Estymacja parametrów

Krzysztof Topolski

Wykład 1

Wrocław, 27 lutego 2014

Próba losowa

Definicja.

Wektor zmiennych losowych $(X_1, X_2, ..., X_n)$ nazywamy **próba losową rozmiaru n z rozkładu o gęstości f(x) (dystrybuancie F(x))** jeśli $X_1, X_2, ..., X_n$ są niezależnymi zmiennymi losowymi o wspólnym rozkładzie z gęstością $f(x)$ (z dystrybuantą $F(x)$).

Przy tak przyjętej definicji rozkład próby losowej $X_1, X_2, ..., X_n$ ma gęstość łączną $f(x_1, x_2, ..., x_n)$ i dystrybuantę łączną odpowiednio postaci

$$f(x_1, x_2, ..., x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^{n} f(x_i).$$

oraz

$$F(x_1, x_2, ..., x_n) = F(x_1)f(x_2)\cdots F(x_n) = \prod_{i=1}^{n} F(x_i).$$

Definicja.

Wektor zmiennych losowych $(X_1, X_2, ..., X_n)$ nazywamy **próba losową rozmiaru n z rozkładu o gęstości f(x) (dystrybuancie F(x))** jeśli $X_1, X_2, ..., X_n$ są niezależnymi zmiennymi losowymi o wspólnym rozkładzie z gęstością $f(x)$ (z dystrybuantą $F(x)$).

Przy tak przyjętej definicji rozkład próby losowej $X_1, X_2, ..., X_n$ ma gęstość łączną $f(x_1, x_2, ..., x_n)$ i dystrybuantę łączną odpowiednio postaci

$$f(x_1, x_2, ..., x_n) = f(x_1)f(x_2)\cdots f(x_n) = \prod_{i=1}^{n} f(x_i).$$

oraz

$$F(x_1, x_2, ..., x_n) = F(x_1)f(x_2)\cdots F(x_n) = \prod_{i=1}^{n} F(x_i).$$

Przykład.

Łączny rozkład $f(x_1, x_2, ..., x_n)$ próby losowej z rozkładu wykładniczego z parametrem β jest postaci

$$f(x_1, x_2, ..., x_n|\beta) = \prod_{i=1}^{n} f(x_i|\beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{-(x_1+...+x_n)/\beta}.$$

Definicja.

Niech $X_1, X_2, ..., X_n$ będzie próbą losową rozmiaru n natomiast $T(x_1, x_2, ..., x_n)$ funkcją przyjmująca wartości rzeczywiste lub wektorowe, której dziedzina zawiera wartości jakie może przyjąć wektor $(X_1, X_2, ..., X_n)$. Zmienną losową lub wektor losowy

$$Y = T(X_1, X_2, ..., X_n)$$

będziemy nazywać *statystyką*, a rozkład Y będziemy nazywać *rozkładem statystyki* Y.

Przykład.

Maksimum z próby

$$X_{(n:n)} = \max(X_1, X_2, ..., X_n).$$

Przykład cd.

Minimum z próby

$$X_{(1:n)} = \min(X_1, X_2, ..., X_n).$$

Przykład.

Maksimum z próby

$$X_{(n:n)} = \max(X_1, X_2, ..., X_n).$$

Przykład cd.

Minimum z próby

$$X_{(1:n)} = \min(X_1, X_2, ..., X_n).$$

Przykład cd.

Niech $X_1, X_2, ..., X_n$ będzie próba losową oznaczmy przez

$$X_{(1:n)} \leq X_{(2:n)} \leq \; ... X_{(k:n)} \leq \; ... X_{(n:n)}$$

próbę uporządkowaną w sposób rosnący. Wektor

$$\left(X_{(1:n)}, X_{(2:n)}, \; ... \; , X_{(k:n)}, \; ... X_{(n:n)}\right)$$

nazywamy wektorem statystyk pozycyjnych, a zmienną losową $X_{(k:n)}$ nazywamy $k-$tą statystyką pozycyjną.

Definicja.

Średnią z próby nazywamy statystykę

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Definicja.

Wariancją z próby nazywamy statystykę

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Definicja.

Średnią z próby nazywamy statystykę

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

Definicja.

Wariancją z próby nazywamy statystykę

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 .$$

Twierdzenie.

Niech $x_1, ..., x_n$ *będą liczbami rzeczywistymi a* $\bar{x} = (x_1 + x_2 + \; ... \; + x_n)/n$ *ich średnią arytmetyczną. Wtedy*

- $\min_a \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2,$
- $(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.$

Dowód.

