8 Minimalne statystyki dostateczne i statystyki swobodne

Niech $X_1,\dots,X_n$ będzie próbą z rozkładu $f(x,\theta)$, $\theta\in\mathbb{R}$.

Def 1

Mówimy, ~~że~~ $T=T(X_1,\dots,X_n)$ jest minimalną statystyką dostateczną dla parametru, jeżeli jest ona funkcją każdej innej statystyki dostatecznej ~~jest funkcją T~~.

Uwaga 1

Na ogół minimalna statystyka dostateczna dla jednowymiarowego parametru jest ~~jedną~~ zmienną losową.

Przykład 1

$X_1,\dots,X_n$ iid $\quad X_i\sim U(\theta-1,\theta+1)$ ~~to~~ czyli

$$f(x,\theta)=\frac{1}{2}\mathbb{1}_{(\theta-1,\theta+1)}(x),\quad \theta\in\mathbb{R}.$$

Łączny rozkład $X_1,\dots,X_n$

$$\left(\frac{1}{2}\right)^n\prod_{i=1}^{n}\mathbb{1}_{(\theta-1,\theta+1)}(x_i)=\left(\frac{1}{2}\right)^n\mathbb{1}_{(\theta-1,\cancel{\theta+1}\,\infty)}(\min\{x_i\})\,\mathbb{1}_{(\cancel{\theta-1}\,-\infty,\theta+1)}(\max\{x_i\})$$

ponieważ $\quad \theta-1\leqslant\min\{x_i\}\leqslant x_j\leqslant\max\{x_i\}\leqslant\theta+1,\ j=1,\dots,n.$

Zatem wektor statystyk $\underline{Y}=(Y_1,Y_n)=(\min\{X_i\},\max\{X_i\})$ jest dostateczny dla θ. Jest to minimalna statystyka dostateczna.

Przykład 2

(i) ENW $\hat{\theta} = \bar{X}$ parametru θ w modelu $N(\theta, \sigma^2)$, σ^2 znane jest minimalną statystyką dostateczną dla θ.

(ii) ENW $\hat{\theta} = \bar{X}$ parametru θ w modelu $Pois(\theta)$ jest minimalną statystyką dostateczną dla θ.

(iii) ENW $\hat{\theta} = Y_n = \max\{X_i\}$ parametru θ w modelu $U(0,\theta)$ jest minimalną stat. dost. dla θ.

(iv) wektor ENW $\underbrace{\hat{\underline{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)}_{}$, gdzie $\hat{\theta}_1 = \bar{X}$, $\hat{\theta}_2 = S^2$ parametrów θ_1 i θ_2 w modelu $N(\theta_1, \theta_2)$ jest minimalną statystyką dostateczną dla $\underline{\theta} = (\theta_1, \theta_2)$.

Uwaga 2

Minimalna statystyka dostateczna nie jest wyznaczona jednoznacznie i nie musi mieć związku z ENW.

Przykład 1 (c.d)

$X_1, \dots, X_n$ iid $X_i \sim U(\theta-1, \theta+1)$.

$\underline{Y} = (Y_1, Y_n)$ jest minimalną stat. dost.

Mamy
$$\theta - 1 < Y_1 < Y_n < \theta + 1$$
równoważnie
$$Y_n - 1 < \theta < Y_1 + 1.$$

Zatem ENW może być dowolna statystyka $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ spełniająca warunek
$$Y_n - 1 < \hat{\theta} < Y_1 + 1.$$

Np
$$\hat{\theta} = \frac{Y_n - 1 + Y_1 + 1}{2} = \frac{Y_n + Y_1}{2}.$$

$\hat{\theta}$ nie jest statystyką dostateczną.

