Лекция 2

Оптимальные, а тем более эффективные оценки существуют не всегда. Поэтому необходимы общие методы получения достаточно хороших оценок. Мы рассмотрим два таких метода.

1. Метод моментов

Пусть $X_1, \ldots, X_n$ — выборка с распределением $F(x,\theta)$, где $\theta \in \Theta \subset \mathbf{R}^m$. Обозначим

$$\mathbf{E}_\theta X_1^k = \alpha_k(\theta).$$

Предположим, что система уравнений

$$\begin{cases} \alpha_{i_1}(\theta) = A_{i_1}, \\ \alpha_{i_2}(\theta) = A_{i_2}, \\ \quad\vdots \\ \alpha_{i_m}(\theta) = A_{i_m}, \end{cases} \quad i_j \neq i_l \Leftarrow j \neq l$$

однозначно разрешима, причём её решение даётся обратимыми функциями

$$\begin{cases} \theta_1 = \psi_1\left(A_{i_1}, \ldots, A_{i_m}\right), \\ \quad\vdots \\ \theta_m = \psi_m\left(A_{i_1}, \ldots, A_{i_m}\right), \end{cases} \quad i_j \neq i_l \Leftarrow j \neq l.$$

Оценки, полученные таким способом, называются точечными оценками, полученными *методом моментов*. Суть метода заключается в том, что выборочные моменты приравниваются теоретическим, откуда получаются значения параметров. Сразу очевиден главный недостаток метода: если какие-либо моменты не существуют, то метод может оказаться не применим, например, параметр выборки, имеющей распределение Коши с плотностью

$$p(x,\theta) = \frac{1}{\pi\left(1 + (x-\theta)^2\right)},$$

не может быть оценён этим методом, так как у случайной величины с распределением Коши отсутствуют моменты всех порядков.

Пример. $X_1, \ldots, X_n \sim \mathrm{U}([\theta_1,\theta_2])$.

$$\begin{array}{r} \mathbf{E}_\theta X_1 = \dfrac{\theta_1+\theta_2}{2}, \\ \mathbf{E}_\theta X_1^2 = \dfrac{\theta_1^2+\theta_1\theta_2+\theta_2^2}{3}, \end{array} \Rightarrow \begin{cases} \dfrac{\theta_1+\theta_2}{2} = A_1 = \dfrac{1}{n}\sum\limits_{i=1}^n X_i = \overline{X}, \\ \dfrac{\theta_1^2+\theta_1\theta_2+\theta_2^2}{3} = A_2 = \dfrac{1}{n}\sum\limits_{i=1}^n X_i^2. \end{cases}$$

Эта система эквивалентна следующей

$$\begin{cases} \theta_1+\theta_2 = 2\overline{X}, \\ \theta_1\theta_2 = 4\overline{X}^2 - 3A_2. \end{cases}$$

θ_1 и θ_2 являются корнями уравнения $t^2 - 2\overline{X}t + 4\overline{X}^2 - 3A_2 = 0$ и

$$\begin{cases} \theta_1 = \overline{X} - \sqrt{3\left(A_2 - \overline{X}\right)^2}, \\ \theta_2 = \overline{X} + \sqrt{3\left(A_2 - \overline{X}\right)^2}, \end{cases} \Leftrightarrow \begin{cases} \theta_1 = \overline{X} - \sqrt{3S^2}, \\ \theta_2 = \overline{X} + \sqrt{3S^2}, \end{cases}$$

Теорема 6. Пусть θ_1, …, θ_m являются непрерывными функциями от моментов

$$\hat{\theta}_i = \psi_i\left(A_{j_1}, \ldots, A_{j_m}\right), \quad i = 1, \ldots, m.$$

Тогда оценки, полученные методом моментов с моментами порядков j_1, …, j_m будут состоятельными и асимптотически несмещёнными.

