Estymacja parametrów

Krzysztof Topolski

Wykład 5

Wrocław, 7 marca 2012

Metoda największej wiarogodności

Niech $X_1, X_2, ..., X_n$ będzie próbą losową z rozkładu o gęstości $f(x \mid \theta_1, ..., \theta_k)$ gdzie $\theta_1, ..., \theta_k$ oznaczają parametry rozkładu.

Rozkład łączny próby, czyli rozkład łączny wektora losowego $(X_1, X_2, ..., X_n)$ jest postaci

$$f(x_1 \mid \theta_1, ..., \theta_k) \times ... \times f(x_n \mid \theta_1, ..., \theta_k),$$

co możemy zapisać jako

$$\prod_{i=1}^{n} f(x_i \mid \theta_1, ..., \theta_k).$$

Metoda największej wiarogodności

Niech $X_1, X_2, ..., X_n$ będzie próbą losową z rozkładu o gęstości $f(x \mid \theta_1, ..., \theta_k)$ gdzie $\theta_1, ..., \theta_k$ oznaczają parametry rozkładu.

Rozkład łączny próby, czyli rozkład łączny wektora losowego $(X_1, X_2, ..., X_n)$ jest postaci

$$f(x_1 \mid \theta_1, ..., \theta_k) \times ... \times f(x_n \mid \theta_1, ..., \theta_k),$$

co możemy zapisać jako

$$\prod_{i=1}^{n} f(x_i \mid \theta_1, ..., \theta_k).$$

Metoda największej wiarogodności

Niech $X_1, X_2, \dots, X_n$ będzie próbą losową z rozkładu o gęstości $f(x \,|\, \theta_1, \dots, \theta_k)$ gdzie $\theta_1, \dots, \theta_k$ oznaczają parametry rozkładu.

Rozkład łączny próby, czyli rozkład łączny wektora losowego $(X_1, X_2, \dots, X_n)$ jest postaci

$$f(x_1 \,|\, \theta_1, \dots, \theta_k) \times \dots \times f(x_n \,|\, \theta_1, \dots, \theta_k),$$

co możemy zapisać jako

$$\prod_{i=1}^{n} f(x_i \,|\, \theta_1, \dots, \theta_k).$$

Funkcja wiarogodności

Funkcja wiarogodności

$$
\begin{aligned}
L(\theta \,|\, x) &= L(\theta_1, \dots, \theta_k \,|\, x_1, \dots, x_n) \\
&= \prod_{i=1}^{n} f(x_i \,|\, \theta_1, \dots, \theta_k).
\end{aligned}
$$

Estymator MLE

Definicja.

Dla każdej realizacji $x = (x_1, ..., x_n)$ próby losowej niech $\hat{\theta}(x)$ będzie wartością parametru θ, dla której funkcja $L(\theta|x)$ osiąga swoje maksimum jako funkcja θ z ustalonym x.

Estymatorem największej wiarogodności parametru θ opartym na próbie $(X_1, ..., X_n)$ jest $\hat{\theta}(X_1, ..., X_n)$.

Czasami estymator wyznaczony metodą największej wiarogodności nazywa się w skrócie estymatorem MLE od angielskiego terminu **(Maximum Likelihood Estymator)**.

Estymator MLE

Definicja.

Dla każdej realizacji $x = (x_1, ..., x_n)$ próby losowej niech $\hat{\theta}(x)$ będzie wartością parametru θ, dla której funkcja $L(\theta|x)$ osiąga swoje maksimum jako funkcja θ z ustalonym x.

Estymatorem największej wiarogodności parametru θ opartym na próbie $(X_1, ..., X_n)$ jest $\hat{\theta}(X_1, ..., X_n)$.

Czasami estymator wyznaczony metodą największej wiarogodności nazywa się w skrócie estymatorem MLE od angielskiego terminu **(Maximum Likelihood Estymator)**.

Estymator MLE

Jak wyznaczyć estymator największej wiarogodności?

Jeśli funkcja wiarogodności jest różniczkowalna względem θ to możemy szukać kandydata na estymator MLE wśród wartości θ, dla których

$$\frac{d}{d\theta}L(\theta|x) = L'(\theta|x) = 0.$$

Estymator MLE

Jak wyznaczyć estymator największej wiarogodności?

Jeśli funkcja wiarogodności jest różniczkowalna względem θ to możemy szukać kandydata na estymator MLE wśród wartości θ, dla których

$$\frac{d}{d\theta}L(\theta|x) = L'(\theta|x) = 0.$$

Przykład (Rozkład normalny)

Niech $X_1, ..., X_n$ będzie próbą z rozkładu normalnego $N(\theta, 1)$.

$$
\begin{aligned}
L(\theta|x) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}} \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}
\end{aligned}
$$

Równanie

$$\frac{d}{d\theta} L(\theta|x) = 0$$

sprowadza się do równania

$$\sum_{i=1}^{n}(x_i - \theta) = 0.$$

Przykład cd.

