Przykład 1

$X_1,\dots,X_n$ iid, $X_i \sim f(x,\theta) = \theta e^{-\theta x}\, \mathbb{1}_{(0,\infty)}(x)$, $\theta>0$.

Chcemy wyznaczyć MVUE parametru θ.

$$L(\theta) = L(\theta, x_1,\dots,x_n) = \theta^n e^{-\theta \sum x_i}$$

(z tw.) Stąd $Y_1 = \sum_{i=1}^{n} X_i$ jest statystyką dostateczną.

$$\ell(\theta) = \log L(\theta) = n\log\theta - \theta\sum x_i$$

$$\ell'(\theta) = \frac{n}{\theta} - \sum x_i = 0 \Rightarrow \theta = \frac{n}{\sum x_i} = \frac{1}{\frac{1}{n}\sum x_i} = \frac{1}{\bar{X}}$$

$$\ell''(\theta) = -\frac{n}{\theta^2} < 0 \Rightarrow \hat{\theta} = \frac{1}{\bar{X}} = Y_2 \quad \text{ENW}$$

Y_2 - jest asymptotycznie nieobciążony

$X_1 \sim \Gamma(1, \frac{1}{\theta})$, stąd $Y_1 \sim \Gamma(n, \frac{1}{\theta})$ oraz

$$E\left[\frac{1}{\bar{X}}\right] = nE\left[\frac{1}{Y_1}\right] = n\int_0^\infty \frac{1}{x}\frac{\theta^n}{\Gamma(n)}x^{n-1}e^{-\theta x}\,dx = \left\{\cancel{\begin{matrix} z=\theta x \\ dz = \theta dx\end{matrix}}\right\} =$$

$$n\int_0^\infty \frac{\theta^n}{\Gamma(n)}x^{n-2}e^{-\theta x}\,dx = n\frac{\Gamma(n-1)\theta}{\Gamma(n)}\int \frac{\theta^{n-1}}{\Gamma(n-1)}x^{n-2}e^{-\theta x}\,dx = \frac{n}{n-1}\theta.$$

Zatem $\frac{n-1}{\sum_{i=1}^{n} X_i}$ jest MVUE parametru θ.

Uwaga 2

Y_2 - estymator θ, $EY_2=\theta$, Y_1 - statystyka dostateczna,

$\varphi(Y_1)=E[Y_2|Y_1]$, $Var\,\varphi(Y_1)\leqslant Var\,Y_1$.

Y_3 - estymator θ, $EY_3=\theta$, Y_3 - nie jest statystyką dostateczną

$\psi(Y_3)=E[\varphi(Y_1)|Y_3]$. $E\psi(Y_3)=\theta$ i $Var\,\psi(Y_3)\leqslant Var\,\varphi(Y_1)$.

Ponieważ Y_3 nie jest statystyką dostateczną, rozkład ~~$\psi(Y_3)$ zależy od θ.~~ warunkowy $\varphi(Y_1)|Y_3$ zależy od θ. Zatem $\psi(Y_3)$ nie jest statystyką ponieważ $\psi(Y_3)$ zależy od θ.

Przykład 2

X_1,X_2,X_3 iid, $X_i\sim Exp(\theta)$, $\theta>0$.

$$(X_1,X_2,X_3)\sim\left(\frac{1}{\theta}\right)^3 e^{-(x_1+x_2+x_3)/\theta},\quad 0<x_i<\infty,\ i=1,2,3.$$

Z tw. o faktoryzacji $Y_1=X_1+X_2+X_3$ jest statystyką dostateczną dla θ.

$$EY_1=E(X_1+X_2+X_3)=3\theta.\ \text{Zatem}\ E\left[\frac{Y_1}{3}\right]=\theta.$$

Niech $Y_2=X_2+X_3$, $Y_3=X_3$. Przekształcenie odwrotne:

$$\begin{aligned}x_1&=y_1-y_2\\ x_2&=y_2-y_3\\ x_3&=y_3\end{aligned}\qquad J=\begin{vmatrix}1&-1&0\\0&1&-1\\0&0&1\end{vmatrix}=1$$

