# IV Metody Największej Wiarogodności

## ~~1 Estymacja metodą największej wiarogodności~~

Załóżmy, że $X_1, \dots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie z gęstością $f(x,\theta)$, $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$.

**Def 1**

Funkcję $L: \Theta \to \mathbb{R}$ ~~parametru~~ $\theta$ postaci

$$L(\theta) = L(\theta, \underline{x}) = \prod_{i=1}^{n} f(x_i, \theta), \ \theta \in \Theta,$$

gdzie $\underline{x} = (x_1, \dots, x_n)$, nazywamy funkcją wiarogodności.

Funkcję $l: \Theta \to \mathbb{R}$ postaci

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta), \ \theta \in \Theta$$

nazywamy logarytmem funkcji wiarogodności (logwiarogodnością).

**Przykład 1**

Niech $X_1, \dots, X_n$ będzie próbą losową z rozkładu

$$p(x) = \begin{cases} \theta^x (1-\theta)^{1-x}, & x = 0,1, \\ 0, & \text{w p.p.} \end{cases}$$

dla $\theta \in [0,1]$. Dla $x_i = 0,1$, $i = 1, \dots, n$,

$$P(\underline{X} = \underline{x}) = P((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i},$$

Zatem

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

**Problem**: Jaka wartość $\theta$ maksymalizowałaby prawdopodobieństwo $L(\theta)$ otrzymania obserwowanej próby $x_1, \dots, x_n$? [Dobry estymator $\theta$?]

Mamy

$$l(\theta) = \log L(\theta) = \left(\sum x_i\right) \log \theta + \left(n - \sum x_i\right) \log(1-\theta)$$

$$\frac{dl(\theta)}{d\theta} = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0, \quad \theta \neq 0, 1$$

$$(1-\theta)\sum x_i - \theta\left(n - \sum x_i\right) = \sum x_i - \theta n = 0 \Rightarrow \hat{\theta} = \frac{1}{n}\sum x_i = \bar{x}.$$

$$\frac{\partial^2 l(\theta)}{\partial \theta^2} = -\frac{\sum x_i}{\theta^2} - \frac{n - \sum x_i}{(1-\theta)^2}\Big|_{\theta = \bar{x}} < 0$$

$$-\frac{n\bar{x}}{\cancel{\bar{x}^2}} - \frac{n-\cancel{n\bar{x}}}{(1-\bar{x})^2} = -\frac{n}{\bar{x}} - \frac{n}{1-\bar{x}} = \frac{-n+\cancel{n\bar{x}}-\cancel{n\bar{x}}}{\bar{x}(1-\bar{x})} = \frac{-n}{\bar{x}(1-\bar{x})} < 0$$ ~~max~~

Statystyka

$$\hat{\theta} = \bar{X}$$

jest nazywana estymatorem największej wiarogodności parametru $\theta$.

---

Niech $\theta_0$ będzie prawdziwą wartością parametru $\theta$.

Założenia (warunki regularności)

(R0) : Jeżeli $\theta \neq \theta'$, to $f(x, \theta) \neq f(x, \theta')$ dla ~~każdego~~ pewnego ~~$\{x_i, i=1,..,n\}$~~ $x \in \mathbb{R}$.

(R1) : ~~Gęstości $f(x_i, \theta)$, $i=1,..,n$ mają ten sam nośnik dla wszystkich $\theta$~~
Nośnik $f$ nie zależy od $\theta$.

(R2) : $\theta_0 \in \operatorname{int} \Theta$.

prawdziwy model jest asymptotycznie odseparowany od fałszywego

**Tw 1**

Przy założeniach (R0) i (R1)

$$\lim_{n\to\infty} P_{\theta_0}[L(\theta_0, \underline{X}) > L(\theta, \underline{X})] = 1, \text{ dla wszystkich } \theta \neq \theta_0.$$

**Dowód**

1. Mamy $L(\theta_0, \underline{X}) > L(\theta, \underline{X}) \iff \prod_{i=1}^n f(\theta_0, X_i) > \prod_{i=1}^n f(\theta, X_i) \iff$

$$\sum \log f(\theta_0, X_i) > \sum \log f(\theta, X_i) \iff \frac{1}{n}\sum_{i=1}^n \log\left(\frac{f(\theta, X_i)}{f(\theta_0, X_i)}\right) < 0.$$