Przykład 3

Rozważmy model przesunięcia
$$X_i = \theta + W_i,$$
gdzie $W_1, \dots, W_n$ są niezależnymi zmiennymi losowymi o gęstości $f(w)$ i ciągłej dystrybuancie $F(w)$. Wektor statystyk pozycyjnych $\underline{Y} = (Y_1, \dots, Y_n)$ jest dostateczny i zupełny ~~dla F~~ dla F. Ponadto, jeżeli

(i) $f(w) \sim N(0,1)$, to $\bar{X}$ jest ENW i ENMW parametru θ oraz ~~Dostatym~~
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ jest minimalną statystyką dostateczną dla θ.

(ii) $f(w) = e^{-w} \mathbb{1}_{(0,\infty)}(w)$, to Y_1 jest ENW i minimalną statystyką dost.

(iii) $f(w) \sim$ Logistyczny, to wektor $\underline{Y}$ jest minimalną statystyką dostateczną

(iv) $f(w) \sim$ Laplace'a, to $Q_2 = Me\{x_i\}$ jest ENW, ale nie jest statystyką dostateczną. Wektor $\underline{Y}$ jest minimalną statystyką dostateczną.

<u>Uwaga 2.</u>

Statystyka dostateczna i zupełna jest minimalną statystyką dostateczną, ale implikacja w drugą stronę nie zachodzi.

<u>Przykład 1 c.d</u>

$$E\left[\frac{Y_n - Y_1}{2} - \frac{n-1}{n+1}\right] = 0 \quad \text{dla każdego } \theta.$$

A zatem istnieje $u(y_1, y_n) \not\equiv 0$, taka że $E[u(Y_1, Y_n)] = 0$ dla każdego θ.

<u>Def 4</u> Niech $X_1, \dots, X_n$ będzie próbą z rozkładu $f(x,\theta), \theta \in \Theta$.
Statystykę $T = T(X_1, \dots, X_n)$ nazywamy statystyką swobodną (dla parametru θ) jeśli jej rozkład nie zależy od (parametru) θ.

<u>Przykład 4</u>

$X_1, \dots, X_n$ iid, $X_i \sim N(\theta, 1)$. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ jest statystyką swobodną dla θ.

<u>Przykład 5 (Statystyki niezmiennicze ze względu na przesunięcie)</u>

$X_1, \dots, X_n$ iid ponieważ

$$X_i = \theta + W_i \quad i = 1, \dots, n \qquad \theta \in \mathbb{R},$$

$W_1, \dots, W_n$ iid, $W_i \sim f(w)$, która nie zależy od θ. Wtedy $X_i \sim f(x_i - \theta)$.

Niech $Z = u(X_1, \dots, X_n)$ będzie taką statystyką, że

$$u(x_1 + d, \dots, x_n + d) = u(x_1, \dots, x_n) \quad \text{dla wszystkich } d \in \mathbb{R}.$$

Stąd

$$Z = u(W_1 + \theta, \dots, W_n + \theta) = u(W_1, \dots, W_n) \text{ jest funkcją } W_1, \dots, W_n$$

i nie zależy od θ. Zatem rozkład Z nie zależy od θ.
Statystykę Z spełniającą powyższy warunek nazywamy statystyką

niezmienniczą na przesunięcia.

(5) Np. S^2, $Y_n - Y_1$, $\frac{1}{n}\sum_{i=1}^{n}|X_i - Me(X_i)|$, $X_1+X_2-X_3-X_4$, $X_1+X_3-2X_2$,

$\frac{1}{n}\sum_{i=1}^{n}|X_i - \min\{X_i\}|$ itp.

Przykład 6 (Statystyki niezmiennicze na zmianę skali)

Rozważmy model z parametrem skali postaci

$$X_i = \theta W_i, \quad i=1,\dots,n, \quad \text{gdzie } \theta > 0, \text{ a } W_1,\dots,W_n$$

są niezależnymi zmiennymi losowymi z rozkładu o gęstości $f(w)$, która nie zależy od θ. Wtedy X_i ma gęstość $\frac{1}{\theta}f\left(\frac{x}{\theta}\right)$. Parametr θ nazywamy parametrem skali. Załóżmy, że $Z = u(X_1,\dots,X_n)$ jest taką statystyką, że

$$u(cx_1,\dots,cx_n) = u(x_1,\dots,x_n) \quad \text{dla wszystkich } c > 0.$$

Wtedy

$$Z = u(X_1,\dots,X_n) = u(\theta W_1,\dots,\theta W_n) = u(W_1,\dots,W_n),$$

a rozkład statystyki Z nie zależy od θ. Statystykę Z spełniającą powyższy warunek nazywamy statystyką niezmienniczą ze względu na zmianę skali.

