0.09700	0.09560	0.12985	0.13443
8	0.13026	0.16261	0.07028	0.06662	0.09451	0.09666
10	0.09932	0.12196	0.05511	0.05133	0.07438	0.07538
12	0.08019	0.09649	0.04530	0.04188	0.06134	0.06176
14	0.06707	0.08041	0.03843	0.03545	0.05220	0.05232
15	0.06225	0.07275	0.03548	0.03249	0.04858	0.04861
16	0.05787	0.06708	0.03335	0.03078	0.04544	0.04540
17	0.05389	0.06252	0.03128	0.02889	0.04267	0.04258
18	0.05082	0.05794	0.02946	0.02722	0.04168	0.04443
19	0.04789	0.05426	0.02783	0.02574	0.03804	0.03788
20	0.04528	0.05098	0.02637	0.02442	0.03608	0.03591
	p = 3.5		p = 4.5		p = 5.0	
6	0.15235	0.15539	0.15928	0.16106	0.16101	0.16242
8	0.11237	0.11448	0.11829	0.11964	0.11982	0.12091
10	0.08900	0.09048	0.09405	0.09507	0.09537	0.09621
12	0.07367	0.07475	0.07804	0.07883	0.07920	0.07986
14	0.06285	0.06367	0.06668	0.06731	0.06770	0.06823
15	0.05856	0.05929	0.06216	0.06272	0.06313	0.06361
16	0.05480	0.05543	0.05821	0.05872	0.05913	0.05956
17	0.05151	0.05206	0.05471	0.05517	0.05561	0.05600
18	0.04774	0.04828	0.05165	0.05206	0.05248	0.05284
19	0.04598	0.04642	0.04811	0.04864	0.04969	0.05002
20	0.04364	0.04403	0.04642	0.04677	0.04718	0.04748

For $p = \infty$, $V(\hat{\mu}_n) = \mathrm{Cov}(\hat{\mu}_n, \bar{y}_n) = \sigma^2/n$.

CHAPTER 11

Applications

11.1 INTRODUCTION

We examined a large number of real life data sets in numerous areas of application. We chose nearly twenty representative ones which we analyse and present in this chapter. We found very few data sets which can be called normal. This is in conformity with the experiences of Pearson (1931), Geary (1947), Elveback et al. (1970) and Spjϕtvoll and Aastveit (1980) and many others. All the three types of distributions prevail in practice: (i) symmetric long-tailed, (ii) skew, and (iii) symmetric short-tailed. Thus, the families of distributions (2.2.9), (2.5.1) and (3.6.2) considered in Chapters 2 – 6 are of great deal of relevance. We were, in fact, pleasantly surprised to find out how useful the short-tailed symmetric distributions (3.6.2) are in modelling the error distributions in the context of regression and autoregression. It seems that if a linear model provides a very good fit with a large value of the coefficient of determination R^2 (≥ 0.90, say), then the error distribution is mostly short-tailed. However, it is not possible to identify the underlying distribution exactly from a sample. Since the MMLE are robust to plausible deviations, it suffices to locate a distribution in reasonable proximity to the true distribution. This can easily be accomplished by constructing Q – Q plots followed by a formal goodness-of-fit test. A viable alternative to a goodness-of-fit test is to determine the value of a shape parameter by maximizing ln L discussed in Section 11.3. Thus, the methodology of modified likelihood beautifully adapts to normal and non-normal samples. Under non-normality, of course, it yields more efficient estimators than the normal-theory estimators. The method also gives precise estimates of the population standard deviation (scale parameter) which should be an important feature of any estimator of σ.

11.2 ESTIMATORS OF LOCATION AND SCALE PARAMETERS

In this section we give several examples to illustrate that from applications of modified likelihood methodology, estimators of the population mean and standard deviation can easily be obtained which are more efficient than the sample mean and standard deviation. It may be remembered that the methodology yields MMLE which are essentially as efficient as the MLE. The computation of the latter is generally problematic. Consequently, the MLE are elusive in most situations. There is no such problem with the MMLE.

Example 11.1: The following well-known data due to Cushney and Peebles (1905) measure the prolongation of sleep by two soporific drugs as ordered differences y: 0.00 0.8 1.0 1.2 1.3 1.3 1.4 1.8 2.4 4.6

The problem is to estimate $\theta = E(y)$ and find the standard error of the estimate.

Under the normality assumption the estimates of the population mean and standard deviation θ and σ, respectively, are

$$\bar{y} = 1.58 \quad \text{and} \quad s = 1.230, \quad \text{with}$$

$$SE(\bar{y}) = \pm\, s/\sqrt{n} = \pm\, 1.230/\sqrt{10} = \pm\, 0.389 \quad \text{and} \quad SE(s) \cong \pm\, s/\sqrt{2n} = \pm\, 0.275.$$

The normal distribution, however, is not a good model for the data. In fact, the Shapiro-Wilk statistic assumes the value

$$W = \{0.574(4.6 - 0.0) + 0.329(2.4 - 0.8) + 0.214(1.8 - 1.0) + 0.122(1.4 - 1.2)\}^2/13.616$$
$$= 0.781$$

which is smaller than the 5 percent significance level 0.830, let alone the 10 percent significance level advocated in Chapter 9. We reject normality.

Using Q-Q plots accompanied with the results of the Z* (= 0.844) test (Chapter 9), we realize that the Generalized Logistic (2.5.1) with b = 8 provides a plausible model; see also Section 11.3. We now calculate the MMLE of the location and scale parameters from (2.5.11). They are

$$\hat{\mu} = -\,0.527 \quad \text{and} \quad \hat{\sigma} = 0.833;$$

σ is the scale parameter of the assumed Generalized Logistic and not its standard deviation. Thus,

$$\hat{\theta} = \hat{\mu} + \{\psi(8) - \psi(1)\}\hat{\sigma} = -\,0.527 + \{2.016 + 0.577\}\,(0.833) = 1.63.$$

The variance-covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is from (2.5.12),

$$V = \frac{\sigma^2}{10}\begin{bmatrix} 3.083 & -1.067 \\ -1.067 & 0.621 \end{bmatrix} \qquad (11.2.1)$$

which gives

$$V(\hat{\theta}) \cong \frac{\sigma^2}{10}\,\{3.083 - 2(2.593)(1.067) + (2.593)^2\,(0.621)\} = 0.117\,\sigma^2,$$

$$SE(\hat{\theta}) \cong \pm\sqrt{\{(0.117)(0.833)^2\}} = \pm\, 0.285$$

The MMLE of the population standard deviation σ_1 is

$$\hat{\sigma}_1 = \sqrt{\{\psi'(8) + \psi'(1)\}\,\hat{\sigma}^2} = 1.111 \quad \text{and} \quad SE(\hat{\sigma}_1) \cong \pm\sqrt{\{0.0621\hat{\sigma}^2\}} = \pm\, 0.207.$$

The MMLE $\hat{\theta}$ and $\hat{\sigma}_1$ are numerically close to $\bar{y}$ and s, respectively, but have considerably smaller standard errors. In fact, the standard error of s is even larger than the value 0.275 above since (equation 1.2.9)

$$V(s) \cong \frac{\sigma^2}{2n}\left(1 + \frac{1}{2}\lambda_4\right) \qquad (11.2.2)$$

and $\lambda_4 = (\mu_4/\mu_2^{\,2}) - 3 > 2$ for the Generalized Logistic with b=8 (Appendix 2D).

Example 11.2: The following data appear in Thode (2002, p.348) and represent the average annual erosion rates y of thirteen states in US:

– 0.4 – 0.5 – 0.9 – 0.5 0.1 – 1.0 0.1 – 1.5 – 4.2 – 0.6 – 2.0 0.7 – 0.1

One wants to estimate $\theta = E(Y)$ and find the standard error of the estimate. Here,

$$\bar{y} = -\,0.831, \quad s = 1.234; \; SE(\bar{y}) = \pm\, 1.234/\sqrt{13} = \pm\, 0.342.$$

By using Q-Q plots, it is seen that the Generalized Logistic with b=0.5 provides a plausible model for the data. As in the previous example, we obtain

$$\hat{\mu} = -0.181 \quad \text{and} \quad \hat{\sigma} = 0.417;$$

$$\hat{\theta} = -0.181 + \{\psi(0.5) - \psi(1)\}\hat{\sigma} = -0.759.$$

The variance-covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is

$$V = \frac{\sigma^2}{10}\begin{bmatrix} 5.113 & 0.294 \\ 0.294 & 0.760 \end{bmatrix} \tag{11.2.3}$$

This gives

$$V(\hat{\theta}) \cong 0.576\,\sigma^2 \quad \text{and} \quad SE(\hat{\theta}) \cong \pm\sqrt{\{(0.576)(0.417)^2\}} = \pm\,0.316$$

The MMLE $\hat{\theta}$ is numerically close to $\bar{y}$ but is clearly more efficient.

The MMLE of the population standard deviation σ_1 is

$$\hat{\sigma}_1 = \sqrt{\{\psi'(0.5) + \psi'(1)\}\hat{\sigma}^2} = \sqrt{\{4.9348 + 1.6449\}\,(0.417)^2} = 1.070$$

and is numerically close to s.

Example 11.3: The following data represent the gain (in pounds) in weight of 20 pigs assigned randomly to two feeds A and B (this occurs as a numerical example in Tiku et al., 1986, p. 157).

A:	0.09	1.43	2.79	1.60	1.71	3.37	2.06	2.67	8.42	3.67
B:	1.96	1.79	2.60	1.40	2.22	3.45	1.16	5.71	2.93	1.40

The problem is to estimate the means μ_A and μ_B and test the null hypothesis H_0: $\mu_A = \mu_B$.

The traditional estimates of μ_A and μ_B are

A: $\bar{y}_1 = 2.781$, $s_1 = 2.238$ and $SE(\bar{y}_1) = \pm\,2.238/\sqrt{10} = \pm\,0.708$,

B: $\bar{y}_2 = 2.462$, $s_2 = 1.353$ and $SE(\bar{y}_2) = \pm\,1.353/\sqrt{10} = \pm\,0.429$.

However, the two samples clearly come from positively skew distributions. Using Q-Q plots, the plausible distributions are Generalized Logistic with b = 8; see also Section 11.3. From (2.5.11), we obtain the following values from the elements of the matrix in (11.2.1). Denoting the scale parameters of the two distributions by σ_1 and σ_2, respectively, we have the following MMLE.

A: $\hat{\mu}_1 = -0.964$, $\hat{\sigma}_1 = 1.474$; $\hat{\theta}_1 = -0.964 + 2.593(1.474) = 2.858$,

$SE(\hat{\theta}_1) \cong \pm\sqrt{\{0.117(1.474)^2\}} = \pm\,0.504$.

B: $\hat{\mu}_2 = 0.192$, $\hat{\sigma}_2 = 0.882$; $\hat{\theta}_2 = 0.192 + 2.593(0.882) = 2.479$,

$SE(\hat{\theta}_2) \cong \pm\sqrt{\{0.117(0.882)^2\}} = \pm\,0.302$;

2.593 is the value of $\psi(8) - \psi(1)$. The MMLE are clearly more efficient.

To test the null hypothesis H_0: $\mu_A = \mu_B$, we have

$$T = (2.858 - 2.479)/\sqrt{\{(0.504)^2 + (0.302)^2\}} = 0.59$$

We conclude that the two feeds are equally effective in increasing the average weight of pigs.

Example 11.4: The following data give the values of 10(y – 2.0), y being the pollution level (measurement of lead), in water samples from two lakes (Tiku et al., 1986, p.280).

Lake 1:	– 1.48	1.25	– 0.51	0.46	0.60	– 4.27	0.63	– 0.14	– 0.38	1.28
	0.93	0.51	1.11	– 0.17	– 0.79	– 1.02	– 0.91	0.10	0.41	1.11
Lake 2:	1.32	1.81	– 0.54	2.68	2.27	2.70	0.78	– 4.62	1.88	0.86
	2.86	0.47	– 0.42	0.16	0.69	0.78	1.72	1.57	2.14	1.62

One wants to test whether the pollution levels in the lakes are the same on the average. Since the power of a test is directly related to estimating efficiency (Sundrum, 1954), it is important to use efficient estimators.

As before Q-Q plots accompanied with the results of goodness-of-fit tests (based on the U* statistic = 0.935 and 0.869 for the two data sets, respectively) indicate that the symmetric family (2.2.9) with p = 3 and p = 2.5, respectively, provide plausible models for the data. In fact, both data sets seem to have an outlier on the left hand side. In calculating the MMLE, however, the family (2.2.9) with small p generates weights which are diminutive at the ends. Thus, the influence of outliers is automatically depleted.

Here, we have the following estimates of the population means and standard deviations, and their standard errors calculated from equations (2.4.8).

Lake 1: $\bar{y}_1 = -0.064$, $s_1 = 1.282$, $SE(\bar{y}_1) = \pm s_1/\sqrt{n} = \pm 0.287$;

$$\hat{\mu}_1 = 0.112,\ \hat{\sigma}_1 = 1.275,\ SE(\hat{\mu}_1) \cong \pm \sqrt{\left\{\frac{(p-3/2)(p+1)}{np(p-1/2)}\hat{\sigma}_1^{\,2}\right\}} = \pm 0.255.$$

Lake 2: $\bar{y}_2 = 1.037$, $s_2 = 1.654$; $SE(\bar{y}_2) = \pm 0.370$;

$\hat{\mu}_2 = 1.297$, $\hat{\sigma}_2 = 1.581$, $SE(\hat{\mu}_2) \cong \pm 0.296$.

The MMLE are clearly more efficient than the sample means.

To test whether the pollution levels in the two lakes are the same on the average, we have

$$T = (1.297 - 0.112)/\sqrt{\{(0.296)^2 + (0.255)^2\}} = 3.03$$

and

$$t = (1.297 - 0.064)/\sqrt{\{(0.287)^2 + (0.370)^2\}} = 2.35.$$

The null distributions of both T and t are approximately normal N(0, 1). The T statistic, however, gives a smaller probability for H_0: $\mu_1 = \mu_2$ to be true.

Tiku et al. (1986, pp. 280-281) advocate censoring the two smallest and two largest observations in both samples and regard the resulting data as Type II censored normal samples. Using the methodology of Chapter 7 (Section 7.3), we get

Lake 1: $\hat{\mu}_c = 0.108$ and $\hat{\sigma}_c = 0.966$

Lake 2: $\hat{\mu}_c = 1.246$ and $\hat{\sigma}_c = 1.261$.

The estimates of μ are essentially the same as the MMLE above. However, the estimates of σ are smaller. This is due to the fact that the estimators of σ based on censored samples including the Huber M-estimators have substantial downward bias (Chapter 8). Deliberate censoring of observations should be avoided (unless they are grossly anomalous) as said in Chapter 8.

It may be noted that the MMLE $\hat{\sigma}_c$ is unbiased (almost) if the true distribution is very close to the assumed distribution (Chapter 7). This is a distinct advantage of the method of modified likelihood estimation.

Example 11.5: Fisher (1936) gives very interesting data on the characteristics of fifty Iris Setosa plants. He gives the measurements of (a) sepal length, (b) sepal width, (c) petal length (d) petal width. The data is reproduced in Hand et al. (1994, pp. 25-26). The problem is to estimate the population means and standard deviations.

Using Q-Q plots followed by goodness-of-fit tets (Chapter 9), we identify plausible models for the data as follows. The sample skewness $\sqrt{b_1}$ and kurtosis b_2 can also be employed to identify the underlying distribution. However, b_2 should be used with caution since its distribution is awkward, even for large n.

(a) The Weibull distribution (2.8.1) with p = 3.0 provides a good model; see Section 11.3. Here, we have the following estimates.

LSE: $\bar{y} = 5.006$, $s = 0.352$, $SE(\bar{y}) = \pm\, 0.352/\sqrt{50} = \pm\, 0.050$

and $SE(s) \cong \pm\,(0.352/\sqrt{100})(0.853) = \pm\, 0.030$, since $\lambda_4 = 2.705$.

The MMLE calculated from (2.8.13) are

$$\hat{\theta} = 4.054 \quad \text{and} \quad \hat{\sigma} = 1.070.$$

This gives the MMLE of $\mu = E(y)$ and the standard deviation σ_1 of Y:

$$\hat{\mu} = \hat{\theta} + \Gamma\left(1+\frac{1}{3}\right)\hat{\sigma} = 5.009 \quad \text{and} \quad \hat{\sigma}_1 = \sqrt{\left\{\Gamma\left(1+\frac{2}{3}\right) - \Gamma^2\left(1+\frac{1}{3}\right)\right\}\hat{\sigma}^2} = 0.347.$$

The variance-covariance matrix of $\hat{\theta}$ and $\hat{\sigma}$ is from (2.8.14),

$$V = \frac{\sigma^2}{9n}\begin{bmatrix} 2.6617 & -2.4027 \\ -2.4027 & 3.1690 \end{bmatrix}.$$

Thus,

$$V(\hat{\mu}) \cong 0.0020\sigma^2 \quad \text{with} \quad SE(\hat{\mu}) \cong \sqrt{\{0.0020(1.070)^2\}} = \pm\, 0.048,$$

and

$$SE(\hat{\sigma}_1) \cong \pm\left[\frac{\hat{\sigma}_1^2}{450}(3.1690)\left\{\Gamma\left(1+\frac{2}{3}\right) - \Gamma^2\left(1+\frac{1}{3}\right)\right\}\right]^{1/2} = \pm\, 0.029.$$

The MMLE and the LSE (and their standard errors) are very close to one another. This is indeed reassuring since the Weibull $W(3, \sigma)$ and normal with a common mean and variance are not very different from one another. In fact, the coefficients of skewness and kurtosis of $W(3, \sigma)$ are 0.168 and 2.752, respectively, which are close to 0 and 3.

(b) The family (2.2.9) with p = 6 (kurtosis 3.86) beautifully models the data. We now have the following values of the mean μ and the standard deviation σ.

LSE: $\bar{y} = 3.428$, $s = 0.379$, $SE(\bar{y}) = \pm\, 0.379/\sqrt{50} = \pm\, 0.054$

and $SE(s) \cong (\pm\, s/\sqrt{100})(1.43) = \pm\, 0.054$,

since the kurtosis of the distribution is 3.86.

MMLE: $\hat{\mu} = 3.423$, $\hat{\sigma} = 0.382$,

$$SE(\hat{\mu}) \cong \pm\sqrt{\left\{\frac{(p-3/2)(p+1)}{50p(p-1/2)}\hat{\sigma}^2\right\}} = \pm\, 0.052,$$

$$SE(\hat{\sigma}) \cong \sqrt{\left\{\frac{p+1}{100(p-1/2)}\hat{\sigma}^2\right\}} = \pm\, 0.043.$$

(c) The family (2.2.9) with p = 5.5 (kurtosis 4.0) beautifully models the data. The estimates and their standard errors are

LSE: $\bar{y} = 1.462$, $s = 0.174$, $SE(\bar{y}) = \pm 0.174/\sqrt{50} = \pm 0.025$,

$SE(s) \cong (0.174/\sqrt{100})(1.5) = \pm 0.026$, since $\lambda_4 = 4.0$.

