## §5. Эффективность оценок

*Лекция 8*

Обсуждая в начале нашего курса общую проблему статистического вывода, мы говорили о главной задаче математической статистики – построении решающих правил $\delta_n = \delta_n(X^{(n)})$, минимизирующих равномерно по всем $\theta \in \Theta$ функцию риска $R(\theta;\, \delta_n)$. К сожалению, без дополнительных ограничений на класс решающих функций эта задача не разрешима. Действительно, рассмотрим проблему оценки параметра $\theta$, в которой пространство решений $\mathcal{D}$ совпадает с параметрическим пространством $\Theta$, а решающая функция $\delta_n = \hat{\theta}_n$ – оценка $\theta$. Возьмем в качестве оценки некоторую фиксированную точку $\theta_0 \in \Theta$, то есть при любом результате $x^{(n)}$ статистического эксперимента будем принимать одно и то же решение $d = \theta_0$. Если функция потерь обладает тем естественным свойством, что $L(\theta,\, \theta) = 0$, каково бы ни было значение $\theta \in \Theta$, то риск такой оценки $R(\theta;\, \theta_0) = L(\theta,\, \theta_0)$ при $\theta = \theta_0$ равен нулю. Таким образом, если мы хотим построить оценку с равномерно минимальным риском в классе всевозможных оценок $\theta$, то мы должны найти оценку $\theta_n^*$ с функцией риска $R(\theta,\, \theta_n^*) \equiv 0$, и понятно, что такой оценки не существует. Поэтому мы будем всегда при поиске оптимальных решений указывать класс оценок, в которых ищется оптимальное решение.

**Определение 5.1.** Оценка $\theta_n^* = \theta_n^*(X^{(n)})$ называется *оптимальной* или *оценкой с равномерно минимальным риском* в классе $\mathcal{K}$ оценок параметра $\theta$, если для любой оценки $\hat{\theta}_n \in \mathcal{K}$ и каждого $\theta \in \Theta$ имеет место неравенство $R(\theta;\, \theta_n^*) \le R(\theta;\, \hat{\theta}_n)$.

Ниже предлагается метод нахождения оптимальных оценок скалярного параметра $\theta$ при квадратичной функции потерь в классе несмещенных оценок: $\mathbf{E}_\theta \hat{\theta}_n(X^{(n)}) = \theta$ при любом $\theta \in \Theta \subseteq \mathrm{R}$, но при дополнительных ограничениях на вероятностную модель и соответствующее семейство распределений оценки. Эти ограничения аналогичны тем условиям регулярности, которые мы накладывали на вероятностную модель при изучении асимптотических свойств оценок максимального правдоподобия. Мы покажем, что квадратичный риск любой несмещенной оценки, удовлетворяющей этим условиям, не может быть меньше $[nI(\theta)]^{-1}$ – асимптотической дисперсии оценки максимального правдоподобия (см. теорема 4.2). Следовательно, метод максимального правдоподобия доставляет асимптотическое решение проблемы оптимальной оценки. Более того, мы покажем, что при наличие достаточных статистик метод максимального правдоподобия может привести и к точному решению проблемы равномерной минимизации функции риска.

Сформулируем условия регулярности, при выполнении которых будет находиться нижняя (достижимая!) граница квадратичного риска оценки.

(B1) Носитель $\mathcal{X}$ распределения $P_\theta$ наблюдаемой случайной величины $X$ не зависит от $\theta \in \Theta$ (условие, совпадающее с (R2) в §4).

(B2) Информация по Фишеру $I(\theta)$ строго положительна при любом $\theta \in \Theta$ (условие, совпадающее с (R5) в §4).

(B3) Равенство
$$\int_{\mathcal{X}^n} f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) = 1$$
можно дифференцировать по $\theta$ под знаком интеграла, то есть
$$\int_{\mathcal{X}^n} f_n'(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) = 0.$$
По аналогии с (R4) в части (5) для этого достаточно потребовать существование такой интегрируемой по мере $\mu$ функции $H(x)$, что в некоторой окрестности любой точки $\theta \in \Theta$ выполняется неравенство $|\,\partial f(x \mid \theta)/\partial\theta\,| \le H(x),\ x \in \mathcal{X}$.

