Лекция 6

Метод максимального правдоподобия

Пусть закон распределения СВ X известен с точностью до неизвестных параметров $\theta_1,\dots,\theta_k$. Если СВ X, порождающая выборку $\{X_1,\dots,X_n\}$, является дискретной, то пусть

$$(1)\qquad p(x;\theta_1,\dots,\theta_k) = P\{X=x\}, \quad x\in\mathcal{X},$$

где $\mathcal{X}$ - множество всех возможных значений СВ X. Если СВ X является непрерывной и имеет плотность распределения $f(x)$, то пусть

$$(2)\qquad p(x;\theta_1,\dots,\theta_k) = f(x), \quad x\in\mathcal{X}.$$

(Правые части равенств (1) и (2) тоже, разумеется, зависят от $\theta_1,\dots,\theta_k$).

<u>Определение 1</u> Функцией правдоподобия выборки $\{X_1,\dots,X_n\}$ называется функция

$$(3)\qquad L_n(\theta_1,\dots,\theta_k;X_1,\dots,X_n) = \prod_{i=1}^{n} p(X_i;\theta_1,\dots,\theta_k),$$

которая при $X_1=x_1,\dots,X_n=x_n$ и фиксированных $\theta_1,\dots,\theta_k$ в непрерывном случае совпадает со значением n-мерной плотности распределения системы независимых СВ $X_1,\dots,X_n$, а в дискретном — со значением вероятности события $\{X_1=x_1,\dots,X_n=x_n\}$.

<u>Определение 2</u>. Пусть $\{X_1, \dots, X_n\}$ – выборка из распределения, зависящего от параметров $\theta_1, \dots, \theta_k$. Пусть $h(x_1, \dots, x_n)$ – точка глобального максимума функции правдоподобия (3) (на множестве возможных значений $\theta_1, \dots, \theta_k$) при фиксированных $X_1 = x_1, \dots, X_n = x_n$;

$$h(x_1, \dots, x_n) = \begin{pmatrix} h_1(x_1, \dots, x_n) \\ \vdots \\ h_k(x_1, \dots, x_n) \end{pmatrix}$$

– k-мерная точка в пространстве параметров $\theta_1, \dots, \theta_k$. Оценки

$$\hat\theta_1 = h_1(X_1, \dots, X_n)$$
$$\dots\dots\dots$$
$$\hat\theta_k = h_k(X_1, \dots, X_n)$$

называются <u>оценками максимального правдоподобия</u> параметров $\theta_1, \dots, \theta_k$.

Таким образом, в техническом плане нахождение оценки максимального правдоподобия сводится к нахождению точки глобального максимума функции k-переменных.

$$L_n(\hat\theta_1, \dots, \hat\theta_k; x_1, \dots, x_n) = \max_{\substack{\text{по всем} \\ \text{возможным} \\ \text{значениям} \\ \theta_1, \dots, \theta_k}} L_n(\theta_1, \dots, \theta_k; x_1, \dots, x_n)$$

Пример 1 Методом максим. правдоподобия найти оценку $\widehat{\theta}$ параметра θ распределения Бернулли:

X	0	1
P	$1-\theta$	θ

;

отрезок $[0;1]$ — множество допустимых значений параметра θ

Имеем:

$$p(x,\theta)=P\{X=x\}=\begin{cases}\theta, & \text{если } x=1,\\ 1-\theta, & \text{если } x=0.\end{cases}$$

Функция правдоподобия

$$L_n(\theta;x_1,\dots,x_n)=\prod_{i=1}^{n}p(x_i,\theta)=\theta^k(1-\theta)^{n-k},$$

где k — число единиц среди $x_1,\dots,x_n$.
Найдём максимум $L_n(\theta;x_1,\dots,x_n)$ при фиксированных $x_1,\dots,x_n$:

$$\frac{\partial L_n}{\partial\theta}=k\theta^{k-1}(1-\theta)^{n-k}+\theta^k(n-k)(1-\theta)^{n-k-1}\cdot(-1)=$$
$$=\theta^{k-1}(1-\theta)^{n-k-1}\left(k(1-\theta)-\theta(n-k)\right)=$$
$$=\theta^{k-1}(1-\theta)^{n-k-1}(k-n\theta);$$

