

Доказательство. Положим $d = \min\{\rho(\mathcal{F}), (3B_1)^{-1}\}$, см. (10.18). На компакте

$$A(d) = \{(\mathbf{x}, \mathbf{y}) : \|\mathbf{x} - \mathbf{y}\|_y \geq d, \mathbf{x} \in K(F), \mathbf{y} \in F\} \quad (11.21)$$

полунепрерывная снизу функция $(\mathbf{x}, \mathbf{y}) \rightarrow (P_{\mathbf{x}}, P_{\mathbf{y}}) \rightarrow I(P_{\mathbf{x}} : P_{\mathbf{y}})$ достигает своего минимума $b(d) = I(P_z : P_t) > 0$, ибо семейство \mathcal{F} — простое, и $I(P_z : P_t) = 0$ может быть лишь при $P_z = P_t$, $z = t$. Примем $\rho_1(\mathcal{F}) = [3b(d)]^{1/2}$, тогда при $r \leq \rho_2(\mathcal{F})$ будет $b(d) \geq r^2/3$. Для проверки выполнения (11.20) рассмотрим два случая. Когда $\|\mathbf{t} - \boldsymbol{\theta}\|_e \geq d$, то $(\mathbf{t}, \boldsymbol{\theta}) \in A(d)$ и $I(P_t : P_{\boldsymbol{\theta}}) \geq b(d) \geq r^2/3$. А если $r \leq \|\mathbf{t} - \boldsymbol{\theta}\|_e \leq d$, $a \leq (3B_1)^{-1}$, $d \leq \rho(\mathcal{F})$, то по лемме 10.5

$$2I(\mathbf{t}, \boldsymbol{\theta}) \geq \|\mathbf{t} - \boldsymbol{\theta}\|_e^2 (1 - 1/3) \geq 2r^2/3.$$

В обоих случаях неравенство (11.20) проверено.

Для компактных гладких семейств оценку наибольшего правдоподобия естественно определять следующим образом.

Определение 11.3. Пусть $\mathcal{F} = \{P_t, t \in F\}$ — гладкое компактное подсемейство гладкого семейства $\mathcal{P} = \{P_t, t \in \Theta \subseteq \mathbb{R}^n\}$, и компактная область $M = [\text{Int } M] \subseteq F$. Назовем оценкой M -максимального правдоподобия $\tau = \tau(\omega^1, \dots, \omega^N)$ параметра неизвестного распределения вероятностей $P_t \in \mathcal{F}$ по P_t^N -выборке $(\omega^1, \dots, \omega^N) = \varepsilon$ точку абсолютного максимума на M гладкой логарифмической функции правдоподобия

$$l^{(N)}(\varepsilon; \mathbf{t}) = \ln p(\omega^1; \mathbf{t}) + \dots + \ln p(\omega^N; \mathbf{t}),$$

если такая точка единственная, и раз и навсегда фиксированную точку $\theta_0 \in M$ в противном случае.

Когда исходы ω^i распределены по закону $P_{\theta}(\cdot)$, вместо $l^{(N)}(\varepsilon; \mathbf{t})$ удобнее рассматривать θ -приведенную функцию

$$g^{(N)}(\varepsilon; \mathbf{t}; \boldsymbol{\theta}) = N^{-1} \sum_i [\ln p(\omega^i; \mathbf{t}) - \ln p(\omega^i; \boldsymbol{\theta})]. \quad (11.22)$$

Очевидно для каждой выборки ε обе функции, получаемые друг из друга изменением масштаба и сдвигом, имеют общие стационарные точки, а их производные отличаются только масштабным множителем N .

Если абсолютный максимум достигается в точке $\mathbf{x} \in \text{Int } M$, градиент логарифмической функции правдоподобия обращается при $\mathbf{t} = \mathbf{x}$ в нуль, т. е. координаты x^1, \dots, x^n удовлетворяют системе уравнений максимального правдоподобия.

$$g_j^{(N)}(\varepsilon; \mathbf{x}) = N^{-1} \sum_i q_j(\omega^i; \mathbf{x}) = 0, \quad j = 1, \dots, n. \quad (11.23)$$

Этой же системе удовлетворяют координаты точек локальных максимумов функции правдоподобия и, вообще, всех ее стационарных точек. Для достаточно «искривленных» семейств последнее множество может быть весьма обширным; поэтому систему (10.23) невыгодно брать за основу при определении оценки.