MMLE: $\hat{\mu} = 1.461$, $\hat{\sigma} = 0.174$, $SE(\hat{\mu}) \cong \pm 0.024$, $SE(\hat{\sigma}) \cong 0.020$.

(d) Generalized Logistic with b = 4 provides a good model. The estimates of the mean $\theta = E(y)$ and the scale parameter σ are

LSE: $\bar{y} = 0.246$, $s = 0.105$, $SE(\bar{y}) = s/\sqrt{50} = \pm 0.0148$,

$\tilde{\sigma} = 0.105/\sqrt{\{\psi'(b) + \psi'(1)\}} = 0.105/1.389 = 0.0756$,

$$SE(\tilde{\sigma}) \cong \pm \frac{0.0756}{\sqrt{100}}(1.879) \cong \pm 0.0142,$$

since $\sqrt{\lambda_4} = 4.758$ for the Generalized Logistic with b = 4.

MMLE: $\hat{\mu} = 0.120$, $\hat{\sigma} = 0.0694$;

the variance-covariance matrix is from (2.5.12),

$$V = \frac{\sigma^2}{n}\begin{bmatrix} 2.244 & -0.687 \\ -0.687 & 0.634 \end{bmatrix}.$$

Thus,

$$\hat{\theta} = \hat{\mu} + \{\psi(b) - \psi(1)\}\hat{\sigma} = 0.120 + 1.8333(0.0694) = 0.247;$$

$$V(\hat{\theta}) = 0.0371\sigma^2 \quad \text{and} \quad SE(\hat{\theta}) = \pm 0.0134,$$

$$V(\hat{\sigma}) = 0.0127\sigma^2 \quad \text{and} \quad SE(\hat{\sigma}) \cong \pm 0.0078.$$

For all the marginal samples above, the MMLE are numerically close to the sample mean and standard deviation but have smaller standard errors. This is due to the fact that the underlying distributions are not normal.

Remark: The real-life data sets in Examples 11.1 – 11.5 are modeled by skew or symmetric long-tailed distributions. The data sets in the next two examples, however, are modeled by short-tailed distributions.

Example 11.6: Kendall and Stuart (1960, p. 407) give 54 deviations from a simple 11-year moving average of marriage rate in England and Wales. They are

– 6	1	12	10	– 6	– 8	– 6	3	4	7	11
3	– 8	– 2	– 3	– 7	3	4	– 5	– 7	1	6
8	9	– 2	– 8	– 10	– 7	0	8	12	7	5
4	– 3	– 6	– 12	– 5	0	5	7	3	– 4	– 8
– 6	– 5	1	6	6	2	– 6	– 5	– 6	1	

They state that the deviations have a symmetric short-tailed distribution. A Q-Q plot accompanied with the information provided by the goodness-of-fit statistic U*, suggest the symmetric short-tailed distribution (3.6.2) with r = 4 and d = 1.5 as a plausible model. Its mean is μ, its scale parameter is σ and its standard deviation is $\sigma_1 = 1.936\sigma$. We now have the following estimates.

LSE: $\bar{y} = -0.0370$ and $s = 6.378$, $SE(\bar{y}) = \pm 6.378/\sqrt{54} = \pm 0.868$.

MMLE: $\hat{\mu} = 0.0949$, $\hat{\sigma} = 3.265$, $\hat{\sigma}_1 = 1.936(3.265) = 6.321$,

$$SE(\hat{\mu}) \cong \pm 3.265/\sqrt{\{54(0.515)\}} = \pm 0.619,$$

since $V(\hat{\mu}) \cong \sigma^2/nD$ from equation (3.7.9) and D = 0.5152 (equation 3.7.10). The ratio of the standard error of $\hat{\theta}_1$ to that of $\tilde{\theta}_1$ is 0.71.

Example 11.7: The following measurements of the weight X and systolic blood pressure Y of 26 randomly chosen patients is given in Montgomery and Peck (1991).

X:	165	167	180	155	212	175	190	210	200	149	158	169	170
	172	159	168	174	183	215	195	180	143	240	235	192	187
Y:	130	133	150	128	151	146	150	140	148	125	133	135	150
	153	128	132	149	158	150	163	156	124	170	165	160	159

Montgomery and Peck assume the distribution of the random vector (X, Y) to be bivariate normal. It follows that both the marginal distributions are normal. A careful examination of the data reveals that the normality assumption is not correct. The information provided by Q-Q plots suggest the following as plausible models:

(i) X has a beta distribution, $z = (x - \mu)/\sigma$,

$$f(z) \propto z^{a-1}(1-z)^{b-1}, \quad 0 < z < 1, \tag{11.2.4}$$

with a = 2.5 and b = 6.5.

(ii) Y has the symmetric short-tailed distribution (3.6.2) with r = 2 and d = 1 ($\lambda = 2$) as in the previous example; μ is the location parameter, σ is the scale parameter and σ_1 is the standard deviation.

The estimation of μ and σ for the beta distribution is deferred to Appendix 11A.1. We have not considered it in Chapter 2.

For the blood pressure data Y, we have the following estimates.

LSE: $\bar{y} = 145.62$, s = 13.422, $SE(\bar{y}) = \pm 13.422/\sqrt{26} = \pm 2.632$.

MMLE: $\hat{\mu} = 144.93$, $\hat{\sigma} = 8.053$, $\hat{\sigma}_1 = 1.679(8.053) = 14.332$,

$$SE(\hat{\mu}) \cong \pm 8.053/\sqrt{\{26(0.6354)\}} = \pm 1.981.$$

The ratio of the standard error of $\hat{\theta}_1$ to that of $\tilde{\theta}_1$ is 0.75.

Remark: For all the data sets above, the plausible distributions are non-normal. However, the first data set in the following example is modeled by a normal distribution reasonably well. We admit, however, that it is hard to find too many such data sets. Most real-life data sets are indeed non-normal. See also Elveback et al. (1970).

Example 11.8: MacGregor et al. (1979) give the supine systolic and diastolic blood pressures of 15 patients with moderate essential hypertension, immediately before and two hours after taking a drug, captopril. The interest is in estimating the average drop in the blood pressures.

From the given data, we calculate the differences 'before' minus 'after' measurements:

Patient No:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Systolic	9	4	21	3	20	31	17	26	26	10	23	33	19	19	23
Diastolic	5	1	3	– 2	11	16	23	19	12	4	8	21	–4	4	18

The Q-Q plot based on a normal distribution has "close to a straight line pattern" for the systolic measurements (Fig. 1). For this data, the MMLE are exactly the same as the sample mean and standard deviation given below.

Systolic: $\bar{y} = 18.93$, $s = 9.027$, $SE(\bar{y}) = \pm 2.33$.

For the diastolic measurements, however, the Q-Q plot based on the symmetric short-tailed distribution (3.6.2) with r = 2 and d = 1.5 ($\lambda = 4$) gives "close to a sraight line pattern" and is, therefore, a plausible model. Thus, we have the following estimates.

LSE: $\bar{y} = 9.267$ and $s = 8.614$, $SE(\bar{y}) = \pm 2.224$.

MMLE: $\hat{\mu} = 9.362$ and $\hat{\sigma} = 4.594$, $SE(\hat{\mu}) = \pm 4.594/\sqrt{15} = \pm 1.186$,

since D = 1 (equation 3.7.10). The ratio of the standard error of $\hat{\mu}$ to that of $\bar{y}$ is 53 percent.

The MMLE of the population standard deviation σ_1 is 1.915 $\hat{\sigma}$ = 8.798 which is numerically close to s.

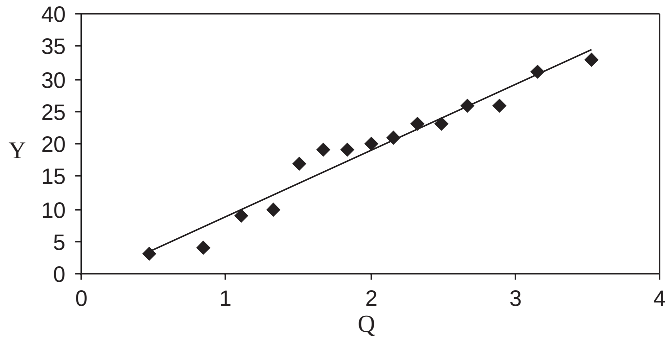

Figure 1 Systolic blood pressure.

Example 11.9: Taufer (2002, Table 5) considers the mileages for 19 military personal carriers that failed in service. They are

y: 162 200 271 320 393 508 539 629 706 778 884
1013 1101 1182 1463 1603 1984 2355 2880

The interest is in estimating $\mu = E(Y)$. The Z statistic (9.8.1) for testing exponentiality (Chapter 9) assumes the value Z = 1.05. Since (equation 9.8.5)

$$\sqrt{3(n-2)}\,(Z-1) = 0.36,$$

we conclude that the exponential $E(\theta, \sigma)$ beautifully models the data. Here, the MMLE are exactly the same as the MLE:

$$\hat{\mu} = \bar{y} = 998.474, \quad \hat{\sigma} = \frac{\sum_{i=1}^{n} y_i - ny_{(1)}}{n-1} = 882.944; \quad SE(\hat{\mu}) = \pm \frac{882.944}{\sqrt{19}} = \pm 202.57.$$

11.3 DETERMINATION OF SHAPE PARAMETERS

In all the examples above we used distributions with three parameters, the location parameter, the scale parameter and the shape parameter. We identified a plausible value of the shape parameter by using Q-Q plots, goodness-of-fit tests, or by matching (approximately) the sample skewness and kurtosis with the corresponding values of a distribution. Alternatively, we determine the shape parameter as follows (Tiku et al., 2000).

Consider, for example, the Generalized Logistic given in (2.5.1). For a given b, calculate $\hat{\mu}$ and $\hat{\sigma}$ from (2.5.11). Calculate

$$(1/n) \ln L = \ln (b/\hat{\sigma}) - (1/n)\sum_{i=1}^{n} \hat{z}_i - (1/n)(b+1)\sum_{i=1}^{n} \ln\{1 + \exp(-\hat{z}_i)\}, \quad (11.3.1)$$

$\hat{z}_i = (y_i - \hat{\mu})/\hat{\sigma}$. Do this for a series of values of b. The value $\hat{b}$ that maximizes ln L is the required estimate. For the two samples in Example 11.3, we have the following values of (1/n) ln L.

b =	0.5	1.0	5.0	6.0	7.0	8.0	9.0	10.0
A:	– 2.24	– 2.10	–1.991	– 1.989	– 1.987	– 1.987	– 1.986	– 1.986
B:	– 1.76	– 1.63	– 1.50	– 1.499	– 1.494	– 1.491	– 1.487	– 1.487

We chose b = 8.0 which is clearly a plausible value of the shape parameter. In fact, the likelihood function L is almost flat for values of b around b = 8. Any value of b close to b = 8 provides a plausible model. Due to the intrinsic robustness of the MMLE, they yield essentially the same estimates and standard errors.

Consider the data set (a) in Example 11.5. We chose the Weibull (2.8.1) with shape parameter p = 3 as a plausible model for the data. Write

$$\hat{z}_i = (y_i - \hat{\theta}_1)/\hat{\sigma} (1 \le i \le n) \quad \text{and} \quad \hat{\theta}_1 = y_{(1)} - \{\Gamma(1 + 1/p)/n^{1/p}\}\hat{\sigma}; \quad (11.3.2)$$

$\hat{\sigma}$ is calculated from (2.8.13) and $\Gamma(1 + 1/p)/n^{1/p}$ is the value of (2.8.6) for i = 1. The MMLE of θ is, in fact, $\hat{\theta}$ given in (2.8.13). For easy computation of ln L, however, we use the estimator $\hat{\theta}_1$. It may be noted that while $\hat{\theta}$ might assume a value greater than the smallest observation, $\hat{\theta}_1$ assumes a value always less than the smallest observation in a sample. For the computation of ln L, therefore, $\hat{\theta}_1$ is preferred. We now calculate the values of

$$(l/n) \ln/L = \ln(p/\hat{\sigma}) + (1/n)(p-1) \sum_{i=1}^{n} \ln \hat{z}_i - (1/n)\sum_{i=1}^{n} \hat{z}_i^p \quad (11.3.3)$$

for a series of values of p. We choose that value which maximizes ln L. For the data set (a) mentioned above, we have the following values of (1/n)ln L.

p =	2.7	2.8	2.9	3.0	3.1	3.2	3.3	3.5
	– 0.362	– 0.360	– 0.358	– 0.357	– 0.357	– 0.358	– 0.359	– 0.364

We choose p = 3 as a plausible value of the shape parameter. We could choose p = 2.9, 3.1 or p = 3.2 as well but due to the intrinsic robustness of the MMLE, they yield essentially the same estimates and standard errors.

11.4 ESTIMATION IN LINEAR REGRESSION MODELS

In this section, we illustrate the usefulness of the modified likelihood methodology for estimating the regression coefficients in linear models.

Consider a response variable Y which depends on a single design variable X. A random error occurs in measuring Y. Thus, Y is a stochastic variable. Assume that the relationship between the two variables is given by

$$Y = \theta_0 + \theta_1 X + e; \quad (11.4.1)$$

(11.4.1) is usually called a simple linear regression model. An experiment is done with n predetermined values x_i $(1 \le i \le n)$. The corresponding responses are denoted by y_i $(1 \le i \le n)$. The linear relationship (11.4.1) implies that

$$y_i = \theta_0 + \theta_1 x_i + e_i \quad (1 \le i \le n). \quad (11.4.2)$$

The errors e_i are assumed to be iid. Traditionally, the distribution of e_i is assumed to be normal. There is now a realization that, in practice, the distribution might not be normal.

It must be mentioned here that the software package EXCEL is a very useful tool to determine whether a linear relationship between Y (or a function of Y) and X is appropriate. It also gives the value of R^2 which is a measure of accuracy of the linear model (11.4.1). A value of R^2 (coefficient of determination) close to 1 indicates that the linear model (11.4.1) is indeed very appropriate.

Error distribution:

Using LSE in non-normal situations results in considerable loss of efficiency as illustrated in Chapter 3. Unfortunately, it is almost impossible to locate the true underlying distribution. For applications of the MMLE, however, it suffices to locate a distribution which is in reasonable proximity to the true distribution. Such a distribution can be located by constructing Q-Q plots of the deviants ($e_i = w_i + \theta_0$)

$$w_i = y_i - \theta_1 x_i \quad (1 \le i \le n)$$

with θ_1 replaced by a reasonably efficient estimate which can also be easily computed, e.g., the LSE $\tilde{\theta}_1 = \sum_{i=1}^n (x_i - \bar{x})y_i / \sum_{i=1}^n (x_i - \bar{x})^2$. The distribution located by using Q-Q plots can be formally verified by doing a goodness-of-fit test. The sample skewness $\sqrt{b_1}$ and kurtosis b_2 can be used to provide complementary information about the shape of the underlying distribution. However, b_2 should be used with caution as a measure of short-or long-tailedness as said earlier. Alternatively, the statistics proposed by Hogg (1967, 1972) may be used. Hamilton (1992, p.16) has very useful Q-Q plots constructed from random samples which identify a variety of distributions, e.g., long-tailed, short-tailed, skew, long-tailed with outliers, etc., see also Weisberg (1980). We particularly mention the Q-Q plot which is indicative of short-tailed distributions: it has points at the ends which are located close to a straight line but has a considerable number of points in the middle which wiggle around. Contrary to the common belief that the deviants w_i (or errors e_i) are normally distributed in which case all ordered deviants $w_{(i)}$ plotted against the quantiles of N(0, 1) are located close to a straight line, we found a considerable number of real life data sets which yield deviants having short-tailed distributions. Consider, for example, Forbes' well known data on air pressure in the Alps and the boiling point of water; X is Boiling Point (°F) and Y is 100 times Log (Pressure).

X:	194.5	194.3	197.9	198.4	199.4	199.9	200.9	201.1	201.4
Y:	131.79	131.79	135.02	135.55	136.46	136.83	137.82	138.00	138.06
X:	201.3	203.6	204.6	209.5	208.6	210.7	211.9	212.2	
Y:	138.05	140.04	142.44	145.47	144.34	146.30	147.54	147.80	

The data has been considered by many authors in a regression analysis. See, for example, Atkinson and Riani (2000).

We calculate the LSE $\tilde{\theta}_1$ (= 0.896) and the deviants

$$w_i = y_i - 0.896 x_i \quad (1 \le i \le 17).$$

Using EXCEL, we construct a Q-Q plot of the ordered deviants $w_{(i)}$ against the quantiles $Q_i = F^{-1}(i/(n+1))$ $(1 \le i \le 17)$, F(z) being the cdf of the standard normal distribution N(0, 1). The plot is given in Figure 2.

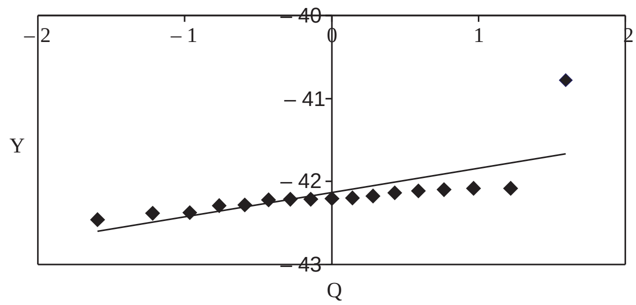

Figure 2 Forbes' data.

It can be seen that the largest deviant $w_{(17)}$ which corresponds to the pair (y_{12}, x_{12}) = (142.44, 204.6) is grossly anomalous. This observation will have to be studied separately. We set it aside for a meaningful regression analysis of the remaining 16 observations. We calculate the LSE $\tilde{\theta}_1$ and the new deviants w_i $(1 \le i \le 16)$. A Q-Q plot of the 16 deviants is given in Figure 3.

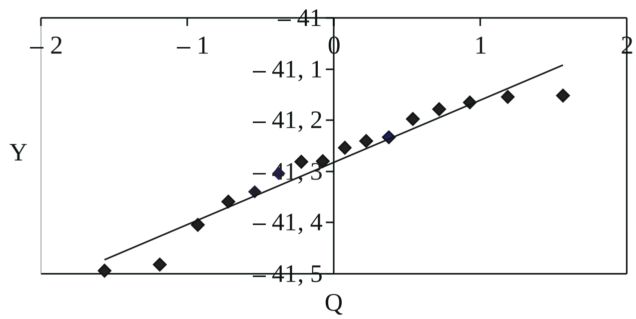

Figure 3 Reduced forbes' data.