2. Ponieważ $\log\left(\frac{f(\theta, X_i)}{f(\theta_0, X_i)}\right)$, $i=1,..,n$ są iid, SPWL implikuje

$$\frac{1}{n}\sum_{i=1}^n \log\left[\frac{f(X_i, \theta)}{f(X_i, \theta_0)}\right] \xrightarrow{P} E_{\theta_0}\left[\log\frac{f(X_1, \theta)}{f(X_1, \theta_0)}\right] \overset{\text{nier. Jensena}}{\underset{\text{ściśle wypukła}}{<}} \log E_{\theta_0}\left[\frac{f(X_1, \theta)}{f(X_1, \theta_0)}\right]$$

3. ~~Ponadto,~~

$$= \log \int_{\mathbb{R}} \frac{f(x_1, \theta)}{f(x_1, \theta_0)} f(x_1, \theta_0)\,dx_1 = \log \int_{\mathbb{R}} f(x_1, \theta)\,dx_1 \overset{R2}{=} \log 1 = 0$$

$E_{\cancel{\theta}}$

**Uwaga 1**

Teza twierdzenia mówi, że, asymptotycznie, funkcja wiarogodności jest maksymalizowana w punkcie $\theta_0$ (czyli prawdziwej wartości parametru $\theta$.

**Def 2**

Mówimy, że $\hat{\theta} = \hat{\theta}(\underline{X})$ jest estymatorem największej wiarogodności (ENW) parametru $\theta$, ~~wtedy i tylko~~ jeżeli

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\, L(\theta, \underline{X})$$

czyli
$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta, \underline{X}).$$

**Uwaga 2**

Estymator NW może nie istnieć lub może nie być wyznaczony jednoznacznie

**Przykład 2**

$X_1, \ldots, X_n$ iid $\quad X_i \sim f(x,\theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}x} \mathbb{1}_{(0,+\infty)}(x)$. $\qquad \frac{1}{\theta} e^{-x/\theta}$

$$L(\theta) = \left(\frac{1}{\theta}\right)^n e^{-\frac{1}{\theta}\sum x_i}$$

$$\ell(\theta) = -n\log\theta - \frac{1}{\theta}\sum x_i$$

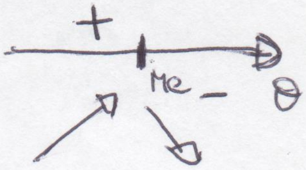
$$\frac{d\ell(\theta)}{d\theta} = \frac{-n}{\theta} + \frac{1}{\theta^2}\sum x_i = 0 \iff n\theta = \sum x_i \iff \underline{\hat{\theta} = \bar{X}}$$

$$\frac{d\ell(\theta)}{d\theta^2} = \frac{1}{\theta^2} - \frac{2}{\theta^3}\sum x_i = \frac{1}{\theta^3}\left(\theta - 2\sum x_i\right)\Bigg|_{\hat{\theta} = \bar{X}} = \frac{1}{\bar{X}^3}(\bar{X} - 2n\bar{X}) < 0.$$

maksimum.

**Przykład 3**

$X_1, \ldots, X_n$ iid $\quad X_i \sim f(x,\theta) = \frac{1}{2} e^{-|x-\theta|}, \quad x, \theta \in \mathbb{R}$.




$$L(\theta) = \left(\frac{1}{2}\right)^n e^{-\sum |x_i - \theta|}$$

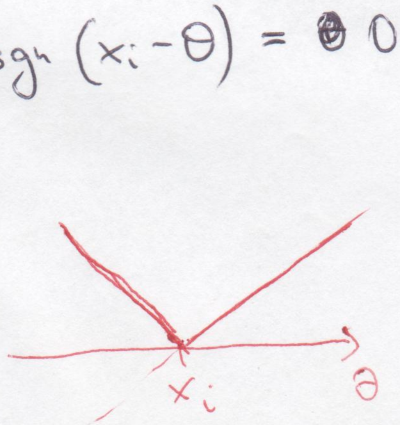
$$\ell(\theta) = -n\log 2 - \sum_{i=1}^{n} |x_i - \theta|$$

$$\ell'(\theta) = \sum_{i=1}^{n} \operatorname{sgn}(x_i - \theta) = 0 \implies \hat{\theta} = \operatorname{med}\{x_1, \ldots, x_n\}$$

