

Если исходы  $(\omega^1, \dots, \omega^N) = \varepsilon$  имеют распределение  $P_\theta^N(\cdot)$ , то для  $\theta$ -приведенной функции

$$E_\theta g^{(N)}(\varepsilon; t; \theta) = E_\theta [\ln p(\omega; t) - \ln p(\omega; \theta)] = -I(P_t: P_\theta). \quad (11.24)$$

Если информационный диаметр  $I(\mathcal{F})$  конечен, т. е. конечны все информационные уклонения  $I(P_t: P_\theta)$ , то по закону больших чисел

$$g^{(N)}(\varepsilon; t, \theta) \Rightarrow -I(P_t: P_\theta) < 0 = I(P_\theta: P_\theta).$$

Естественно ожидать, что вдали от «истинного» значения  $\theta$  параметра с большой вероятностью приведенная логарифмическая функция правдоподобия будет равномерно отрицательной, так что максимум правдоподобия лежит близ  $\theta$  и описывается ближайшим корнем системы (11.23). Ожидаемая картина остается справедливой, но локально, в канонической окрестности  $S_p(\theta)$ . Чтобы не возникало больших флуктуаций случайного поля  $g^{(N)}(\varepsilon; t; \theta)$  вдали от  $\theta$ , нужны дополнительные ограничения, связывающие поведение плотности  $p(\omega; t)$  при далеких друг от друга значений параметра, см. библиографию в [34], [76]. Для нашей узкой цели проще поступить иначе, построив на базе оценки максимума правдоподобия робастную решающую процедуру, асимптотически эффективную без дополнительных ограничений, ср. [10], [118], [127], [130], [131].

Тем самым мы уклонились от детального рассмотрения очень важной проблемы, разрабатываемой Барндорфф—Нильсеном и Блэзилдом, см. [86]—[88], и библиографию там. Если поле информационных количеств Кульбака  $I(P_x: P_y)$  задает *ожидаемую* [expected] информационную геометрию семейства  $\{P_t, t \in F\}$ , то средние (11.22) задают *наблюдаемую* [observed] его геометрию. Поэтому вопрос, когда наблюдаемая геометрия сходится к ожидаемой, является принципиальным. В этой связи, обобщая понятие логарифмической функции правдоподобия, они ввели понятие йока (yoke — парная упряжка, узы, скоба — англ.), связав с каждым йоком свою серию  $\alpha$ -связностей. Тем самым, они построили новые информационные геометрии, правда уже не инвариантные относительно категории статистических решающих правил. Однако такая конструкция является очень естественной для более широкого класса статистических задач, чем обратная задача теории вероятностей, когда в саму постановку задачи входят некоторые дополнительные структуры.

Однако для одной далекой точки  $t$  или конечно числа фиксированных точек обрисованная выше картина поведения случайного поля  $g^{(N)}(\varepsilon; t; \theta)$  оказывается правильной.

Лемма 11.7. Для 2-гладкого простого компактного подсемейства  $\mathcal{F} = \{P_t, t \in F\}$  и любого достаточно малого размера  $a$ ,

$$a \leq \min \{\rho(\mathcal{F}), \rho_1(\mathcal{F})\} \quad (11.25)$$