The plot indicates a short-tailed distribution. In fact, the distribution (3.6.2) with r = 4 and d = 0 (kurtosis $\mu_4/\mu_2^2 = 2.37$) is quite plausible. To verify this, we calculate the MMLE $\hat{\theta}_0$, $\hat{\theta}_1$ and $\hat{\sigma}$ for a series of values of d (r = 4). We write

$$\hat{z}_i = (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)/\hat{\sigma}$$

and calculate the values of lnL, where L is the likelihood function based on the symmetric short-tailed distribution (3.6.2), that is,

$$L = \hat{\sigma}^{-n} C_1{}^n \prod_{i=1}^{n} \left(1 + \frac{\lambda}{2r}\hat{z}_i{}^2\right)^r \phi(\hat{z}_i), \quad \hat{z}_i = (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)/\hat{\sigma}. \tag{11.4.3}$$

Thus, we have the following values calculated from the data with the pair (142.44, 204.6) excluded:

d =	– 2.0	– 1.0	0	1.0	2.0
(1/n) ln L	0.834	0.839	0.847	0.841	0.768

The likelihood function L is clearly maximized at d = 0 (r = 4).

The standard deviation of the distribution is 1.6051σ. The LSE and the MMLE (and their standard errors) are given below.

LSE: $\tilde{\theta}_0 = -41.282$, $\tilde{\theta}_1 = 0.8909$, $s = 0.1361$, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 527.90$

MMLE: $\hat{\theta}_0 = -41.222$, $\hat{\theta}_1 = 0.8905$, $\hat{\sigma} = 0.0690$, $\hat{\sigma}_1 = 0.111$;

$D = 0.4549$ since for $r = 4$ (equation 3.7.7)

$$D = 1 - [\lambda\{1 + a - 3a^2 - 15a^3\}/\{1 + 4a + 18a^2 + 60a^3 + 105a^4\}] \quad (a = \lambda/8). \tag{11.4.4}$$

Thus,

$$SE(\tilde{\theta}_1) = \pm \frac{0.1361}{\sqrt{527.90}} = \pm 0.0059, \qquad SE(\hat{\theta}_1) = \pm \frac{0.0690}{\sqrt{0.4549(527.90)}} = \pm 0.0045.$$

The ratio of the $SE(\hat{\theta}_1)$ to the $SE(\tilde{\theta}_1)$ is 45/59 = 0.76 which is clearly indicative of the fact that the MMLE is more efficient than the LSE.

The other parameters in the model, i.e., θ_0 and σ, are essentially location and scale parameters. The estimation of location and scale parameters has already been considered in Examples 11.1 – 11.9 and the usefulness of the MMLE illustrated; see, however, Section 11.5. For small n, of course, the relative efficiencies of the LSE may be marginally bigger than those given in the Examples. This is due to the fact that the minimum variance bounds are used in calculating the variances of the MMLE.

Example 11.10: Hamilton (1992, p.24) has an interesting data set on the magnitudes and yields of 19 Soviet weapons tests; Y represents the Seismologists' magnitude estimate and X the reported Yield in Kilotons.

X:	29	125	100	4	10	60	10	125	40	90	16	12	23	16
Y:	5.6	6.1	6.0	4.8	5.2	5.8	5.4	6.0	5.7	5.9	5.5	5.3	5.5	5.4
X:	6	8	2	165	140									
Y:	5.1	5.0	4.9	6.1	6.0									

A Q-Q plot of the ordered deviants

$$w_{(i)} = y_{[i]} - \tilde{\theta}_1 x_{[i]}, \quad 1 \le i \le 19,$$

is given in Figure 4.

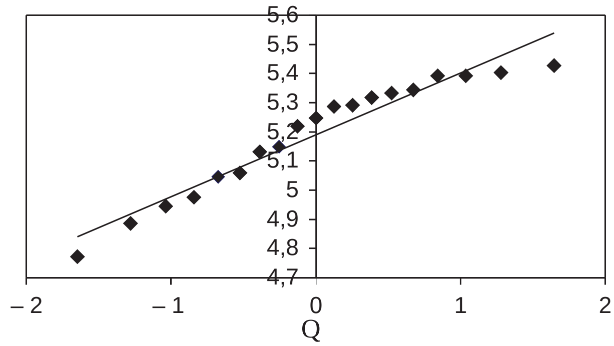

Figure 4 Seismic magnitudes.

The distribution of the errors is clearly negatively skew with a long tail on the left hand side. The Generalized Logistic GL(b, σ) is a candidate. The shape parameter estimated from an equation exactly similar to (11.3.1) with

$$\hat{z}_i = (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)/\hat{\sigma}$$

is b = 0.5. As a result we have the following MMLE; $\sigma_1 = \sqrt{\{\psi'(0.5) + \psi'(1.0)\}\sigma^2} = 2.565\,\sigma$ is the standard deviation of GL(0.5, σ). Realize that for skew distributions, there is a bias correction in the LSE $\tilde{\theta}_0$ (equation 3.2.12).

LSE: $\tilde{\theta}_0 = 5.190$, $\tilde{\theta}_1 = 0.00682$, $s = 0.2002$, $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 53070.42$

MMLE: $\hat{\theta}_0 = 5.320$, $\hat{\theta}_1 = 0.00619$, $\hat{\sigma} = 0.0801$, $\sum_{i=1}^{n}\beta_i(x_{[i]} - \bar{x}_{[.]})^2 = 6074.09$,

$\hat{\sigma}_1 = 0.2054$;

$$SE(\tilde{\theta}_1) = \pm\frac{0.2002}{\sqrt{53070.42}} = \pm\, 0.00087$$

and as in Example 3.4,

$$SE(\hat{\theta}_1) = \pm\frac{0.0801}{\sqrt{1.5(6074.09)}} = \pm\, 0.00084$$

The standard error of $\hat{\theta}_1$ based on the minimum variance bound is

$$\pm\frac{\sqrt{5}(0.0801)}{\sqrt{53070.42}} = \pm\, 0.00078.$$

It is interesting to note that the standard error of the MMLE is, even for small sample sizes, only marginally bigger than that calculated by using the minimum variance bound.

Example 11.11: Hand et al. (1994, p.69) have an interesting data set. It has n = 30 observations on (X, Y): X denotes the average outside temperature in Celsius and Y the gas consumption (1000 cubic feet). The observations were taken over a period of 30 weeks after cavity-wall insulation.

A Q-Q plot of the deviants $W_i = y_i - \tilde{\theta}_1 x_i$ is given in Figure 5. The plot suggests normal $N(\mu, \sigma^2)$ as a plausible distribution. Since a few deviants in the middle wiggle around a straight line, it suggests a short-tailed symmetric distribution as well.

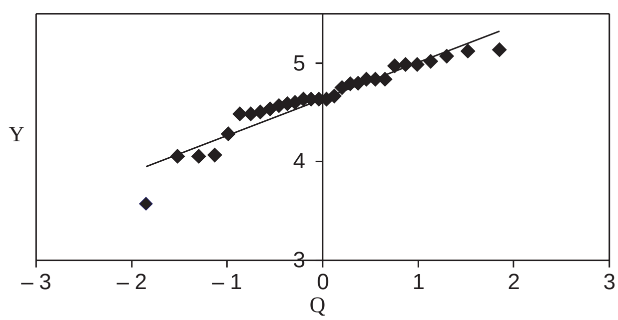

Figure 5 Gas consumption.

We consider the family (3.6.2). We usually choose r = 4. The reason is that with r = 4, the family represents a broad range of symmetric short-tailed distributions for different values of d and affords a great deal of flexibility in choosing an appropriate model. We calculate the values of lnL from equation (11.4.3). Realize that (3.6.2) reduces to normal if λ = 0. We now have the following values (r = 4):

d =	– 20	– 15	– 10	– 1.0	0	1.0
(1/n) ln L	– 0.324	– 0.325	– 0.326	– 0.359	– 0.385	– 0.450

The likelihood function L is almost flat around d = – 15. We choose r = 4 and d = – 15 (λ = 0.21) and have the following MMLE (n = 30).

LSE: $\tilde{\theta}_0 = 4.724$, $\tilde{\theta}_1 = -0.278$, $s = 0.355$ $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 198.53$

$$SE(\tilde{\theta}_1) = \pm \frac{0.355}{\sqrt{198.52}} = \pm 0.0252.$$

MMLE: $\hat{\theta}_0 = 4.726$, $\hat{\theta}_1 = -0.279$, $\hat{\sigma} = 0.319$, $\hat{\sigma}_1 = 1.113(0.319) = 0.355$,

$$SE(\hat{\theta}_1) = \pm \frac{0.319}{\sqrt{0.8073(198.53)}} = \pm 0.0252;$$

0.8073 is the value of D calculated from (11.4.4). It is indeed reassuring that the LSE and the MMLE and their standard errors are essentially equal to one another in situations when the underlying distribution is indistinguishable from a normal.

11.5. MULTIPLE LINEAR REGRESSION

As an extension of a simple linear regression model, we consider a few real life examples in the context of multiple linear regression and illustrate the advantages the modified likelihood methodology has over least squares (Islam et al., 2001; Tiku et al., 2001). Islam and Tiku (2004) study the MMLE, the LSE and the Huber M-estimators of the parameters in multiple linear regression models. They show that the former have considerable advantages. The formulae for computing the MMLE are given in (3.10.2) – (3.10.10) and are natural extensions of those for simple linear regression models; see Appendix 11C.

Example 11.12: The following data give the reaction rate for the catalytic isomerization of n-pentane to isopentane. The data is reproduced in Atkinson and Riani (2000, p. 298).

Run Number	Partial Pressures (psia) x_1	x_2	x_3	Rate y
1	205.8	90.9	37.1	3.541
2	404.8	92.9	36.3	2.397
3	209.7	174.9	49.4	6.694
4	401.6	187.2	44.9	4.722
5	224.9	92.7	116.3	0.593
6	402.6	102.2	128.9	0.268
7	212.7	186.9	134.4	2.797
8	406.2	192.6	134.9	2.451
9	133.3	140.8	87.6	3.196
10	470.9	144.2	86.9	2.021
11	300.0	68.3	81.7	0.896
12	301.6	214.6	101.7	5.084
13	297.3	142.2	10.5	5.686
14	314.0	146.7	157.1	1.193
15	305.7	142.0	86.0	2.648

16	300.1	143.7	90.2	3.303
17	305.4	141.1	87.4	3.054
18	305.2	141.5	87.0	3.302
19	300.1	83.0	66.4	1.271
20	106.6	209.6	33.0	11.648
21	417.2	83.9	32.9	2.002
22	251.0	294.4	41.5	9.604
23	250.3	148.0	14.7	7.754
24	145.1	291.0	50.2	11.590

By using EXCEL, the multiple linear regression model (3.10.1) is found to be quite reasonable. To explore the error distribution, we calculate the LSE $\tilde{\theta}_i$ (i = 1, 2, 3) and the deviants

$$w_i = y_i - \tilde{\theta}_1 x_{1i} - \tilde{\theta}_2 x_{2i} - \tilde{\theta}_3 x_{3i} \quad (1 \le i \le 24).$$

A Q-Q plot of the ordered deviants w(i) is given in Figure 6.

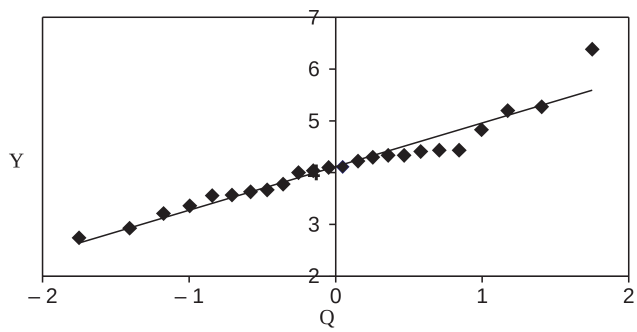

Figure 6 Reaction rates.

The plot clearly indicates a skew distribution with a long tail on the right hand side. One of the distributions in the Generalized Logistic family GL(b, σ) is appropriate. To determine the value of the shape parameter b, we calculate the MMLE of θ_0, θ_i (i = 1, 2, 3) and σ (Appendix 11C) for a given b. We calculate the values of $(x_{ij}' = x_{ij} - \bar{x}_j;\ 1 \le i \le 24,\ 1 \le j \le 3)$

$$\hat{z}_i = (y_i - \hat{\theta}_0 - \hat{\theta}_1 x'_{1i} - \hat{\theta}_2 x'_{2i} - \hat{\theta}_3 x'_{3i})/\hat{\sigma} \quad (1 \le i \le n),\ n = 24,$$

for a series of values of b (integer or half-integer). The maximum of (1/n) ln L is attained at b = 8. With this value of b, we have the following MMLE and the LSE and their standard errors obtained from equations given in Appendix 11C; σ_1 is the standard deviation of the distribution. Realize that

$$\tilde{\theta}_0 = \bar{y} - \{\psi(8) - \psi(1)\}\tilde{\sigma},\ \tilde{\sigma} = \tilde{\sigma}_1/1.3334;$$

see also Islam and Tiku (2004).

Parameter	MMLE	SE	LSE	SE
θ_0	1.943	0.716	2.418	0.921
θ_1	– 0.0065	0.0015	– 0.0089	0.0020
θ_2	0.0342	0.0024	0.0357	0.0032
θ_3	– 0.0361	0.0034	– 0.0386	0.0044
σ_1	0.781	0.074	0.850	0.175*

* Is the value of $(s_e/\sqrt{2n})\ \sqrt{(1+\lambda_4/2)}$, λ_4 = 2.063.

The MMLE are more efficient as expected. The M-estimation is not defined for skew distributions.

Example 11.13: Andrews (1974) considers an interesting data set, the Brownlee's stack loss data on the oxidation of ammonia. There are n = 21 observations $(y_i, x_{1i}, x_{2i}, x_{3i})$ $(1 \le i \le n)$, where

Y is the response and is 10 times the percentage of ammonia escaping up a stack,

X_1 is air flow, X_2 is temperature, and X_3 is acid concentration.

The data is also given in Aitkinson and Riani (2000, p.285).

A multiple linear regression model is appropriate (Andrews, 1974). A Q-Q plot of the residuals

$$e_i = y_i - \tilde{\theta}_0 - \tilde{\theta}_1 x_{1i} - \tilde{\theta}_2 x_{2i} - \tilde{\theta}_3 x_{3i} \quad (1 \le i \le 21)$$

is given in Andrews (1974, p.530). It is seen that the smallest residual $e_{(1)}$ which corresponds to the observation (15, 70, 20, 91) is grossly anomalous. This observation will have to be studied separately. We set aside this observation and calculate the LSE from the remaining 20 observations. A Q-Q plot of the new residuals indicates a long-tailed symmetric distribution. A distribution in the family (3.9.1) is appropriate. To find the most appropriate value of p, we calculate the MMLE of θ_0, θ_i (i = 1, 2, 3) and σ for a series of values of p. We calculate

$$\hat{z}_i = (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_{1i} - \hat{\theta}_2 x_{2i} - \hat{\theta}_3 x_{3i})/\hat{\sigma} \quad (1 \le i \le n)$$

and the corresponding values of (1/n) lnL. The value p = 2 maximizes it. The most plausible distribution, therefore, is the LTS (3.9.1) with p = 2. We calculate the MMLE and their standard errors from the equations given in Appendices 11B and 11C. They are given below together with the LSE and their standard errors; σ is the standard deviation.

Parameter	MMLE	SE	LSE	SE
θ_0	– 40.115	8.382	– 43.704	9.492
θ_1	0.912	0.105	0.889	0.119
θ_2	0.586	0.287	0.817	0.325
θ_3	– 0.113	0.110	– 0.107	0.125
σ	3.209	0.718	2.569	0.406*

*Is considerably larger since the value has to be multiplied by $\sqrt{(1 + \lambda_4/2)}$ and λ_4 is theoretically infinite.

The MMLE above and their standard errors are the same as those in Islam and Tiku (2004) in spite of the fact that they use different values of α_i^* and β_i^* $(1 \le i \le n)$ in their calculations while we use $\alpha_i^* = 0$ and $\beta_i^* = 1/\{1 + (1/k)t_{(i)}^2\}$ if C in the equation 11C.4 (Appendix) assumes a negative value. This is because C is positive for the data above with α_i and β_i calculated from the equations in (2.3.14).

Application of M-estimation results in the exclusion of three more observations besides the one mentioned earlier. These three observations are not grossly anomalous and excluding them is not advisable. In fact, Andrews (1974, p.530) states that "the probability plot of the residuals exhibits only slight anomalies" after excluding the observation (15, 70, 20, 91).

Example 11.14: The data is called Shortleaf pine data and consists of 70 observations (y_i, x_{1i}, x_{2i}), $1 \le i \le 70$, where Y represents the volume of the tree, X_1 its girth and X_2 its height. The data is reproduced in many books, e.g., Atkinson and Riani (2000, pp. 292-293). The interest is in determining y for given x_1 and x_2.

Screening the data through EXCEL, the linear model (n = 69)

$$y_i = \theta_0 + \theta_1 x_{1i} + \theta_{2i} x_{2i} + e_i \quad (1 \le i \le n)$$

is very appropriate when the observation (163.5, 23.4, 104) has been excluded. This observation is grossly anomalous and is set aside. To investigate the error distribution of the remaining 69 observations we construct a Q-Q plot. It is clear that a distribution is the STS family (3.6.2) is appropriate. We choose r = 4. The reason is that with r = 4, (3.6.2) represents a wide variety of short-tailed symmetric distributions with kurtosis decreasing from 3 to 1 as d increases. Realize that no symmetric distribution can have kurtosis less than 1 (Pearson and Tiku, 1970, p. 177).

The value of d that maximizes ln L is d = 1 (r = 4). The resulting estimates and their standard errors are given below. Note that the standard deviation of the chosen STS distribution is $\sigma_1 = 1.799\sigma$.

Parameter	MMLE	SE	LSE	SE
θ_0	– 45.123	3.302	– 42.254	3.982
θ_1	5.482	0.339	5.585	0.409
θ_2	0.265	0.075	0.194	0.090
σ_1	7.903	0.482	7.771	0.495*

* Is the value of $(s_e/\sqrt{2n})\sqrt{1+\lambda_4/2}$, $\sigma_4 = -0.882$.

The MMLE clearly are more precise.

The corresponding M-estimates are – 38.765, 5.115, 0.199 and 2.094. As usual, the M-estimate of σ has enormous downward bias.