Łączny rozkład Y_1,Y_2,Y_3

$$g(y_1,y_2,y_3)=\begin{cases}\left(\frac{1}{\theta}\right)^3 e^{-y_1/\theta}, & 0<y_3<y_2<y_1<+\infty\\ 0, & \text{w p.p.}\end{cases}$$

Rozkład (Y_1,Y_3)

$$g_{13}(y_1,y_3,\theta)=\int_{y_3}^{y_1} g(y_1,y_2,y_3)\,dy_2=\int_{y_3}^{y_1}\left(\frac{1}{\theta}\right)^3 e^{-y_1/\theta}\,dy_2=\left(\frac{1}{\theta}\right)^3 e^{-y_1/\theta}(y_1-y_3),$$
$$0<y_3<y_1<+\infty$$

Rozkład Y_3

wiadomo bo $x_3=y_3$

$$g_3(y_3,\theta)=\cancel{\int_{y_3}^{\infty}\left(\frac{1}{\theta}\right)^3 e^{-y_1/\theta}(y_1-y_3)\,dy_1}=\frac{1}{\theta}e^{-y_3/\theta},\quad 0<y_3<+\infty$$

Rozkład warunkowy Y_1 pod warunkiem Y_3

$$g_{1|3}(y_1|y_3) = \frac{g_{13}(y_1, y_3, \theta)}{g_3(y_3, \theta)} = \frac{\left(\frac{1}{\theta}\right)^3 (y_1 - y_3) e^{-y_1/\theta}}{\left(\frac{1}{\theta}\right) e^{-y_3/\theta}} = \left(\frac{1}{\theta}\right)^2 (y_1 - y_3) e^{-(y_1 - y_3)/\theta}$$

$$\underbrace{0 < y_3 < y_1 < +\infty}$$

Zatem

$$E\left[\frac{Y_1}{3} \middle| Y_3\right] = E\left[\frac{Y_1 - Y_3}{3} \middle| Y_3\right] + E\left[\frac{Y_3}{3} \middle| Y_3\right] =$$

$$= \left(\frac{1}{3}\right) \int_{y_3}^{\infty} \left(\frac{1}{\theta}\right)^2 (y_1 - y_3)^2 e^{-(y_1 - y_3)/\theta} dy_1 + \frac{\cancel{y_3} \; Y_3}{3} = \left\{\begin{matrix} z = y_1 - y_3 \\ dz = dy_1 \end{matrix}\right\}$$

$$= \left(\frac{1}{3}\right) \underbrace{\int_0^{\infty} \left(\frac{1}{\theta}\right)^2 z^2 e^{-z/\theta} dz}_{\Gamma(3, \theta) \,\|\, \Gamma(3)\theta^3} + \frac{\cancel{y_3} \; Y_3}{3} = \frac{1}{3} \cdot \frac{\Gamma(3)\theta^3}{\theta^2} + \frac{\cancel{y_3} \; Y_3}{3} =$$

$$= \frac{2}{3}\theta + \frac{\cancel{y_3} \; Y_3}{3}.$$

Połóżmy, $\Psi(Y_3) = E\left[\frac{Y_1}{3} \middle| Y_3\right] = \frac{2}{3}\theta + \frac{Y_3}{3}.$

$E\Psi(Y_3) = \theta$ i $Var\,\Psi(Y_3) \leq Var\left(\frac{Y_1}{3}\right)$, ale $\Psi(Y_3)$ nie jest statystyką.

4. ~~Zupełność~~ Statystyki zupełne i jednoznaczność estymatorów nieobciążonych o minimalnej wariancji

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

$$f(x, \theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, 2, \dots, \quad \cancel{0 < \theta}. \quad \theta > 0$$

$Y_1 = \sum_{i=1}^n X_i$ jest statystyką dostateczną dla parametru θ oraz

$$g_1(y_1, \theta) = \frac{(n\theta)^{y_1} e^{-n\theta}}{y_1!}, \quad y_1 = 0, 1, 2, \dots$$

Rozważmy rodzinę rozkładów $\{g_1(y_1, \theta) : \cancel{0 < \theta}\ \theta > 0\}$.