$x_i - \theta > 0 \quad 1$

$x_i > \theta \quad 1$

$$\sum_{i=1}^{n} \operatorname{sgn}(x_i - x_{(\frac{n}{2})})$$




$|\theta - x_i|$

## Przykład 4

$X_1,\dots,X_n$ iid, $X_i \sim U(0,\theta)$, $f(x|\theta) = \frac{1}{\theta}\mathbb{1}_{(0,\theta)}(x)$ (R1)! nie jest spełniony

$$L(\theta) = \left(\frac{1}{\theta}\right)^n \prod_{i=1}^n \mathbb{1}_{(0,\theta)}(x_i) = \left(\frac{1}{\theta}\right)^n \cancel{\mathbb{1}_{(0,\theta)}}\, \mathbb{1}_{[0,+\infty)}(x_{(1)})\, \mathbb{1}_{(-\infty,\theta]}(x_{(n)})$$

$$\forall_i \quad 0 \le x_i \le \theta$$

$$0 \le x_{(1)} \ \wedge\ x_{(n)} \le \theta$$

$$\boxed{\hat{\theta} = X_{(n)}}$$

## Tw 2

Niech $X_1,\dots,X_n$ będzie próbą prostą z rozkładu o gęstości $f(x|\theta)$, $\theta \in \Theta$. Niech $g: \Theta \to \mathbb{R}$ będzie dowolną funkcją, a $\eta = g(\theta)$ estymowanym parametrem. Załóżmy, że $\hat{\theta}$ jest ENW parametru $\theta$. Wtedy $g(\hat{\theta})$ jest ENW parametru $\eta = g(\theta)$.

## Dowód

~~I) Załóżmy, że $g$ jest różnowartościowa.~~

~~1. Wtedy~~

$$\cancel{\max_{\eta \in g(\Theta)} L(\eta) = \max_{g(\theta) \in g(\Theta)} L(g(\theta)) = \max_{\eta = g(\theta)} L(\eta) = \max_{\theta \in \Theta} L(g^{-1}(\theta)) = \max_{\eta} L(g^{-1}(\eta))}$$

~~z założenia~~

## Tw 3

Załóżmy, że $X_1,\dots,X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie, ~~których gęstość~~ spełniającymi warunki (R0)-(R2). Ponadto, załóżmy, że $f(x|\theta)$ jest różniczkowalna względem $\theta$. Wtedy, równanie ~~wiarogodności~~

$$\frac{\partial}{\partial\theta} L(\theta) = 0 \quad \text{lub równoważnie} \quad \frac{\partial}{\partial\theta} \ell(\theta) = 0$$

ma rozwiązanie $\hat{\theta}_n$, takie że $\hat{\theta}_n \xrightarrow{P} \theta_0$.

## Tw 4

Przy założeniach tw 3, niech $\hat{\theta}_n$ będzie jedynym rozwiązaniem równania $\frac{\partial}{\partial\theta} L(\theta) = 0$. Wtedy $\hat{\theta}_n$ jest zgodnym estymatorem parametru $\theta_0$.

## 2. ~~Dolne~~ Nierówność ograniczenie Craméra - Rao i asymptotyczna efektywność ENW



Niech $X$ będzie zmienną losową o gęstości $f(x,\theta)$, $\theta \in \Theta$, gdzie $\Theta$ jest zbiorem otwartym.

Założenia (dodatkowe warunki regularności)

(R3) Gęstość $f(x,\theta)$ jest dwukrotnie różniczkowalna względem parametru $\theta$.

(R4) Całkę $\int f(x,\theta)dx$ ~~może b~~ jest dwukrotnie różniczkowalna względem $\theta$ pod znakiem całki.

Uwaga 1

Warunki (R1)-(R4) oznaczają, że parametr $\theta$ nie pojawi się na końcach przedziału, dla którego $f(x,\theta)>0$ oraz, że możemy zamieniać kolejność całkowania z różniczkowania względem $\theta$.

Mamy

$$1 = \int_{-\infty}^{\infty} f(x,\theta)\,dx \qquad \Big/ \frac{\partial}{\partial\theta}$$

$$0 = \int_{-\infty}^{\infty} \frac{\partial f(x,\theta)}{\partial\theta}\,dx$$

równoważnie

$$0 = \int_{-\infty}^{\infty} \frac{\partial f(x,\theta)/\partial\theta}{f(x,\theta)} f(x,\theta)dx = \int_{-\infty}^{\infty} \frac{\partial \log f(x,\theta)}{\partial\theta} f(x,\theta)dx$$

Zatem $$E\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right] = 0$$

Ponadto, różniczkując

$$0 = \int_{-\infty}^{\infty} \frac{\partial \log f(x,\theta)}{\partial\theta} f(x,\theta)dx$$ względem $\theta$ otrzymujemy

$$0 = \int_{-\infty}^{\infty} \frac{\partial^2 \log f(x,\theta)}{\partial\theta^2} f(x,\theta)dx + \int_{-\infty}^{\infty} \frac{\partial \log f(x,\theta)}{\partial\theta} \underbrace{\frac{\partial \log f(x,\theta)}{\partial\theta} f(x,\theta)}_{\frac{\partial f(x,\theta)}{\partial\theta}} dx$$