11.6 AUTOREGRESSION

In this section, we give a few examples to illustrate the usefulness of the modified likelihood methodology to estimate the parameters in autoregressive models.

Example 11.15: Bass and Clarke (1972) give data to determine the effect of advertising on sales and state that the effect of advertising in one period carries over to the next. The data consists of 36 pairs of observations on (X, Y), where Y represents sales and X the advertising. The data is reproduced in Hand et al. (1994, p.83);

X:	15.0	16.0	18.0	27.0	21.0	49.0	21.0	22.0	28.0	36.0	40.0
Y:	12.0	20.5	21.0	15.5	15.3	23.5	24.5	21.3	23.5	28.0	24.0
X:	3.0	21.0	29.0	62.0	65.0	46.0	44.0	33.0	62.0	22.0	12.0
Y:	15.5	17.3	25.3	25.0	36.5	36.5	29.6	30.5	28.0	26.0	21.5
X:	24.0	3.0	5.0	14.0	36.0	40.0	49.0	7.0	52.0	65.0	17.0
Y:	19.7	19.0	16.0	20.7	26.5	30.6	32.3	29.5	28.3	31.3	32.2
X:	5.0	17.0	1.0								
Y:	26.4	23.4	16.4								

As recommended in Appendix 5A, we ignore the first pair (15.0, 12.0). The second pair (16.0, 20.5) is taken to be (x_0, y_0). We have n = 34 additional pairs of observations (x_i, y_i),

$1 \le i \le n$. We assume the autoregressive model (5.2.2) and obtain the following LSE (result of three iterations):

$$\tilde{\delta} = 0.0883 \quad \text{and} \quad \tilde{\phi} = 0.629.$$

We calculate the deviants

$$w_i = y_i - \tilde{\phi}\, y_{i-1} - \tilde{\delta}\,(x_i - \tilde{\phi}\, x_{i-1}), \quad 1 \le i \le 34,$$

and plot $w_{(i)}$ against the quantiles of a standard normal N(0, 1). The resulting Q-Q plot suggests the short-tailed symmetric distribution (5.8.1). We take r = 4 and find that value of d which maximizes lnL using, of course, the MMLE and the values of

$$\hat{z}_i = \{y_i - \hat{\phi} y_{i-1} - \hat{\mu} - \hat{\delta}(x_i - \hat{\phi} x_{i-1})\}/\hat{\sigma} \quad (1 \le i \le 34).$$

The maximum is attained at d = 0 (r = 4) as can be seen from the following values

d =	– 0.5	0	0.5
(1/n) ln L	– 2.768	– 2.765	– 2.767

With r = 4 and d = 0, the standard deviation of the distribution is $\sigma_1 = 1.605\sigma$.

The LSE and the MMLE are given below.

LSE: $\tilde{\delta} = 0.0883$, $\tilde{\phi} = 0.629$, $\tilde{\sigma}_1 = 4.098$, $\{\sum_{i=1}^{n} (u_i - \bar{u})^2\}_{\phi = \tilde{\phi}} = 11698.08$

MMLE: $\hat{\delta} = 0.102$, $\hat{\phi} = 0.639$, $\hat{\sigma} = 2.506$, $\hat{\sigma}_1 = (1.605)2.506 = 4.022$,
$\{\sum_{i=1}^{n} (u_i - \bar{u})^2\}_{\phi = \hat{\phi}} = 11763.61$.

Thus,

$$SE(\tilde{\delta}) = \pm \frac{4.098}{\sqrt{11698.08}} = \pm\, 0.0379$$

$$SE(\hat{\delta}) = \pm \frac{2.506}{\sqrt{0.4549(11763.61)}} = \pm\, 0.0343.$$

Q in (5.9.1) is exactly the same as D in (3.7.7) and (11.4.4). For r = 4 and d = 0, Q = 0.4549. It can be seen that as compared to the LSE, the MMLE indicate a somewhat stronger effect of advertising on the sales and a stronger carry over effect. In fact, T = 0.102/0.0343 = 2.97 is bigger than t = 0.0883/0.0379 = 2.33 and therefore gives a smaller probability for H_0: $\delta = 0$ to be true.

Example 11.16: A well known data set is the so called Gas Furnace data reported in many books, e.g., Box et al. (1994). It consists of 296 observations on (Y, X), where Y is %CO_2 in outlet gas and X is input gas rate. The data has time effect and is presumed to have the autoregressive model (5.2.2). The main interest is in evaluating the dependence of Y on X.

We ignore the first observation (– 0.109, 53.8); (x_0, y_0) is taken to be the second observation (0.000, 53.6). There are n = 294 additional observations (x_i, y_i), $1 \le i \le n$. We calculate the LSE of the parameters in (5.2.2). The LSE are

$$\tilde{\mu} = 0.915, \quad \tilde{\delta} = 0.793, \quad \tilde{\phi} = 0.983, \quad \tilde{\sigma} = 0.707, \quad \{\textstyle\sum_{i=1}^{n} (u_i - \bar{u})^2\}_{\phi = \tilde{\phi}} = 31.765$$

From the Q-Q plot of the deviants

$$\tilde{w}_i = y_i - \tilde{\phi}\, y_{i-1} - \tilde{\delta}\,(x_i - \tilde{\phi}\, x_{i-1}), \quad 1 \le i \le n,$$

a member of the STS family (5.8.1) seems to be most plausible. The maximization of (1/n) lnL yields r = 2 and d = – 1.5. The corresponding MMLE are

$$\hat{\mu} = 0.756, \quad \hat{\delta} = 0.736, \quad \hat{\phi} = 0.986, \quad \hat{\sigma} = 0.560, \quad \hat{\sigma}_1 = 1.2673(0.560) = 7.10,$$

$$Q = 0.6364, \{\textstyle\sum_{i=1}^{n} (u_i - \bar{u})^2\}_{\phi = \hat{\phi}} = 31.834.$$

The standard errors of the estimates of δ are

$$\text{LSE:} \qquad \pm \frac{0.707}{\sqrt{31.765}} = \pm 0.126$$

$$\text{MMLE:} \qquad \pm \frac{0.560}{\sqrt{0.6364(31.834)}} = \pm 0.124.$$

Again, it is reassuring that the MMLE and the LSE and their standard errors are essentially the same when the underlying distribution is indistinguishable from normal.

The estimated model based on the MMLE is

$$y_i - 0.986y_{i-1} = 0.756 + 0.736\,(x_i - 0.98x_{i-1}), \quad 1 \le i \le n.$$

11.7 EXPERIMENTAL DESIGN

In this section we illustrate how beautifully the methodology of modified likelihood adapts to multisample situations in the context of experimental design.

Example 11.17: Till (1974) gives sets of salinity values (parts per thousand) for three separate water masses. The data is reproduced in Hand et al. (1996, p. 201):

I	II	III
37.54	40.17	39.04
37.01	40.80	39.21
36.71	39.76	39.05
37.03	39.70	38.24
37.32	40.79	38.53
37.01	40.44	38.71
37.03	39.79	38.89
37.70	39.38	38.66
37.36		38.51
36.75		40.08
37.45		
38.85		

The Q-Q plots of the three data sets reveal that the observations 38.85 and 40.08 in I and III, respectively, are grossly anomalous. The Q-Q plot of the data set I, for example, is given in Figure 7.

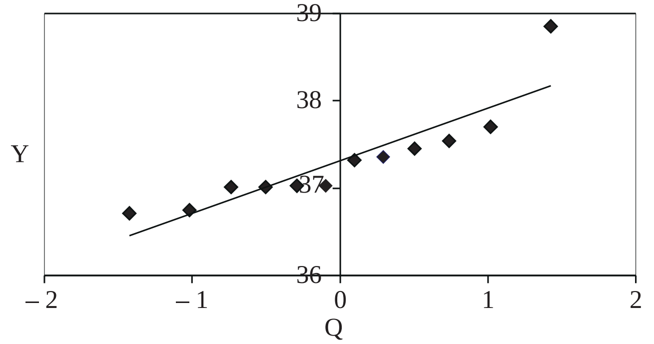

Figure 7 Salinity measurements.

These two observations need to be studied separately. We set them aside for a meaningful analysis of the data. The interest is in finding out if the groups differ in their mean values. The sample sizes are $n_1 = 11$, $n_2 = 8$ and $n_3 = 9$.

To locate the plausible distributions, we use Q-Q plots followed by the techniques of Section 11.3. The distribution (3.6.2) with r = 4 and d = 2 beautifully models all the three data sets. We now have the following estimates and their standard errors; $\sigma_1 = 2.1082\sigma$ is the standard deviation of the distribution and D = 0.6297 calculated from (11.4.4).

	LSE	MMLE
I:	$\bar{y}_1 = 37.174 \quad s_1 = 0.321$	$\hat{\mu}_1 = 37.189 \quad \hat{\sigma} = 0.155 \quad \hat{\sigma}_1 = 0.327$
	$SE(\bar{y}_1) = \pm \frac{0.325}{\sqrt{11}} = \pm 0.0968$	$SE(\hat{\mu}_1) \cong \pm \frac{0.155}{\sqrt{\{0.6297(11)\}}} = \pm 0.0589.$
II:	$\bar{y}_2 = 40.104 \quad s_2 = 0.531$	$\hat{\mu}_2 = 40.134 \quad \hat{\sigma} = 0.252 \quad \hat{\sigma}_2 = 0.531$
	$SE(\bar{y}_2) = \pm \frac{0.531}{\sqrt{8}} = \pm 0.188$	$SE(\hat{\mu}_2) \cong \pm \frac{0.252}{\sqrt{\{0.629(8)\}}} = \pm 0.112$
III:	$\bar{y}_3 = 38.760 \quad s_3 = 0.313$	$\hat{\mu}_3 = 38.764 \quad \hat{\sigma} = 0.512 \quad \hat{\sigma}_3 = 0.320$
	$SE(\bar{y}_3) = \pm \frac{0.313}{\sqrt{9}} = \pm 0.104$	$SE(\hat{\mu}_3) \cong \pm \frac{0.152}{\sqrt{\{0.629(9)\}}} = \pm 0.0638$

Although s_1^2 and s_2^2 are not much different from one another but they are much different from s_2^2 and, similary, for $\hat{\sigma}_i^2$ (i = 1, 2, 3). It is, therefore, advisable not to pool them.

Denote the population means of the three groups by μ_1, μ_2 and μ_3 respectively. To test the null hypothesis

$$H_0: \mu_1 = \mu_2 = \mu_3$$

we simply test the two orthogonal linear contrasts

$$H_{10}: \mu_1 - \mu_2 = 0 \quad \text{and} \quad H_{20}: \mu_1 + \mu_2 - 2\mu_3 = 0.$$

To test H_{10}, we have the following statistics based on the LSE and the MMLE:

$$|t| = \left| \frac{37.174 - 40.104}{\sqrt{\{(0.0968)^2 + (0.188)^2\}}} \right| = \frac{2.98}{0.211} = 13.89$$

$$|T| = \left| \frac{37.189 - 40.134}{\sqrt{\{(0.0589)^2 + (0.112)^2\}}} \right| = \frac{2.845}{0.1265} = 22.49.$$

The null distributions of t and T being standard normal (approximately), both t and T tests rejects H_{10}. The statistic T, however, gives a smaller probability for H_{10} to be true. Equivalently, the T test has higher power than the t test for a given significance level, in agreement with the results of Chapters 2 and 8.

To test H_{20}, we have

$$|t| = \left| \frac{37.174 + 40.104 - 77.52}{\sqrt{\{(0.0968)^2 + (0.188)^2 + 4(0.104)^2\}}} \right| = \frac{0.242}{0.2966} = 0.82$$

$$|T| = \left| \frac{37.189 + 40.134 - 77.528}{\sqrt{\{(0.0589)^2 + (0.112)^2 + 4(0.0638)^2\}}} \right| = \frac{0.205}{0.1797} = 1.14.$$

Both the tests do not reject H_{20}. The null hypothesis H_0, however, is rejected since H_{10} is rejected.

Example 11.18: Mendenhall and Beaver (1992, p. 349) give the following interesting data representing Insurance Claims in dollars submitted by five Health Groups:

Group I	II	III	IV	V
763	1335	596	3742	1632
4365	1262	1448	1833	5078
2144	217	1183	375	3010
1998	4100	3200	2010	671
5412	2948	630	743	2145
957	3210	942	867	4063
1286	867	1285	1233	1232
311	3744	128	1072	1456
863	1635	844	3105	2735
1499	643	1683	1767	767

If one assumes normality and homogeneity of variances, one can calculate the F statistic (Mendenhall and Beaver, p. 352):

$$F = \frac{1685638.62}{1720250.32} = 0.98$$

and conclude that the claims are not different from one another on the average.

A careful examination of the data, however, reveals that the normality assumption is not at all justified. The information provided by Q-Q plots supplemented by the determination of shape parameters as in Section 11.3 gives the most plausible distributions as follows.

Groups I and III: Generalized Logistic GL(4, σ).

Group II: Symmetric short-tailed with $r = 4$ and $d = 0$.

Group IV and V: Beta (a, b) with $a = 3$ and $b = 10$.

The computations of the MMLE for the Beta (a, b) are explained in Appendix 11A.

LSE — MMLE

I: $\bar{y}_1 = 1959.80 \quad s_1 = 1658.78$ — $\hat{\mu}_1 = 0.9391 \; \hat{\sigma} = 1071.286$

$$SE(\bar{y}_1) = \pm \frac{1658.78}{\sqrt{10}} = \pm 524.55 \qquad \hat{\theta}_1 = 0.9391 + 1.8333(1071.286) = 1964.93$$

$$\hat{\sigma}_1 = \sqrt{\{\psi'(4) + \psi'(1)\}}\;\hat{\sigma} = 1.3888(1071.286) = 1487.80$$

As in Example 11.5 (d),

$$SE(\hat{\theta}_1) \cong \pm 1071.286\sqrt{\frac{1.8565}{10}} = \pm 461.59$$

II: $\bar{y}_2 = 1996.10 \quad s_2 = 1383.98$ — $\hat{\mu}_2 = \hat{\theta}_2 = 2137.09 \quad \hat{\sigma} = 790.93$

$$SE(\bar{y}_2) = \pm \frac{1383.98}{\sqrt{10}} = \pm 437.65 \qquad \hat{\sigma}_2 = 1.6051(790.93) = 1269.52$$

$$SE(\hat{\theta}_2) \cong \pm \frac{790.93}{\sqrt{\{10(0.4549)\}}} = \pm 370.83$$

III: $\bar{y}_3 = 1193.90 \quad s_3 = 838.92$ $\qquad \hat{\theta}_3 = 183.331 + 1.8333(559.440) = 1208.95$

$$SE(\bar{y}_3) = \pm \frac{839.92}{\sqrt{10}} = \pm 265.29 \qquad \hat{\sigma}_3 = 1.3888(559.440) = 776.95$$

$$SE(\hat{\theta}_3) \cong 559.440\sqrt{\frac{1.8565}{10}} = 241.05$$

IV: $\bar{y}_4 = 1674.70 \quad s_4 = 1066.33$ $\qquad \hat{\mu}_4 = -217.231 \quad \hat{\sigma} = 8561.374$

$$SE(\bar{y}_4) = \pm \frac{1066.33}{\sqrt{10}} = \pm 337.20 \qquad \hat{\theta}_4 = -217.231 + \frac{3}{13}(8561.374) = 1758.47$$

$$\hat{\sigma}_4 = 0.1126(8561.374) = 964.01$$

$$SE(\hat{\theta}_4) = \pm 8561.374\sqrt{0.00121} = \pm 297.81$$

V: $\bar{y}_5 = 2278.90 \quad s_5 = 1446.88$ $\qquad \hat{\mu}_5 = -264.00 \quad \hat{\sigma} = 11524.69$

$$SE(\bar{y}_5) = \pm \frac{1446.88}{\sqrt{10}} = \pm 457.54 \qquad \hat{\theta}_5 = 2395.54 \quad \hat{\sigma}_5 = 1297.68$$

$$SE(\hat{\theta}_5) \cong 11524.69\sqrt{0.00121} = \pm 400.89.$$

For all the groups, the MMLE are numerically not much different from the corresponding LSE but have considerably smaller standard errors. It must be remembered that for non-normal distributions, the efficiencies of LSE relative to MMLE generally decrease as the sample size n increases (Chapters 2-8).

Since $t_i = \hat{y}_i/\sqrt{(SE)^2}$ and $T_i = \hat{\theta}_i/\sqrt{(SE)^2}$ (i = 1, 2, 3, 4, 5) all have the Student t distribution (approximately) with 9 degrees of freedom, the 95% confidence intervals are the following; $t_{0.025}(9) = 2.262$.

	Based on LSE	Based on MMLE
I:	(773.27, 3146.33)	(920.81, 3009.05)
II:	(1006.14, 2986.06)	(1298.27, 2975.91)
III:	(593.81, 1793.99)	(663.69, 1754.21)
IV:	(912.02, 2437.38)	(1084.82, 2432.12)
V :	(1244.04, 3313.76)	(1488.73, 3302.35)

All the confidence intervals based on the MMLE are shorter than the corresponding intervals based on the LSE.

To reject the null hypothesis H_0: that the Insurance Claims submitted by the five groups are the same on the average, it suffices to reject one linear contrast, e.g., Mean (V)-Mean (III) = 0. Now,

$$t = \frac{2278.90 - 1193.90}{\sqrt{\{(457.54)^2 + (265.29)^2\}}} = 2.05$$

and

$$T = \frac{2395.54 - 1208.95}{\sqrt{\{(400.89)^2 + (241.05)^2\}}} = 2.54;$$

the null distribution of t as well as T are standard normal (approximately). The upper 99 percent point of N(0, 1) being 2.326, the T test clearly rejects the linear contrast. Using the classical F test gives erroneous results. This is due to the fact that the data are not normal.