Доказательство. Согласно нашим предположениям система имеет единственное решение $\hat{\theta}_i = \psi_i\left(A_{j_1}, \ldots, A_{j_m}\right), \quad i = 1, \ldots, m$, причём ψ_i — непрерывные функции. По усиленному закону больших чисел ψ_i сходятся с вероятностью 1 к теоретическим моментам, а из непрерывности функций ψ_i отсюда следует, что оценки, получаемые методом моментов при $n \to \infty$ сходятся с вероятностью 1 к θ_i. Теорема доказана.

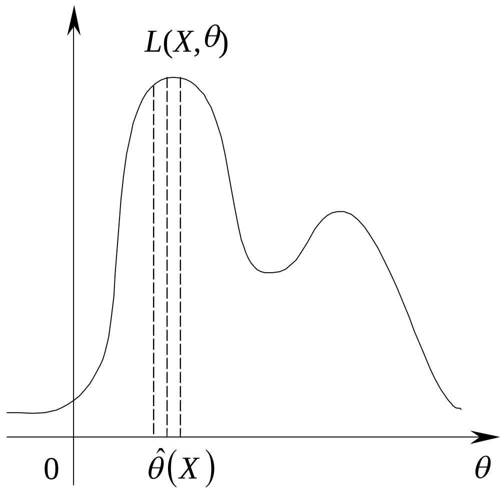
Метод моментов даёт состоятельные оценки, но часто их эффективность и асимптотическая эффективность меньше единицы.

2. Метод максимального правдоподобия

1°. Метод максимального правдоподобия. Пусть $L(X,\theta)$ — функция правдоподобия выборки X.

Определение 1. Оценкой максимального правдоподобия $\hat{\theta}(X)$ параметра θ называется такое значение параметра, что

$$\max L(X,\theta) = L\left(X, \hat{\theta}(X)\right).$$



Справедливо локальное утверждение, что в точках, в которых плотность больше, достигается большее значение вероятности. Таким образом, то, что наблюдается в эксперименте, наиболее вероятно при данном значении параметра.

Примеры. 1. $X_1,\ldots,X_n \sim \mathrm{N}\left(\theta_1,\theta_2^2\right)$, $\theta_1 \in \mathbf{R}$, $\theta_2 > 0$,

$$L\left(X,\theta_1,\theta_2\right)=\frac{1}{\left(\sqrt{2\pi}\theta_2\right)^n}\exp\left[-\frac{1}{2\theta_2^2}\sum_{i=1}^{n}\left(X_i-\theta_1\right)^2\right].$$

Будем рассматривать функцию правдоподобия лишь на множестве, на котором она положительна (так как в нулях заведомо не будет максимума, поскольку для того, чтобы интеграл равнялся единице, должна существовать хотя бы одна точка, в которой значение положительно). Возьмём логарифм функции правдоподобия:

$$\ln L\left(X,\theta_1,\theta_2\right)=-\frac{n}{2}\ln\theta_2^2-\frac{n}{2}\ln 2\pi-\frac{1}{2\theta_2^2}\sum_{i=1}^{n}\left(X_i-\theta_1\right)^2 .$$

Возьмём частные производные и приравняем их нулю (в данном случае корнями системы будут максимумы):