Szukamy rozwiązania równania

$$\sum_{i=1}^{n}(x_i - \theta) = 0.$$

Przekształcając otrzymujemy równanie

$$\sum_{i=1}^{n} x_i - n\theta = 0$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Stąd $\hat{\theta} = \bar{x}$ jest punktem, w którym funkcja wiarogodności może osiągać ekstremum.

Przykład cd.

Szukamy rozwiązania równania

$$\sum_{i=1}^{n}(x_i - \theta) = 0.$$

Przekształcając otrzymujemy równanie

$$\sum_{i=1}^{n} x_i - n\theta = 0$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Stąd $\hat{\theta} = \bar{x}$ jest punktem, w którym funkcja wiarogodności może osiągać ekstremum.

Przykład cd.

Szukamy rozwiązania równania

$$\sum_{i=1}^{n}(x_i - \theta) = 0.$$

Przekształcając otrzymujemy równanie

$$\sum_{i=1}^{n} x_i - n\theta = 0$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Stąd $\hat{\theta} = \bar{x}$ jest punktem, w którym funkcja wiarogodności może osiągać ekstremum.

Przykład cd.

Szukamy rozwiązania równania

$$\sum_{i=1}^{n}(x_i - \theta) = 0.$$

Przekształcając otrzymujemy równanie

$$\sum_{i=1}^{n} x_i - n\theta = 0$$

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Stąd $\hat{\theta} = \bar{x}$ jest punktem, w którym funkcja wiarogodności może osiągać ekstremum.

Przykład cd.

Ponieważ

$$\frac{d^2}{d\theta^2} L(\theta|x)_{\theta=\bar{x}} < 0,$$

oraz nie ma innych kandydatów na maksimum, więc $\bar{X}$ jest globalnym maksimum, a tym samym jest to estymator MLE parametru θ.

Przykład (rozkład 0-1)

Niech $X_1, ..., X_n$ będzie próbą z rozkładu 0-1 z prawdopodobieństwem wartości 1 równym p. Funkcja wiarogodności w tym przypadku ma postać

$$\begin{aligned} L(p|x) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^y(1-p)^{n-y} \end{aligned}$$

gdzie $y = \sum_{i=1}^{n} x_i$.

Przykład cd.

Zamiast rozważania funkcji $L(\theta|x)$ łatwiej jest rozważać jej logarytm, a więc funkcję $\log L(p|x)$.

$$\log L(p|x) = \log(p^y[1-p]^{n-y}) = y\log(p) + (n-y)\log(1-p).$$

Gdy $0 < y < n$, to różniczkując funkcję $\log L(p|x)$ względem p otrzymujemy

$$\frac{d}{dp}\log L(p|x) = y\frac{1}{p} - (n-y)\frac{1}{1-p}.$$

Przyrównując pochodną do zera otrzymujemy równanie

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przykład cd.

Zamiast rozważania funkcji $L(\theta|x)$ łatwiej jest rozważać jej logarytm, a więc funkcję $\log L(p|x)$.

$$\log L(p|x) = \log(p^y[1-p]^{n-y}) = y\log(p) + (n-y)\log(1-p).$$

Gdy $0 < y < n$, to różniczkując funkcję $\log L(p|x)$ względem p otrzymujemy

$$\frac{d}{dp}\log L(p|x) = y\frac{1}{p} - (n-y)\frac{1}{1-p}.$$

Przyrównując pochodną do zera otrzymujemy równanie

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przykład cd.

Zamiast rozważania funkcji $L(\theta|x)$ łatwiej jest rozważać jej logarytm, a więc funkcję $\log L(p|x)$.

$$\log L(p|x) = \log(p^y[1-p]^{n-y}) = y\log(p) + (n-y)\log(1-p).$$

Gdy $0 < y < n$, to różniczkując funkcję $\log L(p|x)$ względem p otrzymujemy

$$\frac{d}{dp}\log L(p|x) = y\frac{1}{p} - (n-y)\frac{1}{1-p}.$$

Przyrównując pochodną do zera otrzymujemy równanie

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przykład cd.