APPENDIX 11A

MMLE FOR THE BETA DISTRIBUTION

Let $y_1, y_2, \ldots, y_n$ be a random sample and write $z_i = (y_i - \mu)/\sigma$. Assuming that the distribution of z is beta(a, b),

$$f(z) \propto z^{a-1}(1-z)^{b-1}, \quad 0 < z < 1 \quad (a > 1, b > 1), \tag{11A.1}$$

the likelihood function is

$$L \propto \left(\frac{1}{\sigma}\right)^n \prod_{i=1}^{n} z_i^{a-1}(1-z)^{b-1}. \tag{11A.2}$$

The likelihood equations are expressions in terms of z_i^{-1} and $(1 - z_i)^{-1}$. They have no explicit solutions. Modified likelihood equations are obtained by expressing them in terms of the ordered variates $z_{(i)} = (y_{(i)} - \mu)/\sigma$ $(1 \le i \le n)$ and using the linear approximations

$$z_{(i)}^{-1} \cong \alpha_{1i} - \beta_{1i} z_{(i)} \quad \text{and} \quad (1 - z_{(i)})^{-1} \cong \alpha_{2i} + \beta_{2i} z_{(i)}.$$

The α_i and β_i coefficients are obtained as usual from the first two terms of Taylor series expansions:

$$\alpha_{1i} = 2t_{(i)}^{-1},\ \beta_{1i} = t_{(i)}^{-2}, \quad \beta_{2i} = (1 - t_{(i)})^{-2} \quad \text{and} \quad \alpha_{2i} = (1 - t_{(i)})^{-1} - \beta_{2i} t_{(i)} \tag{11A.3}$$

where

$$\frac{1}{\beta(a, b)} \int_0^{t_{(i)}} z^{a-1}(1-z)^{b-1}\, dz = \frac{i}{n+1}, \quad 1 \le i \le n;$$

IMSL subroutine in FORTRAN is available to evaluate $t_{(i)}$, $1 \le i \le n$.

The solutions of the modified likelihood equations are the MMLE:

$$\hat{\mu} = K - D\hat{\sigma} \quad \text{and} \quad \hat{\sigma} = \{-B + \sqrt{(B^2 + 4nC)}\}/2\sqrt{\{n(n-1)\}} \tag{11A.4}$$

where

$$K = (1/m)\textstyle\sum_{i=1}^{n} m_i y_{(i)} \quad \left(m = \sum_{i=1}^{n} m_i\right), \quad D = (1/m)\sum_{i=1}^{n} \delta_i,$$

$$B = \textstyle\sum_{i=1}^{n} \delta_i (y_{(i)} - K) \quad \text{and} \quad C = \sum_{i=1}^{n} m_i (y_{(i)} - K)^2 = \sum_{i=1}^{n} m_i y_{(i)}^2 - mK^2; \tag{11A.5}$$

$$m_i = (a-1)\beta_{1i} + (b-1)\beta_{2i} \quad \text{and} \quad \delta_i = (a-1)\alpha_{1i} - (b-1)\alpha_{2i}.$$

The MMLE of $\theta = E(Y)$ is

$$\hat{\theta} = \hat{\mu} + \frac{a}{a+b}\hat{\sigma} \tag{11A.6}$$

The inverse $V = I^{-1}$ of the Fisher information matrix is $(a > 2, b > 2)$,

$$V = \frac{\sigma^2}{2n(a+b-1)} \begin{bmatrix} \dfrac{a(a-2)}{(a+b-2)} & -(a-2) \\ -(a-2) & a+b-4 \end{bmatrix} \tag{11A.7}$$

Simulations reveal that 11A.7 gives close approximations to the variances and the covariance of $\hat{\theta}$ and $\hat{\sigma}$ for all sample sizes $n \ge 10$. For $a = b$, $\hat{\theta}$ and $\hat{\sigma}$ are uncorrelated.

The coefficients of skewness $\mu_3/\mu_2^{3/2}$ and kurtosis μ_4/μ_2^2 of the beta distribution beta(a, b) are given below; (1) = skewness, (2) = kurtosis:

a		b = 2.5	3	5	10	20
2.5	(1)	0	0.124	0.434	0.760	0.979
	(2)	2.250	2.315	2.684	3.424	4.147
3	(1)	– 0.124	0	0.310	0.638	0.860
	(2)	2.315	2.333	2.585	3.197	3.837
4	(1)	– 0.306	– 0.181	0.129	0.459	0.688
	(2)	2.494	2.444	2.523	2.945	3.461
5	(1)	– 0.434	– 0.310	0	0.333	0.566
	(2)	2.684	2.586	2.539	2.823	3.251
10	(1)	– 0.760	– 0.637	– 0.333	0	0.247
	(2)	3.406	3.190	2.818	2.746	2.901

Consider the data on X in Example 11.7. Assuming that the underlying distribution is beta (2.5, 6.5), we have the following estimates of $\theta = E(X)$.

LSE: $\bar{x} = 182.423$, $s = 24.749$, $SE(\bar{x}) = \pm 24.79/\sqrt{26} = \pm 4.854$.

MMLE: $\hat{\theta} = \hat{\mu} + (2.5/9)\hat{\sigma} = 183.447$, $\hat{\sigma} = 166.884$;

since the variance of beta (a, b) is $ab/(a + b)^2(a+b+1)$, the MMLE of the standard deviation is $\hat{\sigma}_1 = \sqrt{0.020\hat{\sigma}^2} = 23.64$.

The variance-covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is from 11A.7,

$$\frac{\hat{\sigma}^2}{16n}\begin{bmatrix} 0.1786 & -0.5 \\ -0.5 & 5 \end{bmatrix}.$$

This gives

$$V(\hat{\theta}) \cong \frac{1}{416}\{0.1786 + 5(0.0772) - (0.2778)\}\sigma^2 = 0.0007\sigma^2,$$

$$SE(\hat{\theta}) \cong \pm 0.026(166.884) = \pm 4.422.$$

The MMLE is clearly more efficient. Incidentally, the choice a = 2.5 and b = 6.5 maximizes lnL, θ and σ equated to $\hat{\theta}$ and $\hat{\sigma}$, respectively.

To verify how close the theoretical variance $0.00070\sigma^2$ is to the true variance, we simulated from 10,000 Monte Carlo runs the variance of $\hat{\theta}$, sample size n = 26. The simulated variance is $0.00076\sigma^2$. The two are very close to one another, in spite of the fact that $\mu < X < \mu + \sigma$, i.e., the range of X depends on the parameters we are estimating.

Linear Regression Model

Assuming that the error e_i in (11.4.2) have a beta (a, b) distribution, the MMLE are

$$\hat{\theta}_0 = \bar{y}_{[.]} - \hat{\theta}_1\bar{x}_{[.]} - (\Delta/m)\hat{\sigma}, \quad \hat{\theta}_1 = K - D\hat{\sigma} \tag{11A.8}$$

$$\hat{\sigma} = \{-B + \sqrt{(B^2 + 4nC)}\} / 2\sqrt{\{n(n-2)\}},$$

where

$$K = \sum_{i=1}^{n} m_i (x_{[i]} - \bar{x}_{[.]})y_{[i]} / \sum_{i=1}^{n} m_i (x_{[i]} - \bar{x}_{[.]})^2$$

$$D = \sum_{i=1}^{n} \delta_i (x_{[i]} - \bar{x}_{[.]}) / \sum_{i=1}^{n} m_i (x_{[i]} - \bar{x}_{[.]})^2$$

$$B = \sum_{i=1}^{n} \delta_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\} \quad \text{and}$$

$$C = \sum_{i=1}^{n} m_i \{y_{[i]} - \bar{y}_{[.]} - K(x_{[i]} - \bar{x}_{[.]})\}^2;$$

$\bar{y}_{[.]} = (1/m)\sum_{i=1}^{n} m_i y_{[i]}$, $\bar{x}_{[.]} = (1/m)\sum_{i=1}^{n} m_i \bar{x}_{[i]}$, $m = \sum_{i=1}^{n} m_i$, $\Delta = \sum_{i=1}^{n} \delta_i$ and m_i and δ_i are the same as in 11A.5.

The Fisher information matrix is ($a > 2$, $b > 2$)

$$I = \frac{n(a+b-1)(a+b-2)}{(a-2)(b-2)\sigma^2} \times \begin{bmatrix} a+b-4 & (a+b-4)(\sum_{i=1}^{n} x_i/n) & a-2 \\ (a+b-4)(\sum_{i=1}^{n} x_i/n) & (a+b-4)(\sum_{i=1}^{n} x_i^2/n) & (a-2)(\sum_{i=1}^{n} x_i/n) \\ a-2 & (a-2)(\sum_{i=1}^{n} x_i/n) & \dfrac{a(a-2)}{a+b-2} \end{bmatrix}$$

The asymptotic variances, in particular, are

$$V(\hat{\theta}_0) \cong \frac{a(a-2)^2\,(b-2)(a+b-4)}{n(a+b-1)(a+b-2)^2}\left[1 + \frac{2(b-2)}{a(a+b-4)}\frac{\bar{x}^2}{s_x^{\,2}}\right]\sigma^2$$

$$V(\hat{\theta}_1) \cong \frac{(a-2)(b-2)}{(a+b-1)(a+b-2)(a+b-4)}\,\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \qquad (11A.9)$$

and

$$V(\hat{\sigma}) \cong \frac{(a+b-4)}{(a+b-1)}\,\frac{\sigma^2}{2n};\ s_x^{\,2} = \sum_{i=1}^{n}(x_i-\bar{x})^2/n.$$

APPENDIX 11B

The equations for calculating the MMLE are given in Section 3.10. The Fisher information matrices are natural extensions of those in (3.7.7), (3.9.8) and (3.14.5); see also Islam and Tiku (2004). They are given below.

Long-tailed symmetric: For the family (3.9.1), the asymptotic variance-covariance matrix is $V = I^{-1}$ where

$$I = \frac{n}{\sigma^2}\,\frac{p(p-1/2)}{(p+1)(p-3/2)}$$

$$\times \begin{bmatrix} 1 & \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_q & 0 \\ \bar{x}_1 & \frac{1}{n}\sum_{i=1}^{n} x_{i1}^{\,2} & \frac{1}{n}\sum_{i=1}^{n} x_{i1}\,x_{i2} & \cdots & \frac{1}{n}\sum_{i=1}^{n} x_{i1}\,x_{iq} & 0 \\ \bar{x}_2 & \frac{1}{n}\sum_{i=1}^{n} x_{i2}\,x_{i1} & \frac{1}{n}\sum_{i=1}^{n} x_{i1}^{\,2} & \cdots & \frac{1}{n}\sum_{i=1}^{n} x_{i2}\,x_{iq} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \bar{x}_q & \frac{1}{n}\sum_{i=1}^{n} x_{iq}\,x_{i1} & \frac{1}{n}\sum_{i=1}^{n} x_{iq}\,x_{i2} & \cdots & \frac{1}{n}\sum_{i=1}^{n} x_{iq}^{\,2}\cdots & 0 \\ 0 & 0 & 0 & & 0 & \dfrac{2(p-3/2)}{p} \end{bmatrix} \qquad (11B.1)$$

Generalized Logistic: For GL(b, σ), the information matrix is the same as 11B.1 with the common multiplier p(p – 1/2)/(p + 1)(p – 3/2) replaced by b/(b+2), and the last row (column) replaced by

$$\left(\delta, \delta\bar{x}_1, \ldots\ldots, \delta\bar{x}_q, \frac{b+2}{b} + \gamma + \delta^2\right); \tag{11B.2}$$

$\delta = \psi(b + 1) - \psi(2)$ and $\gamma = \psi'(b + 1) + \psi'(2)$. The values of $\psi(b)$ and $\psi'(b)$ are given in Appendix 2D. If the design points are taken to be $x_{ij} - \bar{x}_j$ $(1 \le j \le k)$, then the last row (column) reduces to

$$\left(\delta, 0, \ldots\ldots, 0, \frac{b+2}{b} + \gamma + \delta^2\right) \tag{11B.3}$$

As a result, the MMLE $\hat{\theta}_1$ $(1 \le i \le k)$ are uncorrelated with $\hat{\sigma}$ at any rate for large n. This simplifies the distributions of the statistics used for testing the assumed values of θ_i $(1 \le i \le k)$. Also, it makes the computations easier.

Short-tailed symmetric: For the family (3.6.2), the information matrix is the same as 11B.1 with the common multiplier p(p – 1/2)/(p – 3/2)(p+1) replaced by D, and the last element 2(p – 3/2)/p replaced by D*/D. The expressions for D and D* are given in (3.7.8).

Least squares: For the LSE, the information matrix I is the same as 11B.1 with p(p – 1/2)/(p+1)(p – 3/2) replaced by 1, and the last element 2(p – 3/2)/p replaced by $2\left(1 + \frac{1}{2}\lambda_4\right)^{-1}$, $\lambda_4 = (\mu_4/\mu_2^2) - 3$. Realize that

$$V(\tilde{\sigma}) \cong \frac{\sigma^2}{2n}\left(1 + \frac{1}{2}\lambda_4\right). \tag{11B.4}$$

The estimated variance of $\tilde{\sigma}$ is obtained by replacing σ^2 by

$$s_e^2 = \sum_{i=1}^{n}(y_i - \tilde{\theta}_0 - \tilde{\theta}_1 x_{i1} - \ldots\ldots - \tilde{\theta}_q x_{iq})^2/(n - q - 1).$$

For long-tailed distributions $\lambda_4 > 0$. Consequently,

$$|SE(\hat{\sigma})| > s_e/\sqrt{2n}\,. \tag{11B.5}$$

APPENDIX 11C

Explicit expressions for the MMLE are given below.

LTS: For the LTS family (3.9.1), the MMLE are

$$\hat{\theta}_0 = \bar{y}_{[.]} - \sum_{j=1}^{k}\hat{\theta}_j\bar{x}_{[.]j}, \quad \hat{\boldsymbol{\theta}} = K + D\hat{\sigma} \quad \text{and} \tag{11C.1}$$

$$\hat{\sigma} = \{B + \sqrt{B^2 + 4nC}\}/2\sqrt{n(n-k-1)} \tag{11C.2}$$

where

$$\begin{aligned}
&\bar{y}_{[.]} = \sum_{i=1}^{n}\beta_i y_{[i]}/m, \quad \bar{x}_{[.]j} = \sum_{i=1}^{n}\beta_i x_{[i]j}/m \quad \left(m\sum_{i=1}^{n}\beta_i\right),\\
&\mathbf{K} = (X'\beta X)^{-1}(X'\beta Y) = (K_j)\\
&\mathbf{D} = (X'\beta X)^{-1}(X'\alpha 1) = (D_j)\\
&\beta = \mathrm{diag}(\beta_j),\ \alpha = \mathrm{diag}(\alpha_j)\\
&B = \frac{2p}{k}\sum_{i=1}^{n}\alpha_i\{y_{[i]} - \bar{y}_{[.]} - \sum_{i=1}^{n}K_j\,(x_{[i]j} - \bar{x}_{[.]j})\}
\end{aligned} \tag{11C.3}$$

and (11C.4)

$$C = \frac{2p}{k}\sum_{i=1}^{n}\beta_i\{y_{[i]} - \bar{y}_{[.]} - \sum_{i=1}^{n}K_j\,(x_{[i]j} - \bar{x}_{[.]j})\}^2.$$

The coefficients α_i and β_i are given in (2.3.14). If for a sample, C assumes a negative value, α_i and β_i are replaced by $\alpha_i^* = 0$ and $\beta_i^* = 1/\{1 + (1/k)t_{(i)}^2\}$, respectively.

GL(b, σ): For the Generalized Logistic, the MMLE are

$$\hat{\theta}_0 = \bar{y}_{[i]} - \sum_{j=1}^{k}\hat{\theta}_j\bar{x}_{[.]j} - (\Delta / m)\hat{\sigma}, \quad \hat{\boldsymbol{\theta}} = K + D\hat{\sigma} \tag{11C.5}$$

and

$$\hat{\sigma} = \{-B + \sqrt{B^2 + 4nC}\}/2\sqrt{n(n-q-1)}. \tag{11C.6}$$

Here,

$$\alpha_i = (1 + e^t + te^t)/(1 + e^t)^2, \quad \beta_i = e^t/(1 + e^t)^2 \quad (t = t_{(i)}), \tag{11C.7}$$

$$t_{(i)} = -\ln(q_i^{-1/b} - 1),\ q_i = i/(n+1),\ \Delta_i = \alpha_i - (b+1)^{-1} \quad \text{and} \quad \Delta = \sum_{i=1}^{n}\Delta_i;$$

$\bar{y}_{[.]}$, $\bar{x}_{[.]j}$ **K, D,** B and C are exactly the same as in (11C.3) – (11C.4) with the multiplier 2p/k (in B and C) replaced by b + 1. For skew distributions, it is prudent to work with the design points x_{ij} replaced by $x'_{ij} = x_{ij} - \bar{x}_j$ $(1 \le j \le k)$. That makes computations easy.

STS: For the STS family (3.6.2), the MMLE are

$$\hat{\theta}_0 = \bar{y}_{[.]} - \sum_{j=1}^{k}\hat{\theta}_j\bar{x}_{[.]j}, \quad \hat{\boldsymbol{\theta}} = \mathbf{K} - \lambda\mathbf{D}\hat{\sigma} \tag{11C.8}$$

and

$$\hat{\sigma} = \{-\lambda B + \sqrt{(\lambda B)^2 + 4nC}\}/2\sqrt{n(n-q-1)}; \tag{11C.9}$$

$\bar{y}_{[.]}$, $\bar{x}_{[.]j}$, **K, D,** B and C are exactly the same as in (11C.3) – (11C.4) with 2p/k (in B and C) replaced by 1. The coefficients α_i and β_i are given in (3.6.13) for $\lambda = r/(r-d) \le 1$ and in (3.7.5) for $\lambda > 1$.

Note: Realize that in 11C.3 above,

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} \cdots\cdots & x_{1k} \\ x_{21} & x_{22} \cdots\cdots & x_{2k} \\ \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} \cdots\cdots & x_{nk} \end{bmatrix}, \tag{11C.10}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ . \\ . \\ \theta_k \end{bmatrix} \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_k \end{bmatrix}.$$

Bibliography

Abramowitz, M. and Stegun, I. A. (1985). Handbook of Mathematical Functions. Dover: New York.

Abadir, K. M. (1993). On the asymptotic power of unit root tests. Econometric Theory 9, 189-221.

Abadir, K. M. (1995). The limiting distribution of the t ratio under a unit root. Econometric Theory 11, 775-793.

Agresti, A. (1996). Categorical Data analysis. John Wiley: New York.

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989). Statistical Modelling In Glim. Oxford Science: New York.

Akkaya, A. D. (1995). Stochastic modeling of earthquake occurrences and estimation of seismic hazard: A random field approach. Unpublished Ph. D thesis. Middle East Technical University: Ankara.

Akkaya, A. D. and Tiku, M. L. (2001a). Estimating parameters in autoregressive models in non-normal situations: asymmetric innovations. Commun. Stat.-Theory Meth. 30, 517-536.

Akkaya, A. D. and Tiku, M. L. (2001b). Corrigendum: Time series models with asymmetric innovations. Commun. Stat.-Theory Meth. 30, 2227-2230.

Akkaya, A. D. and Tiku, M. L. (2002a). Time series models for short-tailed distributions. (Submitted).

Akkaya, A. D. and Tiku, M. L. (2002b). Autoregressive models with short-tailed distributions. (Submitted).