Np. $\frac{X_1}{X_1+X_2}$, $\frac{X_1^2}{\sum_{i=1}^{n}X_i^2}$, $\frac{\min\{X_i\}}{\max\{X_i\}}$, itp.

Przykład 7 (Statystyki niezmiennicze ~~na zmianę~~, ze względu na przesunięcie i zmianę skali)

Rozważmy model ~~ze względu na zmianę~~ z parametrem przesunięcia i skali postaci

$$X_i = \theta_1 + \theta_2 W_i, \quad i=1,\dots,n, \quad \theta_1 \in \mathbb{R}, \theta_2 > 0,$$

$W_1,\dots,W_n$ iid o gęstości $f(w)$, która nie zależy od θ_1 i θ_2. Wtedy X_i ma gęstość $\frac{1}{\theta_2}f\left(\frac{x-\theta_1}{\theta_2}\right)$. Rozważmy statystykę $Z = u(X_1,\dots,X_n)$, gdzie

$$u(cx_1+d,\dots,cx_n+d) = u(x_1,\dots,x_n).$$

Wtedy

$$Z = u(X_1, \ldots, X_n) = u(\theta_1+\theta_2 W_1, \ldots, \theta_1+\theta_2 W_n) = u(W_1, \ldots, W_n),$$

a rozkład Z nie zależy od θ_1 i θ_2. Statystykę Z spełniającą powyższy warunek nazywamy statystyką niezmienniczą ze względu na przesunięcie i zmianę skali.

Np: $\dfrac{\max\{X_i\} - \min\{X_i\}}{S}$, $\dfrac{\sum_{i=1}^{n-1}(X_{i+1}-X_i)^2}{S^2}$, $\dfrac{(X_i - \bar{X})}{S}$, $\dfrac{|X_i - X_j|}{S}$, $i \neq j$.

9 Dostateczność, zupełność i niezależność

Jeżeli Y_1 jest statystyką dostateczną dla θ, $\theta \in \Theta$, to $h(z|y_1)$ gęstość warunkowa (innej) statystyki Z przy ustalonym $Y_1 = y_1$ nie zależy od θ. Jeżeli, ponadto, Y_1 i Z są niezależne, to gęstość ~~$g_2(z)$ zm. los.~~ Z, $g_2(z) = h(z|y_1)$ i oczywiście nie zależy od θ. Zatem niezależność statystyki Z i ~~stat Y_1~~ statystyki dostatecznej Y_1 (dla parametru θ) oznacza, że rozkład Z nie zależy od θ.

Tzn Z jest statystyką swobodną.

Z drugiej strony, załóżmy, że Z jest statystyką ~~dostateczną~~ swobodną dla θ, a Y_1 statystyką dostateczną dla θ. Czy Z i Y_1 są niezależne?

~~Mamy~~ Niech $g_1(y_1, \theta)$ ~~będzie~~ będzie rozkładem Y. Rozkład brzegowy Z ma postać

$$\int_{-\infty}^{\infty} g_1(y_1,\theta)\, h(z|y_1)\, dy_1 = g_2(z)$$

i, ~~ai~~, z założenia, nie zależy od θ. ~~Ponieważ~~ ~~Jeżeli~~ ~~Ponieważ~~

Ponieważ

$$\int_{-\infty}^{\infty} g_1(y_1,\theta)\, g_2(z)\, dy_1 = g_2(z)$$

~~!~~

zachodzi, to

$$\int_{-\infty}^{\infty} \left[g_2(z) - h(z|y_1)\right] g_1(y_1,\theta)\, dy_1 = 0 \quad \text{dla każdego } \theta \in \Theta.$$

Z założenia $g_2(z) - h(z|y_1)$ nie zależy od θ. (Z swobodna, Y_1 dostat.)