(B4) Оценка $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ должна принадлежать классу оценок $\mathcal{K}'$, среднее значение которых
$$\mathbf{E}_\theta\, \hat{\theta}_n(X^{(n)}) = \int_{\mathcal{X}^n} \hat{\theta}_n(x^{(n)}) f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)})$$
можно дифференцировать по $\theta \in \Theta$ под знаком интеграла.

Конечно, условие (B4) требует комментария. В “высокой” теории статистического вывода приводятся достаточные условия на семейство распределений $\{P_\theta,\, \theta \in \Theta\}$ наблюдаемой случайной величины $X$, которые обеспечивают выполнение условия (B4), но формулировка этих условий и, в особенности, доказательство того, что они влекут (B4), настолько технически и концептуально сложны, что могут составить предмет специального курса. Однако все изучаемые нами в курсе ТВ вероятностные модели, за исключением равномерного распределения, удовлетворяют этим условиям, и поэтому любая оценка их параметров принадлежит классу $\mathcal{K}'$.

Прежде, чем получить основной "технический" результат этого параграфа, вспомним одно замечательное неравенство из курса математического анализа. Это – неравенство Коши–Буняковского, которое в случае интегралов Лебега по вероятностной мере $P$ называется неравенством Шварца. Пусть $Y$ – случайная величина с распределением $P$ и $g$, $h$ – две интегрируемые с квадратом по мере $P$ функции на области $\mathcal{Y}$ значений $Y$. Для этих функций имеет место неравенство $(\mathbf{E}\, g(Y)h(Y))^2 \le \mathbf{E}\, g^2(Y) \cdot \mathbf{E}\, h^2(Y)$ или, что то же,

$$\left(\int_{\mathcal{Y}} g(y)h(y)\, dP(y)\right)^2 \le \int_{\mathcal{Y}} g^2(y) dP(y) \cdot \int_{\mathcal{Y}} h^2(y) dP(y),$$

причем знак равенства достигается тогда и только тогда, когда функции $g$ и $h$ линейно зависимы: существуют такие постоянные $a$ и $b$, что $ag(y) + bh(y) = 0$ для почти всех $y \in \mathcal{Y}$ по мере $P$.

**Теорема 5.1.** (неравенство Рао–Крамера) *При выполнении условий (B1)–(B4) для квадратичного риска любой оценки $\hat{\theta}_n \in \mathcal{K}'$ справедливо неравенство*

$$\mathbf{E}_\theta \left(\hat{\theta}_n(X^{(n)}) - \theta\right)^2 \ge \mathbf{D}_\theta\, \hat{\theta}_n(X^{(n)}) \ge \frac{[\, d\gamma(\theta)/d\theta\,]^2}{nI(\theta)}, \tag{1}$$

*где $\gamma(\theta) = \mathbf{E}_\theta\, \hat{\theta}_n(X^{(n)})$, причем знак равенства между риском и дисперсией оценки $\hat{\theta}_n$ достигается на несмещенных оценках: $\gamma(\theta) = \theta$, а знак равенства во втором неравенстве (1) имеет место тогда и только тогда, когда существует такая параметрическая функция $C(\theta), \theta \in \Theta$, что*

$$\hat{\theta}_n(X^{(n)}) - \gamma(\theta) = C(\theta) \frac{\partial \mathcal{L}(\theta \mid X^{(n)})}{\partial \theta}\, . \tag{2}$$

Д о к а з а т е л ь с т в о.. Продифференцируем обе части равенств

$$\int_{\mathcal{X}^n} f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) = 1,$$

$$\int_{\mathcal{X}^n} \hat{\theta}_n(x^{(n)}) f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) = \gamma(\theta)$$

по параметру $\theta$, занося производные в левых частях под знаки интегралов, что можно сделать благодаря условиям (B3) и (B4). Полученный результат, используя условие (B1), представим в виде

$$\int_{\mathcal{X}^n} \frac{\partial \mathcal{L}(\theta \mid x^{(n)})}{\partial \theta} f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) = 0,$$

$$\int_{\mathcal{X}^n} \hat{\theta}_n(x^{(n)}) \frac{\partial \mathcal{L}(\theta \mid x^{(n)})}{\partial \theta} f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) = \gamma'(\theta).$$