Pierwszą równość

$$\min_a \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2,$$

otrzymujemy dodając i odejmując $\bar{x}$

$$\begin{aligned}\sum_{i=1}^{n}(x_i - a)^2 &= \sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - a)^2 \\ &= \sum_{i=1}^{n}(x_i - \bar{x})^2 + 2\sum_{i=1}^{n}(x_i - \bar{x})(\bar{x} - a) + \sum_{i=1}^{n}(\bar{x} - a)^2 \\ &= \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - a)^2\end{aligned}$$

Prawa strona równości jest dodatnia i minimalna gdy $a = \bar{x}$.

Dowód cd.

Drugą równość

$$(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

otrzymujemy biorąc w równości

$$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + \sum_{i=1}^{n}(\bar{x} - a)^2$$

za $a = 0$.

Lemat.

Niech $X_1, \ldots, X_n$ *będzie próba losową, a* $g(x)$ *funkcją, dla której* $\mathrm{E}\, g(X_1)$ *oraz* $\mathrm{Var}\, g(X_1)$ *istnieją. Wtedy*

$$\mathrm{E}\left(\sum_{i=1}^{n} g(X_i)\right) = n\,\mathrm{E}\, g(X_1),$$

oraz

$$\mathrm{Var}\left(\sum_{i=1}^{n} g(X_i)\right) = n\,\mathrm{Var}\, g(X_1).$$

Twierdzenie.

Niech $X_1, ..., X_n$ będzie próba losową z rozkładu o średniej μ i wariancji $\sigma^2 < \infty$. Wtedy

- $\mathrm{E}\bar{X} = \mu$,
- $\mathrm{Var}\,\bar{X} = \frac{\sigma^2}{n}$,
- $\mathrm{E}S^2 = \sigma^2$.

Dowód.

Niech $g(X_i) = X_i/n$, wtedy $\mathrm{E}g(X_i) = \mu/n$. Na mocy lematu

$$\mathrm{E}\bar{X} = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\,\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\, n\,\mathrm{E}\,X_1 = \mu,$$

co dowodzi pierwszej równości w twierdzeniu.

Dowód cd.

Podobnie dowodzimy równość drugą

$$\mathrm{Var}\bar{X} = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\, n\,\mathrm{Var}\, X_1 = \frac{\sigma^2}{n}.$$

Dowód cd.

Korzystając z twierdzenia 1 dla wariancji z próby, otrzymujemy

$$
\begin{aligned}
\mathrm{E}S^2 &= \mathrm{E}\left(\frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]\right) \\
&= \frac{1}{n-1}(n\,\mathrm{E}\,X_1^2 - n\,\mathrm{E}\,\bar{X}^2) \\
&= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right) \\
&= \frac{1}{n-1}(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2) = \sigma^2.
\end{aligned}
$$

Twierdzenie.

Niech $X_1, ..., X_n$ będzie próba losową z rozkładu normalnego $N(\mu, \sigma^2)$ natomiast

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{oraz} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Wtedy

- $\bar{X}$ *oraz* S^2 *są niezależnymi zmiennymi losowymi,*
- $\bar{X}$ *ma rozkład normalny* $N\left(\mu, \frac{\sigma^2}{n}\right)$,
- $\frac{n-1}{\sigma}S^2$ *ma rozkład* χ^2 *z* $n-1$ *stopniami swobody.*

Rodzina wykładnicza rozkładów

W statystyce matematycznej ważną rolę odgrywają rozkłady prawdopodobieństwa, których gęstość można przedstawić w następującej postaci:

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x)\right).$$

Rodzina wykładnicza rozkładów

W statystyce matematycznej ważną rolę odgrywają rozkłady prawdopodobieństwa, których gęstość można przedstawić w następującej postaci:

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x)\right).$$

Twierdzenie.

Niech $X_1, ..., X_n$ będzie próba losową z rozkładu o gęstości $f(x|\theta)$ postaci

$$f(x|\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)\, t_i(x)\right).$$

Zdefiniujmy statystyki $T_1, ..., T_k$ jako

$$T_i(X_1, ..., X_n) = \sum_{i=1}^{n} t_i(X_j), \quad i = 1, ..., k.$$

Jeśli zbiór $\{(w_1(\theta), w_2(\theta), ..., w_k(\theta)),\ \theta \in \Theta\}$ zawiera otwarty podzbiór $\mathbf{R}^k$, to rozkład wektora losowego $(T_1, ..., T_n)$ jest postaci

$$f_T(u_1, ..., u_k|\theta) = H(u_1, ..., u_k)[c(\theta)]^n \exp\left(\sum_{i=1}^{k} w_i(\theta)\, u_i\right).$$