Akkaya, A. D. and Tiku, M. L. (2003). Robust estimation and hypothesis testing under short-tailedness and inliers. To appear in TEST (2004).

Anderson, R. L. (1949). The problem of autocorrelation in regression analysis. J. Amer. Stat. Assoc. 44, 113-127.

Andrews, D. F. (1974). A robust method for multiple linear regression. Technometrics 16, 523-531.

Andrews, D. F., Bickel, P. J., Hampel, F., Huber, P., Rogers, W., and Tukey, J. W. (1972). Robust Estimation of Location: Survey and Advances. Princeton University press: Princeton, N. J.

Ariyawansa, K. A. and Templeton, J. G. C. (1984). Structural inference on the parameter of the Rayleigh distribution from doubly censored samples. Statist. Hefte 25, 181-199.

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). A First Course In Order Statistics. John Wiley: New York.

Atkinson, A. and Riani, M. (2000). Robust Diagnostic Regression Analysis. Springer, New York.

Bain, L. J. (1972). Inferences based on censored sampling from the Weibull or extreme value distribution. Technometrics 14, 693-702.

Bain, L. J. (1978). Statistical Analysis of Reliability and Life-Testing Models. Marcel Dekker: New York.

Bain, G. and Tiku, M. L. (1997a). Bayesian inference based on robust priors and MML estimators: part I, symmetric location-scale distributions. Statistics 29, 317-345.

Bain, G. and Tiku, M. L. (1997b). Bayesian inference based on robust priors and MML estimators: part II, skew location-scale distributions. Statistics 29, 81-99.

Balakrishnan, N. (1983). Empirical power study of a multi-sample test of exponentiality based on spacings. J. Stat. Comp. Simul. 18, 265-271.

Balakrishnan, N. (1984). Order statistics from the half-logistic distribution. J. Stat. Comp. Simul. 20, 287-309.

Balakrishnan, N. (1989a). Approximate maximum likelihood estimation of the mean and standard deviation of the normal distribution based on type II censored samples. J. Stat. Comput. Simul. 32, 137-148. RETRACTION: J. Stat. Comput. Simul. 43 (1992), 127.

Balakrishnan, N. (1989b). Approximate MLE of the scale parameter of the Rayleigh distribution. IEEE Trans. Reliability 38, 355-357. RETRACTION: IEEE, Trans. Reliability 41 (1992), 271.

Balakrishnan, N. and Leung, W. Y. (1988). Means, variances and covariances of order statistics, BLUE for the Type I generalized logistic distribution and some applications. Commun. Stat. -Simula. 17, 51-84.

Balakrishnan, N. and Chan, P. S. (1992a). Order statistics from extreme value distribution I: Tables of means, variances and covariances. Commun. Stat. Simul. Comput. 21, 1199-1217.

Balakrishnan, N. and Chan, P. S. (1992b). Order statistics from extreme value distribution II: Best linear unbiased estimators and some other uses. Commun. Stat. Simul. Comput. 4, 1219-1246.

Balakrishnan, N. and Ambagaspitiya, R. S. (1989). An empirical power comparison of three tests of exponentiality under mixture and outlier models. Biometrical. J. 31, 49-66.

Balakrishnan, N. and Cohen, A. C. (1990). Order Statistics and Inference: Estimation Methods. Academic Press: Boston.

Barlow, R. E. and Proshan, F. (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Winston: New York.

Barnett, V. D. (1966a). Evaluation of the maximum likelihood estimator when the likelihood equation has multiple roots. Biometrika 53, 151-165.

Barnett, V. D. (1966b). Order statistics estimators of the location of the Cauchy distribution. J. Amer. Stat. Assoc. 61, 1205-1218. Correction 63, 383-385.

Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. John Wiley: Chichester, England.

Barnett, V. and Lewis, T. (1994). Outliers in Statistical Data, 3 rd ed. John Wiley: New York.

Barton, D. E. and Dennis, K. E. (1952). The condition under which Gram Charlier and Edgeworth curves are positive definite and unimodal. Biometrika 39, 425-427.

Bartholomew, D. J. (1957). A problem in life testing. J. Amer. Stat. Assoc. 52, 350-355.

Bartlett, M. S. (1953). Approximate confidence intervals. Biometrika 40, 12-19.

Basu, A. P. and Ghosh, J. K. (1980). Asymptotic properties of a solution to the likelihood equation with life-testing applications. J. Amer. Stat. Assoc. 75, 410-414.

Bass, F. M. and Clarke, D. G. (1972). Testing distributed lag models of advertising effect. J. Marketing Research 9, 298-308.

Beach, C. M. and Mackinnon, J. G. (1978). A maximum likelihood procedure for regression with autocorrelated errors. Econometrica 46, 51-58.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. Technometrics 16, 147-186.

Benjamin, Y. (1983). Is the t test really conservative when parent population is long-tailed? J. Amer. Stat. Assoc. 78, 645-654.

Berkson, J. (1951). Why I prefer logits to probits. Biometrics 7, 327-339.

Bhattacharyya, G. K. (1985). The asymptotics of maximum likelihood and related estimators based on type II censored data. J. Amer. Stat. Assoc. 80, 398-404.

Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformation revisited. J. Amer. Stat. Assoc. 76, 296-311.

Billman, B., Antle, C. and Bain, L. J. (1972). Statistical inferences from censored Weibull samples. Technometrics 14, 831-840.

Birch, J. B. and Myers, R. H. (1982). Robust analysis of covariance. Biometrics 38, 699-713.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). J. R. Stat. Soc. B26, 211-252.

Box, G. E. P. and Tiao, G. C. (1964a). A Bayesian approach to the importance of assumptions applied to the comparison of variances. Biometrika 51, 153-167.

Box, G. E. P. and Tiao, G. C. (1964b). A note on criterion robustness and inference robustness. Biometrika 51, 169-173.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). Time Series Analysis: Forcasting And Control. Prentice Hall: New Jersey.

Bowman, K. O. and Shenton, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrika 62, 243-250.

Bowman, K. O. and Shenton, L. R. (1986). Moment ($\sqrt{b_1}$, b_2) techniques. In Goodness of Fit Techniques. (Eds., D'Agostino, R. B. and Stephens, M. A.) Marcel Dekker: New York.

Bradley, J. V. (1980). Nonrobustness in Z, t and F tests at large sample sizes. Bull. Psychonomic Soc. 16, 333-336.

Brown, M. B. (1975). Exploring interaction effects in the ANOVA. J. R. Stat. Soc. C24, 288-298.

Brownlee, K. A. (1965). Statistical Theory and Methodology in Science and Engineering, 2nd ed. John Wiley: New York.

Chakraborty, R. and Srinivasan, M. R. (1992). A modified "best maximum likelihood" estimator of linear regression with errors in both variables: An application for estimating Genetic Admixture. Biometrical J. 34, 567-576.

Chan, L. K. (1970). Linear estimation of the location and scale parameters of the Cauchy distribution based on sample quantiles. J. Amer. Stat. Assoc. 65, 851-859.

Chernoff, H. and Gastwirth, J. L. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. Ann. Math. Stat. 38, 52-72.

Cleveland, W. S. (1984). Graphical methods for data presentation. Amer. Statist. 38, 270-280.

Cochran, W. G. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. Ann. Math. Stat. 17, 164-177.

Cochran, W. G. (1977). Sampling Techniques, 3rd ed. John Wiley: New York.

Cogger, K. O. (1990). Robust Time Series Analysis-An L1 Approach. In Robust Regression. (Eds., Lawrence, K. D. and Arthur, J. L.). Marcel Dekker, Inc. New York.

Cohen, A. C. (1955). Censored samples from truncated normal distributions. Biometrika 42, 516-519.

Cohen, A. C. (1957). On the solution of estimating equations from truncated and censored samples from normal populations. Biometrika 44, 225-236.

Cohen, A. C. (1959. Simplified estimators for the normal distribution when samples are singly censored or truncated. Technometrics 1, 217-237.

Cohen, A. C. (1963). Progressively censored samples in life testing. Technometrics 5, 327-339.

Cohen, A. C. (1991). Truncated and Censored Samples, Theory and Applications. Marcel Dekker: New York.

Cohen, A. C. and Woodward, J. (1953). Tables of Pearson-Lee-Fisher functions of singly truncated normal distributions. Biometrics 9, 489-497.

Cohen, A. C. and Whitten, B. (1982). Modified maximum likelihood and modified moment estimators for the three-parameter Weibull distribution. Commun. Stat.-Theory Meth. 11, 2631-2656.

Collett, D. (1991). Modelling Binary Data. Chapman and Hall: London.

Cox, D. R. (1959). The analysis of exponentially distributed lifetimes with two types of failures. J. R. Stat. Soc. B21, 411-421.

Cox, D. R. (1961). Tests of separate families of hypotheses. Proc. Fourth Berkeley Symp. 1, 105-123.

Cox, D. R. (1962). Further results on tests of separate families of hypotheses. J. R. Stat. Soc. B24, 406-423.

Cox, D. R. and Lewis, P. A. W. (1966). The Statistical Analysis of Series of Events. Methues: London.

Cressie, N. A. C. (1980). Relaxing assumptions in the one-sample t-tests. Austral. J. Stat. 22, 143-153.

Cryer, J. D. (1986). Time Series Analysis. PWS-KENT Publishing Company: Boston.

Csörgö, M., Seshadri, V. and Yalovsky, M. (1973). Some exact tests for normality in the presence of unknown parameters. J. R. Stat. Soc. B35, 507-522.

Csörgö, M., Seshadri, V. and Yalovsky, M. (1975). Applications of characterizations in the area of goodness-of-fit. In Model Building and Model selection, Vol. 2: Statistical distributions in Scientific Work. (Eds., Patil, G. P., Kotz, S. and Ord, J. K.), 79-90. D. Reidel: Boston.

Cushney, A. R. and Peebles, A. R. (1905). The action of optimal isomers II, Hyocines. J. Physiology 32, 501-510.

D'Agostino, R. B. (1970). Transformations to normality of the null distribution of g_1. Biometrika 57, 679-680.

D'Agostino, R. B. (1971). On omnibus test of normality for moderate and large sample sizes. Biometrika 58, 341-348.

D'Agostino, R. B. and Tietjen, G. L. (1971). Simulation probability points of b_2 for small samples. Biometrika 58, 669-672.

D'Agostino, R. B. and Tietjen, G. L. (1973). Approach to the null distribution of b_1. Biometrika 60, 169-173.

D'Agostino, R. B. and Pearson, E. S. (1973). Tests for departure from normality: Empirical results for the distributions of b_1 and b_2. Biometrika 60, 613-622.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two level experiments. Technometrics 1, 311-342.

David, H. A. (1981). Order Statistics, 2nd ed. John Wiley: New York.

David, H. A. and Johnson, N. L. (1951). The effect of non-normality on the power function of the F-test in the analysis of variance. Biometrika 38, 43-57.

David, H. A. and Johnson, N. L. (1954). Statistical treatment of censored data, I. Fundamental formulae. Biometrika 41, 228-240.

David, H. A., Hartley, H. O. and Pearson, E. S. (1954). The distribution of the ratio, in a single normal sample, of range to standard deviation. Biometrika 41, 482-493.

David, H. A., Kennedy, W. J., and Knight, R. D. (1977). Means, variances and covariance of normal order statistics in the presence of an outlier. In Selected Tables in Mathematical Statistics, Vol. 5, 75-204. American Mathematical Society: Providence, R. I.

David, H. A. and Groeneveld, R. A. (1982). Measures of local variation in a distribution: Expected length of spacings and variances of order statistics. Biometrika 69, 227-232.

Davis, D. J. (1952). An analysis of some failure data. J. Amer. Stat. Assoc. 47, 113-150.

Davis, A. W. (1976). Statistical distributions in univariate and multivariate Edgeworth populations. Biometrika 63, 661-670.

Davis, A. W. and Rensick, S. (1986). Limit theory for the sample covariance and correlation functions of moving averages. Annals Stat. 14, 533-558.

Dempster, A. P. and Rosner, B. (1971). Detection of outliers. In Statistical Decision Theory and Related Topics. Academic Press: New York.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with unit root. J. Amer. Stat. Assoc. 74, 427-431.

Dixon, W. J. (1950). Analysis of extreme values. Annals Math. Stat. 21, 488-506.

Dixon, W. J. (1953). Precessing data for outliers. Biometrics 9, 74-89.

Dixon, W. J. (1960). Simplified estimation from censored normal samples. Annals Math. Stat. 31, 385-391.

Donaldson, T. S. (1968). Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. J. Amer. Stat. Assoc. 63, 660-667.

Downton, F. (1966). Linear estimates of the parameters in the extreme value distribution. Technometrics 8, 3-17.

Draper, N. R. and Smith, H. (1966). Applied Regression Analysis. John Wiley: New York.

Dudewicz, E. and Meulen, E. V. (1981). Entrophy based tests of uniformity. J. Amer. Stat. Assoc., 76, 967-974.

Dunnett, C. W. (1982). Robust multiple comparisons. Commun. Stat.-Theory. Meth. 11, 2611-2629.

Durbin, J. (1960). Estimation of parameters in time-series regression model. J. R. Stat. Soc. B22, 139-153.

Durbin, J. (1975). Kolmogorov-Simirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings. Biometrika 62, 5-22.

Dyer, D. and Harbin, M. S. (1981). An empirical power study of multi-sample tests for exponentiality. J. Stat. Comp. Simul. 12, 277-291.

Edgeworth, F. Y. (1887). The choice of means. Phylosophical Magazine 24, 268-271.

Elveback, L. R., Guillier, C. L. and Keating, F. R. (1970). Health, normality and the Ghost of Gauss. J. Amer. Medical Assoc. 211, 69-75.

Engelhardth, M. and Bain, L. J. (1974). Some results on point estimation for the two parameter Weibull or extreme-value distribution. Technometrics 16, 49-56.

Engelhardth, M. and Bain, L. J. (1979). Prediction limits and two-sample problems with complete or censored Weibull data. Technometrics 21, 233-237.

Epstein, B. (1958). Exponential Distribution and its role in life-testing. Industrial Quality Control 15, 4-9.

Epstein, B. (1960a). Tests for the validity of the assumption that the underlying distribution of life is exponential: Parts I and II. Technometrics 2, 81-101, 167-183.

Epstein, B. (1960b). Estimation of the parameters of two exponential distributions from censored samples. Technometrics 2, 403-406.

Epstein, B. and Sobel, M. (1953). Life-testing. J. Amer. Stat. Assoc. 48, 486-502.

Epstein, B. and Tsao, C. K. (1953). Some tests based on ordered observations from two exponential populations. Ann. Math. Stat. 24, 456-466.

Ferguson, T. S. (1961). On the rejection of outliers. Proc. Fourth Berkeley Symp. I, 253-287.

Filliben, J. J. (1975). The probability plot correlation coefficient test for normality. Technometrics 17, 111-117.

Finney, D. J. (1941). On the distribution of a variate whose logarithm is normally distributed. J. R. Stat. Soc. B7, 155-161.

Finney, D. J. (1947). The estimation from individuals records of the relationship between dose and quantal response. Biometrika 34, 320-334.

Fisher, R. A. (1931). The truncated normal distribution. British Association for the Advancement of Science. Math. Tables 5, 33-34: London.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179-184.

Fisher, R. A. (1966). The Design of Experimets, 8th ed. Oliver and Boyd: Edinburgh.

Fleming, T. R. and Harrington, D. P. (1980). A class of hypothesis tests for one and two sample censored survival data. School of Engineering and Applied Sciences, University of Virginia. DAMACS Tech. Rep. No., 80-89.

Frazer, D. A. S. (1976). Necessary analysis and adaptive inference (with discussion). J. Amer. Stat. Assoc. 71, 99-113.

Fuller, W. A. (1970). Simple estimators for the mean of skewed populations. Technical Report. Iowa State University: Iowa.

Gail, M. H. and Gastwirth, J. L. (1978). A scale-free goodness of fit tests for the experimental distribution based on the Gini statistic. J. R. Stat. Soc. B40, 350-357.

Gan, F. F. and Koehler, K. J. (1990). Goodness-of-fit tests based on P-P probability plots. Technometrics 32, 289-303.

Gayen, A. K. (1949). The distribution of Student's t in random samples of any size drawn from non-normal universe. Biometrika 36, 353-369.

Gayen, A. K. (1950). The distribution of the variance-ratio of random samples of any size drawn from non-normal universes. Biometrika 37, 236-255.

Geary, R. C. (1947). Testing for normality. Biometrika 34, 209-220.

Gehan, E. A. and Thomas, D. G. (1969). The performance of some two-sample tests in small samples with and without censoring. Biometrika 56, 127-132.

Godambe, V. P. (1955). A unified theory of sampling from finite populations. J. Roy. Stat. Soc. B17, 269-278.

Govindarajulu, Z. (1963). On moments of order statistics and quasi-ranges from normal populations. Annals Math. Stat. 34, 633-651.

Govindarajulu, Z. (1964). A supplement to Mendanhall's bibliography on life testing and related topics. J. Amer. Stat. Assoc. 59, 1231-1291.

Govindarajulu, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential population. J. Amer. Stat. Assoc. 61, 248-258.

Gross, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions. J. Amer. Stat. Assoc. 71, 409-416.

Gross, A. M. (1977). Confidence intervals for bisquare regression estimates. J. Amer. Stat. Assoc. 72, 341-354.

Gross, A. J. and Clark, V. A. (1975). Survival Distributions. John Wiley: New York.

Grubbs, F. E. (1950). Sample criteria for testing outlying observations. Annals Math. Stat. 21, 27-58.

Grubbs, F. E. (1969). Procedures for detecting outlying observations. Technometrics 11, 1-21.

Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. Biometrika 39, 260-273.

Gupta, S. S., Qureishi, A. S. and Shah, B. K. (1967). Best linear unbiased estimators of the parameters of the logistic distribution using order statistics. Technometrics 9, 43-56.

Halperin, M. (1952). Maximum likelihood estimation in truncated samples. Annals Math. Stat. 23, 226-238.

Hamilton, L. C. (1992). Regression With Graphics. California: Brooks/Cole Publishing Company: California.

Hamilton, J. D. (1994). Time Series Analysis. Princeton University Press: New Jersey.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J. and Ostrowski, E. (1994). Small Data Sets. Chapman & Hall: New York.

Harter, H. L. (1964). Exact confidence bounds, based on one order statistic, for the parameters of an exponential population. Technometrics 6, 301-317.

Harter, H. L. (1969). Order Statistics and Their Use in Testing and Estimation. U. S. Government Printing Office: Washington, D. C.

Harter, H. L. and Moore, A. H. (1966). Local maximum likelihood estimation of the parameters of three-parameter log-normal populations from complete and censored samples. J. Amer. Stat. Assoc. 61, 842-851.