Równoważnie

$$-\int_{-\infty}^{\infty} \frac{\partial^2 \log f(x,\theta)}{\partial\theta^2} f(x,\theta)dx = E\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right]^2$$

**Def 1**

Informację Fishera dla parametru $\theta$ nazywamy liczbę (wektor)

$$I(\theta) = E\left[\left(\frac{\partial \log f(X,\theta)}{\partial\theta}\right)^2\right].$$

**Wniosek 1** Przy założeniu warunków regularności (R0) – (R4)

$$I(\theta) = -\int_{-\infty}^{\infty} \frac{\partial^2 \log f(x,\theta)}{\partial\theta^2} f(x,\theta)\,dx = Var\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right].$$

**Def 2**

Funkcję

$$\frac{\partial \log f(x|\theta)}{\partial\theta}$$

nazywamy funkcją wynikową.

**Przykład 1**

$X \sim b(1,\theta)$, $\quad f(x|\theta) = \theta^x(1-\theta)^{1-x}$

$$\log f(x,\theta) = x\log\theta + (1-x)\log(1-\theta)$$

$$\frac{\partial \log f(x,\theta)}{\partial\theta} = \frac{x}{\theta} - \frac{1-x}{1-\theta}$$

$$\frac{\partial^2 \log f(x,\theta)}{\partial\theta^2} = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2}.$$

Stąd

$$I(\theta) = -E\left[\frac{-X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\right] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}.$$

**Przykład 2**

$X_1, \dots, X_n$ iid, takie że

$$X_i = \theta + e_i, \quad i = 1, \dots, n, \qquad \text{(model przesunięcia / położenia)}$$

gdzie $e_1, \dots, e_n$ iid $e_i \sim f(x)$. Wtedy $X_i \sim f(x,\theta) = f(x-\theta)$.

Załóżmy, że $f$ spełnia warunki regularności. Wtedy

$$I(\theta) = \int_{-\infty}^{\infty}\left(\frac{f'(x-\theta)}{f(x-\theta)}\right)^2 f(x-\theta)\,dx \overset{\{z = x-\theta,\ dz = dx\}}{=} \int_{-\infty}^{\infty}\left(\frac{f'(z)}{f(z)}\right)^2 f(z)\,dz.$$

Zatem w modelu przesunięcia informacja Fishera nie zależy od $\theta$.

---

Załóżmy, że $X_i$ ma rozkład Laplace'a. $f(x_i,\theta) = \frac{1}{2}e^{-|x_i-\theta|}$

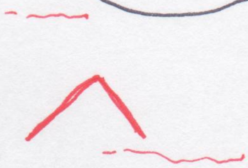
nie spełnia warunków regularności

Ponieważ

$$X_i = \theta + e_i,$$

więc

$$e_i \sim f(z_i) = \frac{1}{2}e^{-|z_i|}$$



Ponadto,

$$f'(z) = -\frac{1}{2}e^{-|z|}\,\mathrm{sgn}(z).$$

Tym samym

$$I(\theta) = \int_{-\infty}^{\infty}\left(\frac{f'(z)}{f(z)}\right)^2 f(z)\,dz = \int_{-\infty}^{\infty} f(z)\,dz = 1.$$

**Uwaga 2**

Jeżeli $X_1,\dots,X_n$ iid, $X_i \sim f(x|\theta)$ oraz $I(\theta)$ jest informacją Fishera $X_1$, to $nI(\theta)$ jest informacją Fishera dla całej próby.

**Dowód**

$$\mathrm{Var}\left(\frac{\partial \log L(\theta,\underline{X})}{\partial\theta}\right) = \mathrm{Var}\left(\sum_{i=1}^{n}\frac{\partial \log f(X_i,\theta)}{\partial\theta}\right) = \sum_{i=1}^{n}\mathrm{Var}\left(\frac{\partial \log f(X_i,\theta)}{\partial\theta}\right) = nI(\theta).$$

**Tw. 1 (Nierówność Craméra-Rao)**

Niech $X_1,\dots,X_n$ będą niezależnymi zmiennymi losowymi z gęstością $f(x,\theta)$, $\theta \in \Theta$. Załóżmy, że warunki regularności (R0) - (R4) są spełnione.

Jeżeli $Y = u(X_1,\dots,X_n)$ jest taką statystyką, że $EY = E[u(X_1,\dots,X_n)] = k(\theta)$, to

$$\mathrm{Var}(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}.$$

**Dowód**

I Załóżmy, że $X_1$ ma rozkład ciągły (dla dyskretnych analogicznie)