Wykład 3: Parametryczna estymacja punktowa

Badamy pewną cechę X. Na przykład X to waga ryb danego gatunku. Zuważmy, że X to zmienna losowa – jej wartość nie jest z góry ustalona, ale zależy od wyboru ryby do pomiaru, momentu przeprowadzenia pomiaru itd. (czyli zależy od zdarzenia elementarnego).

Naszym celem jest opisanie rozkładu X. Aby go osiągnąć, pobieramy próbę, którą oznaczamy

$$X_1, X_2, \ldots, X_n.$$

Przed zebraniem danych elementy próby to zmienne losowe. Zakładamy o nich, że mają ten sam rozkład, co badana cecha populacji X. Jeśli dodatkowo przyjmiemy, że są one niezależne, to będziemy mieć prostą próbę losową, często zwaną po prostu próbą losową.

Definicja. Jeśli $X_1, X_2, \ldots, X_n$ są niezależne i mają ten sam rozkład co cecha populacji X, to $X_1, X_2, \ldots, X_n$ nazywamy *(prostą) próbą losową* z X.

Oczywiście założenie, że pracujemy z prostą próbą losową, musi mieć swoje odzwierciedlenie podczas procesu zbierania danych – do próby powinniśmy wybierać niezależne od siebie obserwacje i każda z nich powinna dobrze reprezentować badaną populację.

W wyniku zebrania danych otrzymujemy *realizację próby losowej*, czyli n ustalonych wartości, które oznaczamy $x_1, x_2, \ldots, x_n$.

Na podstawie próby losowej $X_1, X_2, \ldots, X_n$ chcemy opisać rozkład X. Rozważymy podejście parametryczne, w którym zakłada się, że X ma rozkład o dystrybuancie o znanej postaci ale o nieznanych parametrach. W tym wykładzie zajmiemy się estymacją punktową.

Estymacja punktowa polega na oszacowaniu nieznanego parametru rozkładu 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.

METODY WYZNACZANIA ESTYMATORÓW

Powstaje pytanie jak znajdować funkcę t z definicji estymatora. Znanych jest wiele metod, poniżej przedstawiamy dwie z nich.

1. Metoda momentów

Jeśli $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ jest k-wymiarowym wektorem, to wyznaczamy $EX, EX^2, \ldots, EX^k$. Wszystkie te momenty zależą od nieznanych parametrów $\theta_1, \theta_2, \ldots, \theta_k$:

$$\begin{cases} EX & = & \mu_1(\theta_1, \theta_2, \ldots, \theta_k) \\ EX^2 & = & \mu_2(\theta_1, \theta_2, \ldots, \theta_k) \\ \vdots & & \\ EX^k & = & \mu_k(\theta_1, \theta_2, \ldots, \theta_k) \end{cases}.$$

Powyższy układ równań rozwiązujemy ze względu na $\theta_1, \theta_2, \ldots, \theta_k$ (zakładamy, że istnieje dokładnie jedno rozwiązanie):

$$\begin{cases} \theta_1 & = & g_1(EX, EX^2, \ldots, EX^k) \\ \theta_2 & = & g_2(EX, EX^2, \ldots, EX^k) \\ \vdots & & \\ \theta_k & = & g_k(EX, EX^2, \ldots, EX^k) \end{cases}.$$

Następnie momenty teoretyczne $EX, EX^2, \ldots, EX^k$ zamieniamy na momenty empiryczne M_1, M_2, …, M_k, gdzie $M_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$. Uzyskane w ten sposób funkcje to szukane estymatory:

$$\begin{cases} \hat{\theta}_1 &= g_1(M_1, M_2, \ldots, M_k) \\ \hat{\theta}_2 &= g_2(M_1, M_2, \ldots, M_k) \\ \vdots & \\ \hat{\theta}_k &= g_k(M_1, M_2, \ldots, M_k) \end{cases}.$$

Przykład 3.1. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu wykładniczego $Exp(\lambda)$. Wyznaczymy estymator parametru λ stosując metodę momentów.

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

Przykład 3. 2. 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 momentów wyznaczymy estymator parametru p.

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

Przykład 3.3. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$. Wyznaczymy estymator parametru (μ, σ^2) stosując metodę momentów.

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

Pojęcie *estymator wyznaczony metodą momentów* nie jest jednoznaczne. W wyżej opisanej metodzie momenty $EX, EX^2, \ldots, EX^k$ możemy zastąpić przez momenty centralne $E(X - EX)^2$, $E(X - EX)^3$, …, $E(X - EX)^{k+1}$ a momenty empiryczne $M_1, M_2, \ldots, M_k$ - na centralne momenty empiryczne $\tilde{M}_2, \tilde{M}_3, \ldots, \tilde{M}_{k+1}$, gdzie $\tilde{M}_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^r$. Tak otrzymane estymatory też będziemy nazywać estymatorami wyznaczonymi metodą momentów. Ponadto zamiast kolejnych pierwszych momentów możemy użyć innych i znowu otrzymane estymatory nazwiemy estymatorami wyznaczonymi metodą momentów.

2. 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 3.4. Niech $X_1, X_2, \ldots, X_n$ będzie prostą próbą losową z rozkładu wykładniczego $Exp(\lambda)$. Wyznaczymy estymator największej wiarygodności parametru λ.

$$\hat{\lambda}_{NW} = \frac{1}{\bar{X}}$$

Przykład 3. 5. 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 3.6. 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).

Rozwiązanie. 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*

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

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

$$\left.\begin{array}{rcl} \sum_{i=1}^{n} x_i - n\mu &=& 0 \\ \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 &=& n \end{array}\right\} \Rightarrow \left.\begin{array}{l} \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 \end{array}\right\}.$$

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 3.1. 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 3.7. 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", "cauchy"
- `start` to lista z początkowymi ocenami parametrów rozkładu, np.

  ```
  > fitdistr(x=dane, densfun="gamma", start=list(shape=2,rate=0.7))
  ```

  listy tej nie należy podawać dla rozkładów: "normal" , "log-normal", "Poisson", "geometric", "exponential" i innych, dla których znane są dokładne wzory na estymatory parametrów otrzymane metodą największej wiarygodności; dla pozostałych rozkładów stosuje się metody aproksymacyjne wyznaczania tych estymatorów i wtedy listę z początkowymi ocenami parametrów można podawać,
- `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 drugi parametr muszą być dodatnie, zatem argument `lower` to wektor złożony z dwóch zer:

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

Przykład 3. 8. Posługując się pakietem R, wyznaczymy estymatory największej wiarygodności parametrów a i d rozkładu Cauchy'ego o gęstości

$$f(x) = \frac{d}{\pi(d^2 + (x-a)^2)}, \quad \text{gdzie} \quad d > 0, \quad a \in \mathbb{R}, \quad x \in \mathbb{R},$$

dla następującej próby z tego rozkładu:

-1.11, 2.90, -1.10, -3.80, -1.66, -3.07, -3.25, -3.01,
-1.27, -5.86, -14.99, -10.23, -3.91, 99.34, -3.56, -5.39.