Załóżmy, że $u(Y_1)$ jest funkcją Y_1, taką że $E[u(Y_1)]=0$ dla każdego $\theta>0$. Pokażemy, że ten warunek implikuje, że $u(y_1)=0$ dla $y_1=0,1,2,\dots$ Tzn, że

$$E[u(Y_1)]=0 \text{ dla } \theta>0 \Rightarrow 0=u(0)=u(1)=\dots$$

Dla wszystkich $\theta>0$, mamy

$$0=E[u(Y_1)]=\sum_{y_1=0}^{\infty} u(y_1)\frac{(n\theta)^{y_1}e^{-n\theta}}{y_1!}=e^{-n\theta}\left[u(0)+u(1)\frac{n\theta}{1!}+u(2)\frac{(n\theta)^2}{2!}+\dots\right]$$

Ponieważ $e^{-n\theta}>0$, to

$$0=u(0)+[u(1)n]\theta+\left[\frac{n^2 u(2)}{2}\right]\theta^2+\dots \qquad \text{wielomian}(\theta)$$

[Wielomian znika, to znikają współczynniki]. Zatem

$$u(0)=0,\ u(1)n=0,\ \frac{n^2(u(2))}{2}=0, \dots$$

Tym samym $0=u(0)=u(1)=u(2)=\dots$ // Z wyjątkiem punktów miary 0

<u>Def 1</u>

Niech Z będzie zmienną losową o rozkładzie z rodziny $\{h(z,\theta), \theta\in\Theta\}$.

Jeżeli

$$E_\theta[u(Z)]=0 \text{ dla każdego } \theta\in\Theta \Rightarrow u(z)\equiv 0$$

z wyjątkiem zbioru punktów miary zero,

to rodzinę $\{h(z,\theta),\theta\in\Theta\}$ nazywamy rodziną zupełną.

<u>Przykład 1</u>

$\{h(z,\theta): \theta\in\Theta\}$ rodzina rozkładów zmiennej losowej Z postaci

$$h(z,\theta)=\frac{1}{\theta}e^{-z/\theta}, \quad z>0.$$

$s=\frac{1}{\theta}$ $F(s)=\int_0^\infty e^{-st}f(t)\,dt$

Załóżmy, że $E[u(Z)]$ dla każdego θ. Tzn

$$\int_0^\infty u(z)\frac{1}{\theta}e^{-z/\theta}dz=0=\frac{1}{\theta}\underbrace{\int_0^\infty u(z)e^{-z/\theta}dz}_{\text{Transformata Laplace'a } u(z)} \quad \text{dla } \theta>0.$$

stąd $u(z)=0$

Zatem rodzina rozkładów $\{h(z,\theta),\ \theta>0\}$ jest zupełna.

Niech $X_1,\dots,X_n$ będzie próbą z rozkładu $f(x,\theta)$, a $Y_1 = u_1(X_1,\dots,X_n)$ statystyką dostateczną dla parametru θ. Niech $f_{Y_1}(y_1,\theta)$ będzie rozkładem Y_1. Jeżeli Y_2 jest nieobciążonym estymatorem θ, który nie jest funkcją Y_1, to $\varphi(Y_1) = E[Y_2 | Y_1]$ jest nieobciążonym estymatorem θ. Przypuśćmy, że istnieje ~~inna~~ funkcja $\psi \neq \varphi$, taka że $E[\psi(Y_1)] = \theta$ dla każdego $\theta \in \Theta$. Stąd

$$E[\varphi(Y_1) - \psi(Y_1)] = 0 \quad \text{dla } \theta\in\Theta.$$

Jeżeli rodzina $\{f_{Y_1}(y_1,\theta) : \theta \in \Theta\}$ jest zupełna, to $\varphi(y_1) - \psi(y_1) = 0$ poza zbiorem miary zero. Równoważnie $\varphi(y_1) = \psi(y_1)$ poza zbiorem miary zero. Zatem $\varphi(Y_1)$ jest jedyną funkcją Y_1, ~~która~~ taką że $E[\psi(Y_1)] = \theta$. Z tw. Rao-Blackwella $\varphi(Y_1)$ jest MVUE θ.