при $\theta\in(0,1)$ $\quad \frac{\partial L_n}{\partial\theta}=0 \Longleftrightarrow \theta=\frac{k}{n}$; видно, что $\theta=\frac{k}{n}$ — это точка максимума на $(0,1)$.
Поскольку $L_n=0$ при $\theta=0$ и $\theta=1$, то $\widehat{\theta}=\frac{k}{n}$ — искомая оценка. Поскольку $k=X_1+\dots+X_n$, то окончательно

$$\widehat{\theta}=\frac{1}{n}\sum_{i=1}^{n}X_i,\ \text{т.е. } \widehat{\theta}=\overline{X_n} \text{ - выборочное среднее}$$

<u>Пример 2</u> Методом максим. правдоподобия найти оценку $\hat{\lambda}$ параметра λ показательного распределения с плотностью

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

промежуток $(0;+\infty)$ – множество возможных значений λ

Вычисляем функцию правдоподобия

$$L_n(\lambda; x_1,\dots,x_n)=\prod_{i=1}^{n}\left(\lambda e^{-\lambda x_i}\right)=\lambda^n e^{-\lambda\sum_{i=1}^{n}x_i}.$$

Поскольку $L_n(\lambda; x_1,\dots,x_n)>0$ на области определения, то вместо максимума функции $L_n(\lambda; x_1,\dots,x_n)$ можно искать максимум функции $\ln L_n(\lambda; x_1,\dots,x_n)$:

$$\frac{\partial}{\partial\lambda}\ln L_n(\lambda; x_1,\dots,x_n)=\frac{\partial}{\lambda}\left(n\ln\lambda-\lambda\sum_{i=1}^{n}x_i\right)=$$

$$=n\cdot\frac{1}{\lambda}-\sum_{i=1}^{n}x_i\ ,$$

так что нулевое значение производной будет при $\lambda=\dfrac{1}{\frac{1}{n}\sum_{i=1}^{n}x_i}$ и это именно точка максимума.

Значит, искомая оценка

$$\hat{\lambda}=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}X_i}=\frac{1}{\overline{X}_n}.$$

Пример 3 Выборка $\{X_1, \ldots, X_n\}$ порождена СВ X, равномерно распределенной на отрезке $[\theta_1, \theta_2]$, т.е. плотность распределения

$$f(x) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1}, & x \in [\theta_1, \theta_2], \\ 0, & x \notin [\theta_1, \theta_2]. \end{cases}$$

Методом максим. правдоподобия найти оценки $\hat{\theta}_1$ и $\hat{\theta}_2$ неизвестных параметров θ_1 и $\theta_2 > \theta_1$.

Выпишем функцию правдоподобия (см. (3))

$$L_n(\theta_1, \theta_2; X_1, \ldots, X_n) = \prod_{i=1}^{n} f(\theta_1, \theta_2, X_i) =$$

$$= \begin{cases} \left(\dfrac{1}{\theta_2 - \theta_1}\right)^n, & \text{если все } X_1, \ldots, X_n \in [\theta_1, \theta_2] \\ 0, & \text{если } \exists i \in \{1, 2, \ldots, n\} \; X_i \notin [\theta_1, \theta_2] \end{cases}$$

Заметим, что если $X_{(1)}, \ldots, X_{(n)}$ – порядковые статистики, а $x_{(1)}, \ldots, x_{(n)}$ – их соответствующие реализации, то всегда

$$L_n(\theta_1, \theta_2; x_1, x_2, \ldots, x_n) \leq \underbrace{\left(\frac{1}{x_{(n)} - x_{(1)}}\right)^n}_{\text{Обозначим через } L_{max}}.$$

Теперь заметим, что

$$L_n(\theta_1, \theta_2; x_1, x_2, \dots, x_n)\Big|_{\substack{\theta_1 = x_{(1)} \\ \theta_2 = x_{(n)}}} = L_{max}$$

Это означает, что искомыми оценками для θ_1 и θ_2 являются статистики

$$\hat{\theta}_1 = X_{(1)} \quad и \quad \hat{\theta}_2 = X_{(n)}.$$

Если реализацией выборки $\{X_1, \dots, X_n\}$ является последовательность чисел $(x_1, \dots, x_n)$, то реализацией оценок $\hat{\theta}_1$ и $\hat{\theta}_2$ будут числа $x_{(1)}$ и $x_{(n)}$, где

$$x_{(1)} = \min(x_1, \dots, x_n), \quad x_{(n)} = \max(x_1, \dots, x_n).$$

В заключение, в качестве теоретического результата отметим (без доказательства), что при определённых условиях гладкости, налагаемых на функцию правдоподобия, оценки, полученные по методу максимального правдоподобия, являются всегда асимптотически несмещенными и состоятельными.