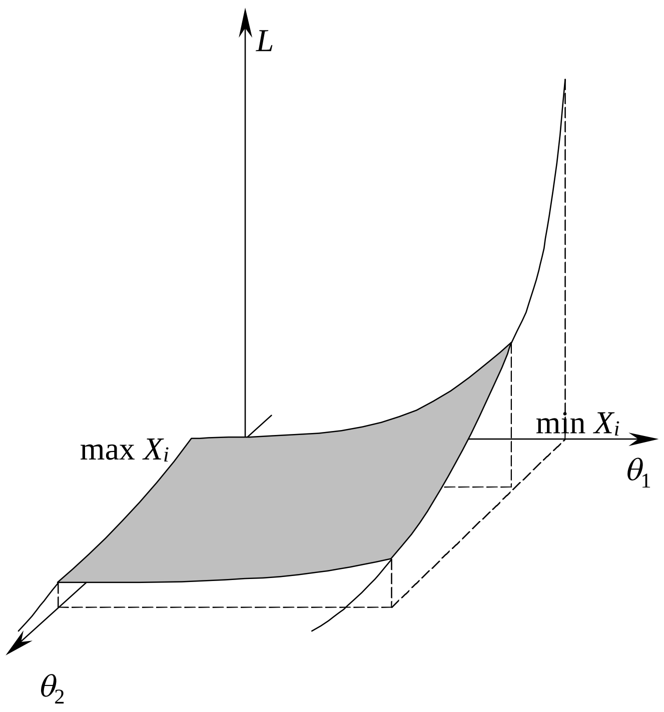
$$\begin{cases}\dfrac{\partial\ln L}{\partial\theta_1}=\dfrac{1}{\theta_2^2}\displaystyle\sum_{i=1}^{n}\left(X_i-\theta_1\right)=0, & (1)\\[2ex] \dfrac{\partial\ln L}{\partial\left(\theta_2^2\right)}=-\dfrac{n}{2\theta_2^2}+\dfrac{1}{2\theta_2^4}\displaystyle\sum_{i=1}^{n}\left(X_i-\theta_1\right)^2=0, & (2)\end{cases}$$

$$(1)\Rightarrow\theta_1=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}=\hat{\theta}_1(X),$$
$$(2)\Rightarrow\theta_2^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=S^2=\hat{\theta}_2(X).$$

Заметим, что $\hat{\theta}_1(X)$ и $\hat{\theta}_2(X)$ — достаточные статистики.

2. $X_1, \ldots, X_n \sim \mathrm{U}([\theta_1,\theta_2])$, $\theta_1 < \theta_2$, $H(x)=\begin{cases}1, & x\ge 0,\\ 0, & x<0,\end{cases}$

$$L\left(X,\theta_1,\theta_2\right)=\frac{1}{\left(\theta_2-\theta_1\right)^n}H\left(\theta_2-\max X_i\right)H\left(\min X_i-\theta_1\right)$$

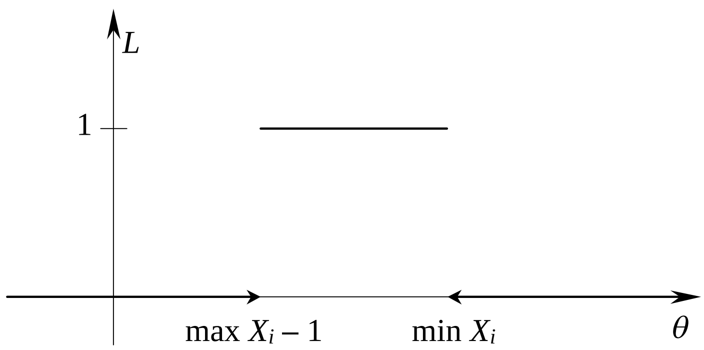


Функция правдоподобия по θ_1 возрастает, а по θ_2 — убывает, следовательно, оценками максимального правдоподобия являются $\hat{\theta}_1 = \min X_i$, $\hat{\theta}_2 = \max X_i$.

Оценка максимального правдоподобия может быть не единственной:

3. $X_1, \ldots, X_n \sim U([\theta, \theta + 1])$,

$$L(X,\theta) = H(\theta + 1 - \max X_i)\cdot H(\min X_i - \theta).$$



В этом случае оценкой максимального правдоподобия будет любое значение параметра, лежащее между $\max X_i - 1$ и $\min X_i$, то есть $\alpha\cdot(\max X_i - 1) + (1 - \alpha)\cdot\min X_i$, $0 \le \alpha \le 1$. Если дополнительно потребовать несмещённость оценки, то

$$\mathbf{E}_\theta T(X) = \theta \Rightarrow \alpha = \frac{1}{2}.$$

2°. Свойства метода максимального правдоподобия.