Tw. 1 Lehmanna - Scheffé'go

Niech $X_1,\dots,X_n$, n ustalone, będzie próbą z rozkładu $f(x,\theta)$, $\theta\in\Theta$. Niech $Y_1 = u_1(X_1,\dots,X_n)$ będzie statystyką dostateczną dla θ, a $\{f_{Y_1}(y_1,\theta),\ \theta\in\Theta\}$ rodziną zupełną. Jeżeli istnieje funkcja φ taka, że $E\varphi(Y_1)$ jest nieobciążonym estymatorem θ, to $\varphi(Y_1)$ jest ENMW wyznaczonym jednoznacznie. (z dokładnością do zbiorów miary zero)

404

5. Wykładnicza rodzina rozkładów

Rozważmy rodzinę rozkładów $\{f(x,\theta) : \theta\in\Theta\}$, gdzie $\Theta = \{\theta : \gamma<\theta<\delta\}$, gdzie γ i δ są znanymi stałymi (±∞ dopuszczalne), a

$$f(x,\theta) = \begin{cases} \exp[p(\theta)K(x) + S(x) + q(\theta)], & x\in S,\\ 0, & \text{w p.p.}\end{cases}$$

gdzie S jest nośnikiem X.

Def 1

Mówimy, że gęstość $f(x,\theta)$ należy do regularnej rodziny wykładniczej, jeżeli

(i) S, nośnik X, nie zależy od θ,

(ii) $p(\theta)$ jest {nietrywialną} funkcją θ.

(iii) - jeżeli X jest ciągłą zmienną losową, to $K'(x) \not\equiv 0$ i $S(x)$ są ciągłymi funkcjami, $x \in S$

- jeżeli X jest zmienną losową dyskretną, to $K(x)$ jest nietrywialną funkcją $x \in S$.

Przykład 1

(i) Rodzina rozkładów $\{f(x,\theta) : 0 < \theta < \infty\}$, gdzie

$$f(x,\theta) = \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{x^2}{2\theta}} = \exp\left[-\frac{1}{2\theta}x^2 - \log\sqrt{2\pi\theta}\right], \quad x \in \mathbb{R}$$

jest regularną rodziną wykładniczą typu ciągłego.

(ii) Rodzina rozkładów jednostajnych $\{f(x,\theta) ; 0 < \theta < +\infty\}$, gdzie

$$f(x,\theta) = \begin{cases} \exp\{-\log\theta\}, & x \in (0,\theta), \\ 0 & \text{w p.p.}, \end{cases}$$

nie jest regularną rodziną wykładniczą.

Niech $X_1, \dots, X_n$ będzie próbą z regularnej rodziny wykładniczej. Łączny rozkład $X_1, \dots, X_n$ ma postać

$$\exp\left[p(\theta)\sum_{i=1}^n K(x_i) + \sum_{i=1}^n S(x_i) + nq(\theta)\right] \quad \text{dla } x_i \in S,\ i = 1,\dots,n.$$

równoważnie

$$\exp\left[p(\theta)\sum_{i=1}^n K(x_i) + \overset{nq(\theta)}{\cancel{\sum_{i=1}^n S(x_i)}}\right] \cdot \exp\left[\sum_{i=1}^n S(x_i)\right]$$

Zatem, z tw. o faktoryzacji $Y_1 = \sum_{i=1}^n K(X_i)$ jest statystyką dostateczną dla θ.