Вычтем из второго равенства первое, умножив его предварительно на $\gamma(\theta)$ :

$$\int_{\mathcal{X}^n} \left(\hat{\theta}_n(x^{(n)}) - \gamma(\theta)\right) \frac{\partial \mathcal{L}(\theta \mid x^{(n)})}{\partial \theta} f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) = \gamma'(\theta).$$

Применим к левой части полученного равенства неравенство Шварца, полагая $y = x^{(n)}$, $\mathcal{Y} = \mathcal{X}^n$, $g(x^{(n)}) = \hat{\theta}_n(x^{(n)}) - \gamma(\theta)$, $h(x^{(n)}) = \partial \mathcal{L}(\theta \mid x^{(n)})/\partial \theta$. В результате получим неравенство

$$(\gamma\prime(\theta))^2 \leq \int_{\mathcal{X}^n} \left(\hat{\theta}_n(x^{(n)}) - \gamma(\theta)\right)^2 f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}) \cdot$$

$$\int_{\mathcal{X}^n} \left(\frac{\partial \mathcal{L}(\theta \mid x^{(n)})}{\partial \theta}\right)^2 f_n(x^{(n)} \mid \theta)\, d\mu_n(x^{(n)}), \tag{3}$$

в котором знак равенства достигается тогда и только тогда, когда выполняется соотношение (2).

Мы получили неравенства (1), поскольку первое из них очевидно (на дисперсии достигается минимум всевозможных средних квадратичных уклонений случайной величины от постоянной). Второе неравенство в (1) есть следствие неравенства (3), ибо первый интеграл в правой части (3) равен $\mathbf{D}_\theta \hat{\theta}_n$, а второй интеграл определяет фишеровскую информацию $I_n(\theta)$, содержащуюся в выборке. Наконец, из пункта $2^0$ леммы 4.2 следует, что $I_n(\theta) = nI(\theta)$.

**Следствие 5.1.** *Если $\hat{\theta}_n$ принадлежит подклассу $\mathcal{K} \subseteq \mathcal{K}'$ несмещенных оценок класса $\mathcal{K}'$, то ее квадратичный риск*

$$R(\theta;\, \hat{\theta}_n) = \mathbf{D}_\theta \hat{\theta}_n \,\geq\, [n\, I(\theta)]^{-1}\,, \tag{4}$$

*причем знак равенства тогда и только тогда, когда выполняется равенство (2) с $\gamma(\theta) = \theta$.*

Понятно, что это следствие есть частный случай доказанной теоремы. Оно указывает неконструктивный путь к построению несмещенных оценок с равномерно минимальным риском. Достаточно вычислить производную в правой части равенства (2) и затем подбирать статистику $\hat{\theta}_n = \hat{\theta}_n(X^{(n)})$ и параметрическую функцию $C(\theta)$, для которых имеет место равенство

$$\hat{\theta}_n(X^{(n)}) - \theta = C(\theta)\frac{\partial \mathcal{L}(\theta \mid X^{(n)})}{\partial \theta}\,.$$

Обычно это можно сделать в случае статистических структур, обладающих достаточными статистиками, где, в силу теоремы факторизации (теорема 2.1 из §2), функция правдоподобия $L(\theta \,|\, X^{(n)}) = g_\theta(T(X^{(n)}))h(X^{(n)})$, и последнее равенство имеет вид

$$\hat{\theta}_n(X^{(n)}) - \theta = C(\theta)\frac{\partial \ln g_\theta(T(X^{(n)}))}{\partial \theta}\,. \tag{5}$$

Например, для показательного распределения с функцией плотности $f(x \,|\, \theta) = \theta^{-1}\exp\{-x/\theta\},\ x > 0$, функция

$$\ln g_\theta(X^{(n)}) = -n \ln\, \theta - \theta^{-1}\textstyle\sum_1^n X,$$

ее производная

$$\partial \ln g_\theta(T(X^{(n)}))/\partial\theta = -n/\theta + \textstyle\sum_1^n X/\theta^2,$$

и равенство (5) выполняется при $C(\theta) = \theta^2/n$ и $\hat{\theta}_n = \overline{X}$. Таким образом, выборочное среднее $\overline{X}$ есть несмещенная оценка с равномерно минимальным риском для параметра $\theta$ показательного распределения. Напомним, что $\overline{X}$ оценка $\theta$ как по методу моментов, так и по методу максимального правдоподобия.