Если $\hat{\theta}(X)$ — оценка максимального правдоподобия θ и $\tau(\theta)$ — взаимно однозначная функция θ, то оценкой максимального правдоподобия функции $\tau(\theta)$ является функция $\tau\left(\hat{\theta}(X)\right)$.

Если существует достаточная статистика $T(X)$, то оценка максимального правдоподобия есть функция $T(X)$: $\hat{\theta}(X) = \varphi(T(X))$.

(это следует из критерия факторизации:

$$L(X,\theta) = \underbrace{g(T(X),\theta)}\cdot \underbrace{h(X)}_{\text{константа по } \theta} \quad).$$

Если существует эффективная оценка параметра θ, то она совпадает с оценкой максимального правдоподобия: $\exists \tilde{\theta}(X) \Rightarrow \hat{\theta}(X) = \tilde{\theta}(X)$. Действительно,

$$\tilde{\theta}(X) - \theta = a_n(\theta)\cdot \frac{\partial \ln L(X,\theta)}{\partial \theta} \qquad \forall \theta .$$

Пусть теперь $\theta = \hat{\theta}(X)$ — оценка максимального правдоподобия. Тогда $\tilde{\theta}(X) - \hat{\theta}(X) = a_n \cdot 0$, так как это точка, в которой достигается максимум функции правдоподобия.

Определение 2. Последовательность случайных величин $\{\xi_n\}$ *асимптотически нормальна*, если существуют такие a_n, $b_n \in \mathbf{R}$, что

$$\frac{\xi_n - a_n}{b_n} \Rightarrow \xi \sim \mathrm{N}\ (0,1).$$

Например, если $\{\xi_n\}$ удовлетворяет условиям центральной предельной теоремы и $S_n =$ $= \xi_1 + \ldots + \xi_n$, то

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \Rightarrow \mathrm{N}\ (0,1),$$

следовательно, S_n — асимптотически нормальная последовательность.

По аналогии оценка $T_n(X)$ называется *асимптотически нормальной*, если

$$\frac{T_n(X) - a_n(\theta)}{b_n(\theta)} \Rightarrow \mathrm{N}\ (0,1).$$

Рассмотрим неравенство Рао-Крамера:

$$\mathbf{D}_\theta T_n(X) \geq \frac{(\tau'(\theta))^2}{\mathbf{E}_\theta U^2(X,\theta)},$$

$$U = \frac{\partial \ln L(X,\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln p(X_i,\theta)}{\partial \theta} \Rightarrow \mathbf{E}_\theta U^2(X,\theta) = n\mathbf{E}_\theta \frac{\partial p(X_1,\theta)}{\partial \theta} = n \cdot i(\theta),$$

где $i(\theta)$ — некоторая функция.

Определение 3. Оценка $T_n(X)$ функции $\tau(\theta)$ называется *асимптотически эффективной*, если

$$\frac{\mathbf{D}_\theta T_n(X)\cdot ni(\theta)}{(\tau'(\theta))^2} \xrightarrow[n\to\infty]{} 1.$$

Пусть выполняются следующие два условия:

1) функция правдоподобия удовлетворяет условиям регулярности для первых двух производных (условиям регулярности второго порядка),

2) оценка максимального правдоподобия $\hat{\theta}(X)$ для всех θ существует, единственна и достигается во внутренней точке множества Θ.

Тогда оценка максимального правдоподобия $\hat{\theta}(X)$

a) асимптотически несмещена,

b) состоятельна,

c) асимптотически эффективна и

d) асимптотически нормальна:

$$\sqrt{n \cdot i(\theta)}\left(\hat{\theta}(X) - \theta\right) \Rightarrow \mathrm{N}\ (0,1).$$