2. ESTYMACJA PUNKTOWA

Estymacja punktowa polega na oszacowaniu nieznanego parametru θ za pomocą funkcji mierzalnej, której argumetami są elementy próby losowej $X_1, X_2, \ldots, X_n$. Oszacowanie takie będziemy nazywać estymatorem θ i oznaczać $\hat{\theta}$.

Definicja. *Estymatorem (punktowym)* θ, oznaczanym $\hat{\theta}$, nazywamy dowolną funkcję mierzalną próby $t(X_1, X_2, \ldots, X_n)$, która służy do szacowania θ.

Metoda największej wiarygodności

Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu o gęstości $f_\theta(x)$, zaś $x_1, x_2, \ldots, x_n$ - jej realizacją. Wtedy funkcję

$$L(\theta; x_1, x_2, \ldots, x_n) = f_\theta(x_1) f_\theta(x_2) \ldots f_\theta(x_n)$$

nazywamy **funkcją wiarygodności** rozważanego eksperymentu.

Analogicznie, jeśli $X_1, X_2, \ldots, X_n$ jest prostą próbą losową z rozkładu o masie prawdopodobieństwa $p_\theta(x)$, zaś $x_1, x_2, \ldots, x_n$ - jej realizacją, to funkcją wiarygodności nazywamy

$$L(\theta; x_1, x_2, \ldots, x_n) = p_\theta(x_1) p_\theta(x_2) \ldots p_\theta(x_n).$$

Definicja. *Estymator największej wiarygodności*, oznaczany $\hat{\theta}_{NW}$, to wartość parametru θ, która, przy ustalonej realizacji próby $x_1, x_2, \ldots, x_n$, maksymalizuje funkcję wiarygodności:

$$\hat{\theta}_{NW} = \arg\max_{\theta \in \Theta} L(\theta; x_1, x_2, \ldots, x_n). \tag{1}$$

Dla uproszczenia rachunków, maksymalizację funkcji wiarygodności często warto zastąpić maksymalizacją jej logarytmu naturalnego:

$$\hat{\theta}_{NW} = \arg\max_{\theta \in \Theta} \ln L(\theta; x_1, x_2, \ldots, x_n). \tag{2}$$

Ponieważ funkcja $f(x) = \ln x$ jest ściśle rosnąca, wzory (1) i (2) są równoważne.

Przykład 1. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu Poissona $Pois(\lambda)$, $\lambda > 0$. Wyznaczymy estymator największej wiarygodności parametru λ.

Rozkład Poissona $Pois(\lambda)$ to rozkład dyskretny o funkcji masy prawdopodobieństwa danej wzorem

$$p_\lambda(x) = e^\lambda \frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots.$$

Stąd

$$L(\lambda; x_1, x_2, \ldots, x_n) = p_\lambda(x_1) p_\lambda(x_2) \ldots p_\lambda(x_n) = e^{-n\lambda} \frac{\lambda^{x_1 + x_2 + \ldots + x_n}}{x_1! x_2! \ldots x_n!} = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! x_2! \ldots x_n!}$$

i

$$\ln L(\lambda; x_1, x_2, \ldots, x_n) = -n\lambda + \left(\sum_{i=1}^n x_i \right) \ln \lambda - \ln(x_1! x_2! \ldots x_n!).$$

Jeśli funkcja różniczkowalna osiąga w pewnym punkcie ekstremum, to pochodna w tym punkcie się zeruje. Dlatego szukamy λ takiej by $\frac{d}{d\lambda} \ln L(\lambda; x_1, x_2, \ldots, x_n) = 0$. Mamy

$$\frac{d}{d\lambda} \ln L(\lambda; x_1, x_2, \ldots, x_n) = -n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0 \quad \Leftrightarrow \quad \lambda = \frac{\sum_{i=1}^n x_i}{n}.$$

Skoro $\frac{d^2}{d\lambda^2} \ln L(\lambda; x_1, x_2, \ldots, x_n) = -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0$ dla każdego $\lambda \neq 0$ (a nas interesują jedynie $\lambda > 0$), to funkcja $\ln L$ jest wklęsła i jedyny otrzymany przez nas punkt podejrzany o ekstremum, musi być punktem, w którym osiągana jest największa wartość tej funkcji. Zatem szukany estymator to

$$\hat{\lambda}_{NW} = \frac{\sum_{i=1}^n X_i}{n} \overset{ozn.}{=} \bar{X}.$$

Przykład 2. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$. Wyznaczymy estymator największej wiarygodności parametru (μ, σ^2).