Легко понять, что если в (4) достигается знак равенства, то $\hat{\theta}_n$ – оптимальная оценка в классе $\mathcal{K}$, но обратное, вообще говоря, может и не выполняться – мы не располагаем утверждением, что любая оптимальная оценка имеет квадратичный риск, равный $[n\, I(\theta)]^{-1}$. Чтобы подчеркнуть это различие и указать в дальнейшем более конструктивный метод построения оптимальных оценок, введем еще одно определение, рассмотрев более общую задачу несмещенной оценки некоторой параметрической функции $\gamma(\theta)$.

**Определение 5.2.** Несмещенная оценка $\hat{\gamma}_n = \hat{\gamma}_n(X^{(n)})$ параметрической функции $\gamma(\theta)$ называется *эффективной* в классе $\mathcal{K}$, если ее квадратичный риск $R(\gamma;\, \hat{\gamma}_n) = \mathbf{E}_\theta\, \left(\hat{\gamma}(X^{(n)}) - \gamma(\theta)\right)^2 = \mathbf{D}_\theta\, \hat{\gamma}_n(X^{(n)}) = \gamma'(\theta)/n\, I(\theta)$, то есть (см. теорему 5.1 с $\hat{\theta}_n = \hat{\gamma}_n$) выполняется равенство

$$\hat{\gamma}_n(X^{(n)}) - \gamma(\theta) = C(\theta)\frac{\partial \mathcal{L}(\theta \mid X^{(n)})}{\partial \theta}. \qquad (6)$$

Оценка $\hat{\gamma}_n$ называется *асимптотически эффективной* в классе $\mathcal{K}'$, если $\mathbf{E}_\theta\, \hat{\gamma}_n(X^{(n)}) \sim \gamma(\theta)$ и $\mathbf{D}_\theta\, \hat{\gamma}_n(X^{(n)}) \sim \gamma'(\theta)/n\, I(\theta)$, когда $n \to \infty$.

В силу теоремы 4.2 оценка по методу максимального правдоподобия скалярного параметра $\theta$ (в данном случае $\gamma(\theta) = \theta$) является асимптотически эффективной оценкой в классе $\mathcal{K}'$. Покажем, что она дает решение проблемы построения эффективной оценки в классе $\mathcal{K}$.

Пусть $\hat{\theta}_n$ – оценка максимального правдоподобия параметра $\theta$. Определим оценку $\gamma(\hat{\theta}_n)$ параметрической функции $\gamma(\theta)$ с помощью подстановки вместо $\theta$ ее оценки $\hat{\theta}_n$.

**Теорема 5.2.** *Если $\gamma(\hat{\theta}_n)$ есть несмещенная оценка параметрической функции $\gamma(\theta)$ и эффективная в классе $\mathcal{K}$ оценка $\gamma_n^*$ параметрической функции $\gamma(\theta)$ существует, то при выполнении условий регулярности (R1)–(R5) и (B4) почти наверное $\gamma_n^*(X^{(n)}) = \gamma(\hat{\theta}_n(X^{(n)}))$.*

Доказательство.. Если $\gamma_n^*$ – эффективная оценка $\gamma(\theta)$, то она удовлетворяет равенству (6):

$$\gamma_n^*(X^{(n)}) - \gamma(\theta) = C(\theta)\partial \mathcal{L}(\theta \mid X^{(n)})/\partial \theta, \qquad (7)$$

каково бы ни было $\theta \in \Theta$. Но если $\hat{\theta}_n$ – оценка по методу максимального правдоподобия, то $\partial \mathcal{L}(\hat{\theta}_n \mid X^{(n)})/\partial \theta = 0$, так что равенство (7) при $\theta = \hat{\theta}_n$ превращается в равенство $\gamma_n^*(X^{(n)}) - \gamma(\hat{\theta}_n) = 0$ почти наверное по вероятности $P_\theta^n$.

Из доказанной теоремы немедленно вытекает, что выборочное среднее $\overline{X}$ есть эффективная (следовательно, и оптимальная) несмещенная оценка параметра $\theta$ таких распределений, как двухточечное, биномиальное при известном $m$, Пуассона, показательное; $\overline{X}$ есть также несмещенная оценка с равномерно минимальным квадратичным риском среднего значения $\mu$ нормального $(\mu,\, \sigma^2)$ распределения.