Jeżeli rodzina $\{g_1(y_1,\theta) : \theta \in \Theta\}$ jest zupełna, to

$$g_2(z) - h(z|y_1) = 0 \quad \text{lub} \quad g_2(z) = h(z|y_1).$$

Tzn. łączny rozkład Y_1 i Z

$$g_1(y_1,\theta)\cdot h(z|y_1) = g_1(y_1,\theta)\, g_2(z).$$

Zatem Y_1 i Z są niezależne. Tym samym zachodzi następujące twierdzenie.

Twierdzenie 1 (Basu)

Niech $X_1,\ldots,X_n$ będzie próbą z rozkładu $f(x,\theta)$, $\theta\in\Theta$, gdzie Θ jest przedziałem. Załóżmy, że Y_1 jest zupełną statystyką dostateczną dla θ. Niech $Z = u(X_1,\ldots,X_n)$ będzie statystyką, która nie jest funkcją Y_1. Jeżeli rozkład Z nie zależy od θ, to Z i Y_1 są niezależne.

Uwaga 1

Twierdzenie zachodzi również dla $\underline{\theta}\in\Theta\subseteq\mathbb{R}^p$, $p>1$.

Przykład 1

$X_1,\ldots,X_n$ iid, $X_i\sim N(\mu,\sigma^2)$

$\bar{X}$ – statystyka dostateczna i zupełna dla μ (przy ustalonym znanym σ^2)

S^2 – statystyka swobodna dla μ (ponieważ niezmiennicza na przesunięcia)

Zatem $\bar{X}$ i S^2 niezależne.

Przykład 2

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

Mamy $f(x,\theta) = f(x-\theta)$, gdzie $f(\omega) = e^{-\omega}\mathbb{1}_{(0,\infty)}(\omega)$.

$Y_1 = \min\{X_i\}$ jest ~~zupeł~~ statystyką dostateczną i zupełną dla θ.
Zatem Y_1 jest niezależna od każdej statystyki, która jest niezmiennicza na przesunięcie. Np. S^2, $Y_n - Y_1$, $\frac{1}{n}\sum_{i=1}^{n}(X_i - \min\{X_i\})$

Przykład 3

X_1, X_2 niezależne z rozkładu $f(x,\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}\mathbb{1}_{(0,\infty)}(x)$, $\theta>0$.

Mamy $f(x,\theta) = \frac{1}{\theta}f\left(\frac{x}{\theta}\right)$, gdzie $f(\omega) = e^{-\omega}\mathbb{1}_{(0,\infty)}(\omega)$

$Y_1 = X_1 + X_2$ jest zupełną statystyką dostateczną dla θ. Zatem Y_1 jest niezależna od każdej statystyki, która jest niezmiennicza na zmianę skali. Np. $\frac{X_1}{X_2}$, $\frac{X_1}{X_1+X_2}$

// F/ Beta

Przykład 4

$X_1, \dots, X_n$ iid, $X_i \sim N(\theta_1, \theta_2)$, $\theta_1 \in \mathbb{R}$, $\theta_2 > 0$.

$(\bar{X}, S^2)$ - statystyka dostateczna i zupełna dla (θ_1, θ_2).

Rozważmy statystykę

$$Z = \frac{\sum_{i=1}^{n-1} (X_{i+1} - X_i)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = u(X_1, \dots, X_n),$$

która spełnia warunek $u(cx_1 + d, \dots, cx_n + d) = u(x_1, \dots, x_n)$ dla $c > 0, d \in \mathbb{R}$.

Zatem Z jest statystyką swobodną dla (θ_1, θ_2). Tym samym

$Z \perp (\bar{X}, S^2)$.