Harter, H. L. and Moore, A. H. (1967). Maximum likelihood estimation from censored samples of the parameters of a logistic distribution. J. Amer. Stat. Assoc. 62, 675-684.

Harter, H. L. and Moore, A. H. (1968). Maximum likelihood estimation, from doubly censored samples, of the parameters of the first asymptotic distribution of extreme value. J. Amer. Stat. Assoc. 63, 889-901.

Harter, H. L. and Moore, A. H. (1976). An evaluation of exponential and Weibull test plans. IEEE Trans. Reliab. R25, 100-104.

Hawkins, D. M. (1977). Comment on “A new statistic for testing suspected outliers”. Commun. Stat. A6, 435-438.

Hawkins, D. M. (1979). Fractiles of an extended multiple outlier test. J. Stat. Comp. Simul. 8, 227-236.

Hawkins, D. M. (1980). Identification of Outliers. Chapman and Hall: London.

Hill, B. M. (1963). The three parameter lognormal distribution and Bayesian analysis of a point sourse epidemic. J. Amer. Stat. Assoc. 58, 72-84.

Hill, Sean, M. (1995). Regression and experimental design based on censored samples from long-tailed symmetric distributions. Unpublished Ph. D thesis. McMaster University: Ontario, Canada.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1983). Exploring Data Table, Trends, and Shapes. John Wiley: New York.

Hoeffding, W. (1953). On the distribution of the expected value of the order statistics. Annals Math. Stat. 24, 93-100.

Hogg, R. V. (1967). Some observations on robust estimation. J. Stat. Amer. Assoc. 62, 1179-1186.

Hogg, R. V. (1972). More light on the kurtosis and the related statistics. J. Stat. Amer. Assoc. 67, 422-424.

Hogg, R. V. (1974). Adaptive robust procedures: a partial review and some suggestions for future applications and theory. J. Amer. Stat. Assoc. 69, 909-927.

Hogg, R. V. (1982). On adaptive statistical inferences. Commun. Stat.-Theory. Meth. 11, 2531-2542.

Hogg, R. V., Fisher, D. M. and Randless, R. H. (1975). A two-sample adaptive distribution- free test. J. Amer. Stat. Assoc. 70, 656-661.

Hogg, R. V., and Craig, A. T. (1978). Introduction to Mathematical Statistics, 4th ed. McMillan: New York.

Hogg, R. V. Horn, P. S. and Lenth, R. V. (1984). On adaptive estimation. J. Stat. Plann. Inf. 9, 333-343.

Hosmer, D. W. and Lemeshow, S. (1989). Applied Logistic Regression. John Wiley: New York.

Huang, J. S. (1975). A note on order statistics from Pareto distribution. Scand. Actuar. J. 187-190.

Huber, P. J. (1964). Robust estimation of a location parameter. Annals Math. Stat. 35, 73-101.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. Annals Stat. 1, 799-821.

Huber, P. J. (1977). Robust Statistical Procedures. Regional Conference Series in Applied Mathematics No. 27, Soc. Industr. Appl. Math., Philadelphia.

Huber, P. J. (1981). Robust Statistics. John Wiley: New York.

Irwin, J. O. (1942). The distribution of the logarithm of survival times when the true law is exponential. J. Hyg. 42, 328-333.

Islam, M. Q., Tiku, M. L. and Yildirim, F. (2001). Non-normal Regression: Part I: Skew Distributions. Commun. Stat.-Theory Meth. 30, 993-1020.

Islam, M. Q., Tiku, M. L. (2004). Multiple linear regression model under non-normality. Commun. Stat.-Theory Meth. (to appear).

Ito, K. (1980). Robustness of ANOVA and MANOVA test procedures. In Handbook of Statistics, Vol. 1 (Ed., Krishnaiah, P. R.). North Holland: Amsterdam.

Jain, R. B. (1981). Percentage points of many outlier detectiom procedures. Technometrics 23, 71-75.

Jain, R. B. and Pingel, L. A. (1981). A procedure for estimating the number of outliers. Commun. Stat.-Theory Meth. A10, 1029-1041.

Jain, R. B. and Pingel, L. A. and Davidson, J. L. (1982). A unified approach for estimation and detection of outliers. Commun. Stat.-Theory Meth. A11, 2953-2976.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. Biometrika 36, 149-176.

Johnson, N. L. (1965). Tables to facilitate fitting S_u frequency curves. Biometrika 52, 547-558.

Johnson, N. L. (1974). Some optimality results for one or two sample procedures based on the smallest r order statistics. In Reliability and Biometry (Eds., Proschan, F. and Serfling, R. J.). SIAM: Philadelphia.

Johnson, N. L., Nixon, E., Amos, D. E. and Pearson, E. S. (1963). Table of percentage points of Pearson curves, for given $\sqrt{\beta_1}$ and β_2, expressed in standard measure. Biometrika 50, 459-498.

Johnson, R. C. and Johnson, N. L. (1979). Survival Models and Data Analysis. John Wiley: New York.

Joiner, B. L. and Rosenblatt, J. R. (1971). Some properties of the range in samples from Tukey's symmetric lambda distributions. J. Amer. Stat. Assoc. 66, 394-399.

Kambo, N. S. (1978). Maximum likelihood estimators of the location and scale parameters of the exponential distribution from a censored sample. Commun. Stat.-Theory Meth. A7, 1129-1132.

Kambo, N. S. and Awad, A. M. (1985). Testing equality of location parameters of exponential distributions. Commun. Stat.-Theory Meth. 14, 567-583.

Kanjo, A. J. (1993). Testing for new in better than used. Commun. Stat.-Theory Meth. 22, 311-321.

Karlin, S. (1966). A First course In stochastic Processes. Academic Press: New York.

Kendall, M. G. and Stuart, A. (1968). The Advanced Theory of Statistics, Vol. 1. McMillan: New York.

Kendall, M. G. and Stuart, A. (1969). The Advanced Theory of Statistics, Vol. 2. McMillan: New York.

Kendall, M. G. and Stuart, A. (1979). The Advanced Theory of Statistics, Vol. 3, McMillan: New York.

Khatri, C. G. (1981). Power of a test for location parameters of two exponential distributions. Aligarh. J. Stat. 1, 8-12.

Kleinbaum, D. G. (1994). Logistic Regression. Springer: New York.

Koul, H. L. (1978). Testing for new is better than used in expectation. Commun. Stat.-Theory Meth. 7, 685-701.

Kramer, W. (1980). Finite sample efficiency of ordinary least squares in the linear regression with autocorrelated errors. J. Amer. Stat. Assoc. 75, 1005-1009.

Kumar, S. and Patel, H. I. (1971). A test for a comparison of two exponential distributions. Technometrics 13, 183-189.

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35, 139-156.

Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modelling using the t distribution. J. Amer. Stat. Assoc. 84, 881-896.

Lawless, J. F. (1971). A prediction problem concerning samples from the exponential distribution, with applications in life-testing. Technometrics 13, 725-730.

Lawless, J. F. (1977). Prediction intervals for the two-parameter exponential distribution. Technometrics 19, 469-472.

Lawless, J. F. (1982). Statistical Models and Methods for Life Data. John Wiley: New York.

Lawless, J. F. and Singhal, K. (1980). Analysis of data from life-test experiments under exponential model. Naval. Res. Logist. Quart. 27, 323-334.

Lee, A. F. S. and D'Agostino, R. B. (1976). Levels of significance of some two- sample tests when observations are from compound normal distributions. Commun. Stat. A5, 342-352.

Lee, A. F. S. and Gurland, J. (1977). One-sample t-test when sampling from a mixture of normal distributions. Annals. Stat. 5, 803-807.

Lee, A. F. S., Kapadia, C. H., and Dwight, B. B. (1980). On estimating the scale parameter of the Rayleigh distribution from doubly censored samples. Statist. Hefte 21, 14-29.

Lehmann, E. L. (1959). Testing Statsitical Hypotheses. John Wiley: New York.

Lieblien, J. (1953). On the exact evaluation of the variances and covariances of order statistics in samples from the extreme value distribution. Annals Math. Stat. 24, 282-287.

Lilliefors, H. W. (1967). On the Kolmogorov-Simirnov test for normality with mean and variance unknown. J. Amer. Stat. Assoc. 62, 399-402.

Littell, C. R., McClave, J. R. and Offen, W. W. (1979). Goodness-of-fit tests for two-parameter Weibull distribution. Commun. Stat.-Simul. Comput. B8, 257-269.

Lloyd, E. H. (1952). Least squares estimators of location and scale parameters using order statistics. Biometrika 39, 88-95.

Locke, C. and Spurrier, J. D. (1977). The use of U-statistics for testing normality against alternatives with both tails heavy or both tails light. Biometrika 64, 638-640.

Lockhart, R. A., O'Reilly, F. J. and Stephens, M. A. (1986). Tests of fit based on normalized spacings. J. R. Stat. Soc., B, 48, 344-352.

Low, B. B. (1959). Mathematics. Neil and Company: Edinburgh.

MacGregor, G. A., Markandu, N. R., Roulston, J. E. and Jones, J. C. (1979). Essential hypertention: effect of an oral inhibitor of angiotensin-converting enzyme. British Medical J. 2, 1106-1109.

Magee, L., Ullah, A. and Srivastava, V. K. (1987). Efficiency of estimators in the regression model with first-order autoregressive errors. Specification Analysis in the Linear Model, Internat. Lib. Econom. Routledge and Kegan Paul, 81-98: London.

Malik, H. J. (1967). Exact moments of order statistics from a power-function distribution. Skand. Aktuarietidskr 7, 144-157.

Mendenhall, W. and Hader, R. J. (1958). Estimation of parameters of mixed exponential distributed failure times from censored life test data. Biometrika 45, 504-520.

Mendenhall, W. and Beaver, R. G. (1992). Business Statistics. Boston: PWS-KENT Publishing Company: Boston.

Maller, R. A. (1989). Regression with autoregressive errors-some asymptotic results. Statistics 20, 23-39.

Mann, N. R. (1969). Optimum estimators for linear function of location and scale parameters. Annals Math. Stat. 40, 2149-2155.

Mann, N. R. (1972). Design of over-stress life-test experiments: Optimum estimators for linear function of location and scale parametrs. Annals Math. Stat. 40, 2149-2155.

Mann, N. R. (1982). Optimal outlier tests for a Weibull model to identify process changes or to predict failure times. TIMS/Studies in the Management Sciences 19, 261-279. North Holland: Amsterdam.

Mann, N. R., Scheuer, E. M. and Fertig, K. W. (1973). A new goodness-of-fit test for the two-parameter Weibull or extreme-value distribution. Commun. Stat. 2, 383-400.

Mantel, N. and Myers, M. (1971). Problem of convergence of maximum likelihood iterative procedures in multiparameter situation. J. Amer. Stat. Assoc. 66, 484-491.

Mann, N. R. and Singapurwalla, N. D. (1983). Life testing. In Encyclopedia of Statistical Sciences, Vol. 4. New York: John Wiley.

Mardia, K. V. (1980). Tests of univariate and multivariate normality. Handbook of Statistics, vol. 1, 279-320. (Ed. Krishnaiah, P. R.). North Holland: Amsterdam.

Martin, R. D. and Yohai, V. J. (1985). Robustness in time series and estimating ARMA models. In Handbook of Statistics 5, 119-155. (Ed. Hannan, E. J.). North Holland: Amsterdam.

McCool, J. I. (1982). Censored data. In Encyclopedia of Statistical Sciences, Vol. 1. New York: John Wiley.

Mecker, W. Q. and Escobar, L. A. (1988). Statistical Methods for Reliability Data. John Wiley: New York.

Menon, M. V. (1963). Estimation of the shape and scale parameters of the Weibull distribution. Technometrics 5, 175-182.

Meyer, P. L. (1970). Introductory Probability and Statistical Applications. Addison-Wesley: Don Mills, Ontario.

Montgomery, D. C. and Peck, E. A. (1992). Introduction to Linear Regression Analysis. New York: John Wiley.

Mosteller, F. and Tukey, J. W. (1977). Data Analysis and Regression. Addison-Wesley: Reading, M. A.

Mulholland, H. P. (1977). On the distribution of b_1 for samples of size at most 25, with tables. Biometrika 64, 401-409.

Murani, A., Lavagnini, I. And Buttazzonl, C. (1986). Statistical study of air pollution concentrations via generalized gamma. J. Air Pollution Control Assoc. 36, 1050-1054.

Nagaraja, N. K. and Fuller, W. A. (1992). Least squares estimation of the linear model with autoregressive errors. New Directions in Time Series Analysis, Part I, IMA Vol. Math. Appl. 45, 215-225. Springer: New York.

Nelson, W. B. (1970). Statistical methods for accelerated life test data- the inverse power law model. General Electric Co. Tech. Rep. 71- C011. Schenectady: New York.

Nelson, W. B. (1972). Theory and applications of hazard plotting for censored failure data. Technometrics 14, 945-965.

Oral, E. (2002). Modified maximum likelihood estimators in binary regression. Unpublished Ph. D Thesis. Hacettepe University: Ankara.

Oral, E. and Tiku, M. L. (2003). Binary regression with stochastic covariates. Biometrical J. (Submitted).

O'Reilly, F. J. and Stephens, M. A. (1988). Transforming censored samples for testing fit. Technometrics 30, 79-86.

Patnaik, P. B. (1949). The noncentral χ^2 and F distributions and their applications. Biometrika 36, 202-232.

Pearl, R. (1940). Medical Biometry and Statistics. W. B. Saunders: Philedelphia.

Pearson, E. S. (1931). The analysis of variance in case of non-normal variation. Biometrika 23, 114-133.

Pearson, E. S. (1959). Note on an approximation to the distribution of non-central χ^2. Biometrika 46, 364.

Pearson, E. S. (1963). Some problems arising in approximating to probability distributions using moments. Biometrika 50, 95-112.

Pearson, E. S. and Adyanthaya, N. K. (1929). The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. Biometrika 21, 259-286.

Pearson, E. S. and Chandra Sekar, C. (1936). The efficiency of statistical tools and a criterion for the rejection of outlying observations. Biometrika 28, 308-320.

Pearson, E. S. and Hartley, H. O. (1972). Biometrika Tables for Statisticians, Vol. 2. Cambridge University Press: Cambridge.

Pearson, E. S. and Tiku, M. L. (1970). Some notes on the relationship between the distributions of central and non-central F. Biometrika 57, 175-179.

Pearson, K. and Lee, A. (1908). On the generalized probable error in multiple normal correlation. Biometrika 6, 69-86.

Perks, W. F. (1932). On some experiments in the graduation of mortality statistics. J. Inst. Actuar. 58, 12-57.

Pettitt, A. N. and Stephens, M. A. (1976). Modified Cramer-von Mises statistics for censored data. Biometrika 63, 291-298.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. Biometrika 75, 335-346.

Posten, H. O. (1978). The robustness of the two-sample t-test over the Pearson system. J. Stat. Comp. Simul. 6, 295-311.

Posten, H. O. (1982). Two-sample Wilcoxon power over the Pearson system and comparison with the t-test. J. Stat. Comp. Simul. 16, 1-18.

Posten, H. O., Yeh, H. C. and Owen, D. B. (1982). Robustness of the two-sample t-test under violations of the homogeneity of variance assumption. Commun. Stat.- Theory Meth. A11, 109-126.

Prescott, P. (1974). Variances and covariances of order statistics from the gamma distribution. Biometrika 61, 607-613.

Prescott, P. (1979). Critical values for a sequential test for many outliers. Appl. Statist. 28, 36-39.

Puthenpura, S. and Sinha, N. K. (1986). Modified maximum likelihood method for the robust estimation of system parametrs from very noisy data. Automatica 22, 231-235.

Pyke, R. (1965). Spacings (with discussion). J. R. Stat. Soc. B27, 395-439.

Rao, C. R. (1975). Linear Statistical Inference and its Applications. John Wiley: New York.

Rao, J. N. K. (1975). On the Foundation of Survey Sampling. A Survey of Statistical Designs and Linear Models. (Ed., Srivastava, J. N.). North Holland: Amsterdam.

Rasch, D. (1980). Probleme der angewanten Statistik, Robusttheit I: Fz. Fur Tierproduktion: Dummerstorf-Rostock, Heft 4.

Rasch, D. (1983). Robustness of sequential tests. In Robustness of Statistical Methods and Nonparametric Statistics (Eds., Rasch, D. and Tiku, M. L.). VEB Deutscher Verlag der Wissenschaften: East Berlin.

Reed, L. J. and Berkson, J. (1929). The application of the logistic function to experimental data. J. Physical Chemistry 33, 760-779.

Rey, W. J. J. (1983). Introduction to Robust and Quasi-Robust Statistical Methods. Springer-Verlag: New York.

Robinson, L. D., Dorroh, J. R., Lien, D. and Tiku, M. L. (1998). The effects of covariate adjustment in Generalized Linear Models. Commun. Stat.-Theory Meth., 27, 1653-1675.

Rosner, B. (1975). On the detection of many outliers. Technometrics 17, 221-227.

Rosner, B. (1977). Percentage points for the RTS many outlier procedure. Technometrics 19, 307-312.

Rousseeuw, P.J. and Leroy, A. M. (1987). Robust Regression and Outlier Detection. John Wiley: New York.

Roy, J. and Tiku, M. L. (1962). A Laguerre series approximation to the sampling distribution of the variance. Sankhya 24, 181-184.

Saleh, M. D., A. K. (1994). Book review of Order Statistics and Inference. Metrika 41, 307-309.

Samford, M. R. and Taylor, J. (1959). Censored observations in randomized block experiments. J. R. Stat. Soc. B21, 214-237.

Sansing, R. C. and Owen, D. B. (1974). The density of the t-statistic for non-normal distributions. Commun. Stat. 3, 139-155.

Sansone, G. (1959). Orthogonal Functions. Interscience Publishers: New York.

Sarhan, A. E. and Greenberg, B. G., eds. (1962). Contributions to Order Statistics. John Wiley: New York.

Sazak, S. (2003). Stochastic Regression. Unpublished Ph. D. thesis. Middle East Technical University: Ankara.

Schäffler, S. (1991). Maximum likelihood estimation for linear regression model with autoregressive errors. Statistics 22, 191-198.

Scheffé, H. (1959). The Analysis of Variance. John Wiley: New York.

Schneider, H. (1986). Truncated and Censored Samples from Normal Populations. Marcel Dekker: New York.

Schrader, R. M. and McKean, J. W. (1977). Robust analysis of variance. Commun. Stat-Theory Meth. A6, 879-894.

Seshadri, V., Csörgö, M. and Stephens, M. A. (1969). Testing for the exponential distribution using Kolmogorov-type statistics. J. R. Stat. Soc. B31, 499-509.

