Wykład 2: Parametryczna 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.

Powstaje pytanie jak znajdować funkcę t z definicji estymatora. Znanych jest wiele metod, poniżej przedstawiamy jedną z nich, uznawaną za najlepszą. Jest to tzw. **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 2. 1. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu dwupunktowego z prawdopodobieństwem sukcesu p, gdzie $p \in (0, 1)$. Metodą największej wiarygodności wyznaczymy estymator parametru p.

$$\hat{p}_{NW} = \bar{X}.$$

Przykład 2.2. 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).

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

Twierdzenie 2. 1. Jeśli $\hat{\theta}_{NW}$ jest estymatorem największej wiarygodności parametru θ i g jest funkcją mierzalną, to $g(\hat{\theta}_{NW})$ jest estymatorem największej wiarygodności parametru $g(\theta)$.

Przykład 2.3. 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)`