Zamiast rozważania funkcji $L(\theta|x)$ łatwiej jest rozważać jej logarytm, a więc funkcję $\log L(p|x)$.

$$\log L(p|x) = \log(p^y[1-p]^{n-y}) = y\log(p) + (n-y)\log(1-p).$$

Gdy $0 < y < n$, to różniczkując funkcję $\log L(p|x)$ względem p otrzymujemy

$$\frac{d}{dp}\log L(p|x) = y\frac{1}{p} - (n-y)\frac{1}{1-p}.$$

Przyrównując pochodną do zera otrzymujemy równanie

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przykład cd.

Szukamy rozwiązania równania

$$y\frac{1}{p} - (n - y)\frac{1}{1-p} = 0.$$

Przekształcając otrzymujemy kolejno równania

$$\frac{y}{p} = \frac{n-y}{1-p}$$

$$p(n-y) = y(1-p) = y - yp$$

$$pn - py = y - yp$$

$$p = \frac{y}{n} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Przykład cd.

Szukamy rozwiązania równania

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przekształcając otrzymujemy kolejno równania

$$\frac{y}{p} = \frac{n-y}{1-p}$$

$$p(n-y) = y(1-p) = y - yp$$

$$pn - py = y - yp$$

$$p = \frac{y}{n} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Przykład cd.

Szukamy rozwiązania równania

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przekształcając otrzymujemy kolejno równania

$$\frac{y}{p} = \frac{n-y}{1-p}$$

$$p(n-y) = y(1-p) = y - yp$$

$$pn - py = y - yp$$

$$p = \frac{y}{n} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Przykład cd.

Szukamy rozwiązania równania

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przekształcając otrzymujemy kolejno równania

$$\frac{y}{p} = \frac{n-y}{1-p}$$

$$p(n-y) = y(1-p) = y - yp$$

$$pn - py = y - yp$$

$$p = \frac{y}{n} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Przykład cd.

Szukamy rozwiązania równania

$$y\frac{1}{p} - (n-y)\frac{1}{1-p} = 0.$$

Przekształcając otrzymujemy kolejno równania

$$\frac{y}{p} = \frac{n-y}{1-p}$$

$$p(n-y) = y(1-p) = y - yp$$

$$pn - py = y - yp$$

$$p = \frac{y}{n} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Przykład cd.

Łatwo pokazać, że dla $p = \frac{1}{n}\sum_{i=1}^{n} x_i$, funkcja $\log L(p|x)$ osiąga maksimum globalne. Ponieważ funkcja logarytmiczna przy podstawie większej niż 1 jest funkcją rosnącą, to w punkcie $p = \frac{1}{n}\sum_{i=1}^{n} x_i$ maksimum globalne osiąga również funkcja wiarogodności $L(p|x)$.

Stąd

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

jest estymatorem MLE parametru p.

Estymator MLE o ograniczonym zasięgu

Niech $X_1, ..., X_n$ będzie próbą z rozkładu normalnego $N(\theta, 1)$ o której dodatkowo wiemy, że θ jest dodatnia.

Bez tego dodatkowego ograniczenia estymatorem wyznaczonym metodą największej wiarogodności parametru θ jest $\bar{X}$, średnia z próby.

Gdy $\bar{X}$ przyjmuje wartość ujemną otrzymujemy estymator spoza zakresu wartości parametru θ.

Estymator MLE o ograniczonym zasięgu

Niech $X_1, ..., X_n$ będzie próbą z rozkładu normalnego $N(\theta, 1)$ o której dodatkowo wiemy, że θ jest dodatnia.
Bez tego dodatkowego ograniczenia estymatorem wyznaczonym metodą największej wiarogodności parametru θ jest $\bar{X}$, średnia z próby.
Gdy $\bar{X}$ przyjmuje wartość ujemną otrzymujemy estymator spoza zakresu wartości parametru θ.

Estymator MLE o ograniczonym zasięgu

Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu normalnego $N(\theta, 1)$ o której dodatkowo wiemy, że θ jest dodatnia.
Bez tego dodatkowego ograniczenia estymatorem wyznaczonym metodą największej wiarogodności parametru θ jest $\bar{X}$, średnia z próby.
Gdy $\bar{X}$ przyjmuje wartość ujemną otrzymujemy estymator spoza zakresu wartości parametru θ.

Estymator MLE o ograniczonym zasięgu

Gdy $\bar{x}$ jest ujemne, to dla $\theta \geq 0$ funkcja wiarogodności $L(\theta|x)$ jest funkcją malejącą ze względu na θ, a więc osiąga swoje maksimum dla

$$\hat{\theta} = 0.$$

Tak więc w przypadku dodatkowego założenia $\theta \geq 0$ estymatorem MLE parametru θ jest

$$\hat{\theta} = \begin{cases} \hat{X}, & \text{gdy} \quad \hat{X} \geq 0 \\ 0, & \text{gdy} \quad \hat{X} < 0 \end{cases}$$