Şenoğlu, B. and Tiku, M. L. (2001). Analysis of variance in experimental design with non-normal error distributions. Commun. Stat.-Theory Meth. 30, 1335-1352.

Şenoğlu, B. and Tiku, M. L. (2002). Linear contrast in experimental design with non-identical error distributions. Biometrical J. 44, 359-374.

Şenoğlu, B. and Sürücü, B. (2004). Goodness-of-fit tests based on Kullback-Liebler information. IEEE Trans. Reliability. (To appear in the Sept. Issue).

Shah, B. K. (1966). On the bivariate moments of order statistics from a logistic distribution. Annals Math. Stat. 37, 1002-1010.

Shah, B. K. (1970). Note on moments of logistic order statistics. Annals Math.

Stat. 41, 2151-2152.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normallity (complete samples). Biometrika 52, 591-611.

Shapiro, S. S. and Wilk, M. B. (1972). An analysis of variance test for exponential distribution. Technometrics 14, 355-370.

Shetty, B. N. and Joshi, P. C. (1989). Likelihood ratio test for testing equality of location parameters of two exponential distributions from doubly censored samples. Commun. Stat.-Theory Meth. 18, 2063-2072.

Siddiqui, M. M. (1962). Some problems connected with Rayleigh distributions. J. Res. Nat. Bur. Stand. 660, 167-174.

Singh, C. (1972). Order Statistics from nonnormal populations. Biometrika 59, 229-233.

Singh, D., Singh, P. and Kumar, P. (1982). Handbook on Sampling Methods. Indian Agricultural Statistics Research Institute: New Delhi.

Smith, W. B., Zeis, C. D. and Syler, G. W. (1973). Three parameter lognormal estimation from censored data. J. Indian Stat. Assoc. 11, 15-31.

Smith, R. M. and Bian, L. J. (1976). Correlation type goodness-of-fit statistics with censored sampling. Commun. Stat. A5, 119-132.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. Biometrika 72, 67-90.

Spjϕtvoll, E. and Aastveit, A. H. (1980). Comparison of robust estimators on some data from field experiments. Scand. J. Stat. 7, 1-13.

Sprott, D. A. (1978). Robust and non-parametric procedures are not the only nor the safe alternatives to normality. Canadian J. Psychol. 32, 180-185.

Sprott, D. A. (1982). Robustness and maximum likelihood estimation. Commun. Stat.-Theory Meth. 11, 2513-2529.

Srinivasan, R. (1970). An approach to testing the goodness-of-fit of incompletely specified distributions. Biometrika 57, 605-611.

Srivastava, A. B. L. (1958). Efffect of non-normality on the power function of t-test. Biometrika 45, 421-429.

Srivastava, A. B. L. (1959). Efffect of non-normality on the power of the analysis of variance test. Biometrika 46, 114-122.

Stephens, M. A. (1974). EDF statistics for goodness-of-fit and some comparisons. J. Amer. Stat. Assoc. 69, 730-737.

Stephens, M. A. (1977). Goodness-of-fit for the extreme value distribution. Biometrika 64, 583-588.

Stephens, M. A. (1986). Tests based on EDF statistics. Goodness of Fit Techniques. (Eds., D'Agostino and Stephens). Marcel Dekker: New York.

Stuart, A. and Ord, J. K. (1987). Kendall's Advanced Theory of Statistics. Vol. 1, 5th ed. Oxford University Press: New York.

Subrahmaniam, K., Subrahmanian, K. and Messeri, J. Y. (1975). On the robustness of some tests of significance in sampling from a compound normal population. J. Amer. Stat. Assoc. 70, 435-438.

Sundrum, R. M. (1954). On the relation between estimating efficiency and the power of tests. Biometrika 41, 542-548.

Sürücü, B. (2002). Goodness-of-fit tests and outlier detection. Unpublished Ph. D thesis. Middle East Technical University: Ankara.

Sürücü, B. (2003). A power comparison of goodness-of-fit tests. J. Comput. Simul. (Submitted).

Szegö, G. (1959). Orthogonal Polynomials. American Math. Soc. Colloguium Publications: New York.

Tallis, G. M., and Light, R. (1968). The use of fractional moments for estimating the parameters of a mixed exponential distribution. Technometrics 10, 161-175.

Taylor, J. (1973). The analysis of design experiments with censored observations. Biometrics 29, 35-43.

Tan, W. Y. (1977). On the distribution of quadratic forms in normal random variables. Canadian J. Statist. 5, 241-250.

Tan, W. Y. (1982a). Sampling distributions and robustness of t, F and variance- ratio in two samples and ANOVA models with respect to departure from normality. Commun. Stat. Theory Meth. 11, 2485-2511.

Tan, W. Y. (1982b) . On approximating probability distributions of a statistic of Brown-Forsyth from normal and nonnormal universes. J. Stat. Comput. Simul. 16, 35-56.

Tan, W. Y. (1985) . On Tiku's robust procedure-a Bayesian insight. J. Stat. Plann. Inf. 11, 329-340.

Tan, W. Y. and Wong, S. P. (1977) . On the Roy-Tiku approximation to the distribution of sample variances from non-normal universes. J. Amer. Stat. Assoc. 72, 875-881.

Tan, W. Y. and Wong, S. P. (1980) . On approximating the null and non-normal distributions of the F-ratios in unbalanced random effects models from non-normal universe. J. Amer. Stat. Assoc. 75, 655-662.

Tan, W. Y. and Balakhrishnan, N. (1986). Bayesian insight into Tiku's robust procedure based on asymmetric censored samples. J. Stat. Comput. Simul. 24, 17-31.

Tan, W. Y. and Lin, V. (1993). Some robust procedures for estimating parameters in an autoregressive model. Sankya B, 55, 415-435.

Tan, W. Y. and Tiku, M. L. (1999). Sampling Distributions in terms of Laguerre Polynomials with Applications. New Age International (Formerly, Wiley Eastern): New Delhi.

Taufer, E. (2002). On entrophy based tests for exponentiality. Commun. Stat.-Simula. 31, 189-200.

Taylor, J. (1973). The analysis of designed experiments with censored observations. Biometrics 29, 35-43.

Thode, H. C. (2002). Testing For Normality. Marcel Dekker: New York.

Thode, H. C., Liu, H-K. and Finch, S. J. (1983). Power of tests of normality for detecting scale contaminated normal samples. Commun. Stat.-Simul. Comput., 12, 675-695.

Thomas, D. R., Bain, L. J. and Antle, C. E. (1969). Inferences on the parameters of the Weibull distribution. Technometrics 11, 445-460.

Tiao, G. C. and Tan, W. Y. (1966). Bayesian analysis of random effect models in the analysis of variance, II: effect of autocorrelated errors.Biometrika 53, 477-495.

Tietjen, G. L. (1986). The analysis and detection of outliers. Goodness-of-Fit Techniques. (Eds., D'Agostino and Stephens). Marcel Dekker: New York.

Tietjen, G. L. and Moore, R. H. (1972). Some Grubbs-type statistics for the detection of several outliers. Technometrics 14, 583-597.

Tietjen, G. L., Kahaner, D. K. and Beckman, R. J. (1977). Variances and covariances of the normal order statistics for sample size 2 to 50. Selected Tables in Mathematical Statistics, Vol. 5, 1-73. American Mathematical Society. Providence, RI.

Till, R. (1974). Statistical Methods For the Earth Scientist. MacMillan: London

Tiku, M. L. (1963). A Laguerre product series approximation to one-way classification variance-ratio distributon. J. Ind. Soc. Agric. Stat. 15, 223-231.

Tiku, M. L. (1964). Approximating the general non-normal variance-ratio sampling distributions. Biometrika 51, 83-95.

Tiku, M. L. (1965). Laguerre series forms of non-central χ^2 and F distributions. Biometrika 52, 415-427.

Tiku, M. L. (1966a). Distribution of the derivative of the likelihood function. Nature 210, No. 5037, 766.

Tiku, M. L. (1966b). Usefulness of three-moment χ^2 and t approximations. J.Ind. Soc. Agric. Stat. 18, 4-16.

Tiku, M. L. (1967a). Estimating the mean and standard deviation from a censored normal sample. Biometrika 54, 155-165.

Tiku, M. L. (1967b). A note on estimating the location and scale parameters of the exponential distribution from a censored sample. Austral. J. Statist. 9, 49-54.

Tiku, M. L. (1968a). Estimating the parameters of log-normal distribution from censored samples. J. Amer. Stat. Assoc. 63, 134-140.

Tiku, M. L. (1968b). Estimating the parameters of normal and logistic distributions from censored samples. Austral. J. Statist. 10, 64-74.

Tiku, M. L. (1968c). Estimating the mean and standard deviation from progressively censored normal samples. J. Ind. Soc. Agric. Stat. 20, 20-25.

Tiku, M. L. (1970). Monte Carlo study of some simple estimators in censored normal samples. Biometrika 57, 207-210.

Tiku, M. L. (1971a). Student's t distribution under non-normal situations. Australian J. Stat. 13, 142-148.

Tiku, M. L. (1971b). Power function of the F-test under non-normal situations. J. Amer. Stat. Assoc. 66, 913-916.

Tiku, M. L. (1973). Testing group effects from Type II censored normal samples in experimental design. Biometrics 29, 25-33.

Tiku, M. L. (1974a). A new statistic for testing for normality. Commun. Stat. 3, 223-232.

Tiku, M. L. (1974b). Testing normality and exponentiality in multi-sample situations. Commun. Stat. 3, 777-794.

Tiku, M. L. (1975a). Laguerre series forms of the distributions of classical test statistics and their robustness in non-normal situations. In Applied Statistics (Ed., Gupta, R. P). Elsevier: New York.

Tiku, M. L. (1975b). A new statistic for testing an assumed distribution. Statistical Distributions in Scientific Work, Vol. 2, 113-124 (Eds. Patil, G. P., Kotz, S. and Ord, J. K.). D. Reidel Publishing Company: New York.

Tiku, M. L. (1975c). A new statistic for testing suspected outliers. Commun. Stat. 4, 737-752.

Tiku, M. L. (1977). Rejoinder: ”Comment on “A new statistic for testing suspected outliers”. Commun. Stat.-Theory Meth. A6, 1417-1422.

Tiku, M. L. (1978). Linear regression model with censored observations. Commun. Stat.-Theory Meth. A7, 1219-1232.

Tiku, M. L. (1980a). Robustness of MML estimators based on censored samples and robust test statistics. J. Stat. Plann. Inf. 4, 123-143.

Tiku, M. L. (1980b). Goodness of fit statistics based on the spacings of complete or censored samples. Austral. J. Statist. 22, 260-275.

Tiku, M. L. (1981a). A goodness of fit statistic based on the sample spacings for testing a symmetric distribution against symmetric alternatives. Austral. J. Statist. 23, 149-158.

Tiku, M. L. (1981b). Testing equality of location parameters of two exponential distributions. Aligarh J. Statist. 1, 1-7. (Invited paper)

Tiku, M. L. (1982). Testing linear contrasts of means in experimental design without assuming normality and homogeneity of variances. Biometrical J. 24, 613-627. (Invited paper).

Tiku, M. L. (1983). Exact efficiencies of some robust estimators in sample survey. Commun. Stat.-Theory Meth. 12, 2043-2051.

Tiku, M. L. (1985a). Noncentral chi-square distribution. Encyclopedia of Statistical Sciences, Vol. 6. (Eds., Johnson, N. L. and Kotz, S.). John Wiley: New York.

Tiku, M. L. (1985b). Noncentral F-distribution. Encyclopedia of Statistical Sciences, Vol. 6. (Eds., Johnson, N. L. and Kotz, S.). John Wiley: New York.

Tiku, M. L.(1988). Order statistics in goodness-of-fit tests. Commun. Stat.-Theory Meth. 17, 2369-2387.

Tiku, M. L.(1996). Robust estimation and testing the mean vector. Proc. Sixth Lucacs Symp., pp. 263-272. Bowling Green State Univ. (Eds., Girko V. L and Gupta, A. K).

Tiku, M. L.and Jones, P. W. (1971). Best linear unbiased estimators for a distribution similar to the logistic. In proceedings “Statistics 71 Canada”. (Ed., Carter et al.), pp. 412-419.

Tiku, M. L., Rai, K. and Mead, E. (1974). A new statistic for testing exponentiality. Commun. Stat. 3, 485-493.

Tiku, M. L.and Stewart, D. E. (1977). Estimating and testing group effects from Type I censored normal samples in experimental design. Commun. Stat.-Theory Meth. A6, 1485-1501.

Tiku, M. L.and Yip, D. Y. N. (1978). A four-moment aproximation based on the F distribution. Austral. J. Statist. 20, 257-261.

Tiku, M. L.and Tamhankar, M. V. (1980). Improved tests of exponentiality in single- and multi-sample situations. J. Ind. Soc. Agric. Stat. 22, 40-50.

Tiku, M. L.and Singh, M. (1981a). Testing outliers in multivariate data. Statistical Distributions in Scientific Work, Vol. 5, 203-218. (Eds. Taillie, C. et al.). D. Reidel Publishing Company: New York.

Tiku, M. L. and Singh, M. (1981b). Testing the two parameter Weibull distribution. Commun. Stat.-Theory Meth. A10, 907-918.

Tiku, M.L.and Kumra, S. (1981). Expected values and variances and covariances of order statistics for a family of symmetric distributions (Student's t). In Selected Tables in Mathematical Statistics, Vol. 8. American Mathematical society: Providence, RI; 1985, 141-270.

Tiku, M. L., Tan, W. Y. and Balakrihnan, N. (1986). Robust Inference. Marcel Dekker: New York.

Tiku, M. L.and Vaughan, D. C. (1991). Testing equality of location parameters of two exponential distributions from censored samples. Commun. Statist.- Theory Meth. 20, 929-944.

Tiku, M. L.and Suresh, R. P. (1992). A new method of estimation for location and scale parameters. J. Stat. Plann. Inf. 30, 281-292.

Tiku, M. L.and Kambo, N. S. (1992). Estimation and hypoyhesis testing for a new family of bivariate non-normal distributions. Commun. Stat.-Theory Meth. 21, 1683-1705.

Tiku, M. L.and Vellaisamy, P. (1996). Improving efficiency of survey sample procedures through order statistics. J. Ind. Soc. Agric. Stat. 49 (Golden Jubilee Number), 363-385.

Tiku, M. L.and Vaughan, D. C. (1997). Logistic and nonlogistic density functions in binary regression with nonstochastic covariates. Biometrical J. 39, 883-898.

Tiku, M. L.and Wong, W. K. (1998). Testing for a unit root in an AR(1) model using three and four moment approximations. Commun. Stat.-Simula. 27, 185-198.

Tiku, M. L., Wong, W. K. and Bian, G. (1999). Time series models with asymmetric innovations. Commun. Stat.-Theory Meth. 28, 1131-1160.

Tiku, M. L.and Vaughan, D. C. (1999). A family of short-tailed symmetric distributions. Technical Report: McMaster University, Canada.

Tiku, M. L., Wong, W. K., Vaughan, D. C. and Bian, G. (2000). Time series models in non-normal situations: symmetric innovations. J. Time Series Analysis 21, 571-596.

Tiku, M. L., Islam, M. Q. and Selçuk, A. (2001). Nonnormal regression. II. Symmetric Distributons. Commun. Stat.-Theory Meth. 30, 1021-1045.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley: Reading, Mass.

Turker, O. (2002). Autoregressive Models: Statistical Inference and Applications. Unpublished Ph. D thesis: Middle East Technical University: Ankara.

Turker, O. and Akkaya, A.D. (2004). Multiple autoregressive model. (To be published).

Vasicek, O. (1976). A test for normality based on sample entropy. J. R. Stat. Soc. B38, 54-59.

Vaughan, D. C. (1992a). On the Tiku-Suresh method of estimation. Commun. Stat.-Theory Meth. 21, 451-469.

Vaughan, D. C. (1992b). Expected values, variances and covariances of order statistics for Student's t distribution with two degrees of freedom. Commun. Stat. Simula. 21, 391-404.

Vaughan, D. C. (1994). The exact values of the expected values, variances and covariances of the order statistics from the Cauchy distribution. J. Stat. Comput. Simul. 49, 21-32.

Vaughan, D. C. (2002). The Generalized Secant Hyperbolic distribution and its properties. Commun. Stat.- Theory Meth. 31, 219-238.

Vaughan, D. C. and Tiku, M. L. (1993). Testing the equality of location parameters of exponential populations from censored samples. Commun. Stat. -Theory Meth. 22, 2567-2581.

Vaughan, D. C. and Tiku, M. L. (2000). Estimation and hypothesis testing for a non-normal bivariate distribution with applications. J. Mathematical and Computer Modelling 32, 53-67.

Velu, R. and Gregory, C. (1987). Reduced rank regression with autoregressive errors. Econometrics 35, 317-335.

Verrill, S. and Johnson, R. A. (1987). The asymptotic equivalence of some modified Shapiro-Wilk statistics - complete and censored sample cases. Annals of Statistics 15, 413-419.

Vinod, H. D. and Shenton, L. R. (1996). Exact moment for autoregressive and random walk models for a zero or stationary initial value. Econometric Theory 12, 481-499.

Wang, Y. H. and Chang, S. A. (1977). A new approach to nonparametric tests of exponential distribution with unknown parameters. In The Theory and Applications of Reliablity, Vol. 1. (Eds., Tsokos, C. P. and Shimi, I. N.). Academic Press: New York.
Weisberg, S. (1980). Applied Linear Regression. John Wiley: New York.

Weiss, G. (1990). Least absolute error estimation in the presence of serial correlation. Journal of Econometrics 44, 127-158.

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. J. Amer. Stat. Assoc. 73, 812-815.

White, J. S. (1969). The moments of the log-Weibull order statistics. Technometrics 11, 373-386.

Wilks, S. S. (1962). Mathematical Statistics. John Wiley: New York.

Wilks, S. S. (1963). Multivariate statistical outliers. Sankhya A25, 407-426.

Wolynetz, M. S. (1974). Analysis of type I censored normally distributed data. Unpublished Ph. D thesis. University of Waterloo: Ontario, Canada.

Wright, F. T., Engelhardt, M. and Bian, L. J. (1978). Inferences for the two-parameter exponential distribution under type I censored sampling. J. Amer. Stat. Assoc. 73, 650-655.

Zacks, S. (1971). The Theory of Statistical Inference. John Wiley: New York.

Zhoo, L. C. and Zhang, H. (2001). A central limit theorem for testing exponentiality. Commun. Stat.-Theory Meth. 30, 1163-1170.

Index

A

B

C

D

E

F

G

H

I

K

L

M

O

P

Q

R

S

T

U

V

W