Gęstość rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$ dana jest wzorem $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$. Stąd

$$\begin{aligned} L(\mu, \sigma^2; x_1, x_2, \ldots, x_n) &= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right) = \\ &= (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right), \end{aligned}$$

$$\ln L(\mu, \sigma^2; x_1, x_2, \ldots, x_n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2.$$

Jeśli różniczkowalna funkcja dwóch zmiennych osiąga w pewnym punkcie ekstremum, to zerują się w nim pochodne cząstkowe. Dlatego szukamy μ i σ^2 takich by

$$\left.\begin{aligned} \tfrac{\partial}{\partial \mu} \ln L(\mu, \sigma^2; x_1, x_2, \ldots, x_n) &= 0 \\ \tfrac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2; x_1, x_2, \ldots, x_n) &= 0 \end{aligned}\right\}. \tag{3}$$

Liczymy pochodne cząstkowe

$$\frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2; x_1, x_2, \ldots, x_n) = -\frac{1}{2\sigma^2}\sum_{i=1}^n 2(x_i-\mu)(-1) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu),$$

$$\frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2; x_1, x_2, \ldots, x_n) = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i-\mu)^2$$

i z (3) otrzymujemy

$$\left.\begin{aligned} \textstyle\sum_{i=1}^n x_i - n\mu &= 0 \\ \tfrac{1}{\sigma^2}\textstyle\sum_{i=1}^n (x_i-\mu)^2 &= 0 \end{aligned}\right\} \Rightarrow \left.\begin{aligned} &\mu = \tfrac{1}{n}\textstyle\sum_{i=1}^n x_i \\ &\sigma^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^n (x_i-\mu)^2 \end{aligned}\right\}.$$

Można pokazać, że funkcja $\ln L(\mu, \sigma^2; x_1, x_2, \ldots, x_n)$, dla ustalonych wartości $x_1, x_2, \ldots, x_n$, w punkcie $(\frac{1}{n}\sum_{i=1}^n x_i, \frac{1}{n}\sum_{i=1}^n (x_i-\mu)^2)$ rzeczywiście osiąga wartość największą (rachunki te pomijamy). Stąd szukany estymator $(\hat{\mu}_{NW}, \hat{\sigma^2}_{NW})$ jest postaci:

$$\hat{\mu}_{NW} = \bar{X},\ \hat{\sigma^2}_{NW} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Twierdzenie 1. Jeśli $\hat{\theta}$ jest estymatorem największej wiarygodności parametru θ i g jest funkcją mierzalną, to $g(\hat{\theta})$ jest estymatorem największej wiarygodności parametru $g(\theta)$.

Przykład 3. Pokazaliśmy, że w przypadku próby losowej z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$ mamy

$$\hat{\mu}_{NW} = \bar{X},\ \hat{\sigma^2}_{NW} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Z powyższego twierdzenia natychmiast wynika,że estymatorem największej wiarygodności parametru (μ, σ) jest

$$\hat{\mu}_{NW} = \bar{X},\ \hat{\sigma}_{NW} = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}.$$

Metoda największej wiarygodności jest zaimplementowana w R.

```
> install.packages("MASS")
> library("MASS")
> fitdistr(x, densfun, start, lower, upper)
```

gdzie

- `x` to realizacja próby losowej,
- `densfun` to nazwa rozkładu, np. densfun=”normal”, ”log-normal”, ”Poisson”, ”geometric”, ”exponential”,
- `start` to lista z początkowymi ocenami parametrów rozkładu, np.

  ```
  > fitdistr(x=dane, densfun="gamma", start=list(shape=3,scale=0.5));
  ```

  listy tej nie należy podawać dla rozkładów: ”normal”, ”log-normal”, ”Poisson”, ”geometric”, ”exponential”, bo dla nich znane są dokładne wzory na estymatory parametrów otrzymane metodą największej wiarygodności; dla pozostałych rozkładów stosuje się metody aproksymacyjne wyznaczania tych estymatorów.
- `lower` i `upper` to dolne i górne ograniczenia na parametry rozkładu, warto je podawać, gdy używana jest aproksymacyjna metoda wyznaczania estymatorów tych parametrów; np. w przypadku rozkładu gamma zarówno parametr kształtu jak i parametr skali muszą być dodatnie, zatem argument `lower` to wektor złożony z dwóch zer:

  ```
  lower=c(0,0)
  ```