

# Estymatory

SW (na podstawie materiałów W. Kotłowskiego)

# Definicje

Estymator  $\hat{\theta}$  parametru  $\theta$  to funkcja próby  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ , która szacuje nieznaną wartość parametru rozkładu  $\theta$ .

Estymator  $\hat{\theta}$  jako funkcja próby jest **zmienną losową**, ma więc swój rozkład, wartość oczekiwaną, wariancję, itp.

## Estymator nieobciążony

Estymator  $\hat{\theta}$  parametru  $\theta$  jest **nieobciążony**, jeśli

$$E[\hat{\theta}] = \theta,$$

czyli obciążenie  $E[\hat{\theta}] - \theta$  jest równe 0.

**Zadanie:** Pokaż, że estymator  $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  wartości oczekiwanej  $\mu$  jest nieobciążony.

## Estymator nieobciążony

Estymator  $\hat{\theta}$  parametru  $\theta$  jest **nieobciążony**, jeśli

$$E[\hat{\theta}] = \theta,$$

czyli obciążenie  $E[\hat{\theta}] - \theta$  jest równe 0.

**Zadanie:** Pokaż, że estymator  $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  wartości oczekiwanej  $\mu$  jest nieobciążony.

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \underbrace{E[X_i]}_{\mu} = \mu.$$

## Estymator obciążony

**Zadanie:** Wykaż, że estymator wariancji  $\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  jest obciążony.

## Estymator obciążony

**Zadanie:** Wykaż, że estymator wariancji  $\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  jest obciążony.

Przypomnienie...

$$\sigma^2 = E[X^2] - E^2[X]$$

$$E[X^2] = \sigma^2 + E^2[X]$$

## Estymator obciążony

**Zadanie:** Wykaż, że estymator wariancji  $\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  jest obciążony.

$$\begin{aligned} E \left[ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right] &= E \left[ \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \right] \\ &= E \left[ \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X} \frac{1}{n} \sum_{i=1}^n X_i + \bar{X}^2 \right] \\ &= E \left[ \frac{1}{n} \sum_{i=1}^n X_i^2 - 2\bar{X}^2 + \bar{X}^2 \right] = E \left[ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right] \end{aligned}$$

## Estymator obciążony

**Zadanie:** Wykaż, że estymator wariancji  $\hat{s}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  jest obciążony.

$$\begin{aligned} E \left[ \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 \right] &= \frac{1}{n} \sum_{i=1}^n \underbrace{E[X_i^2]}_{\sigma^2 + E^2[X_i] = \sigma^2 + \mu^2} - \underbrace{E[\bar{X}^2]}_{\frac{\sigma^2}{n} + E^2[\bar{X}] = \frac{\sigma^2}{n} + \mu^2} \\ &= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n} \sigma^2 \neq \sigma^2 \end{aligned}$$

Obciążenie estymatora  $\hat{s}^2$  ( $b(\theta) = E[\hat{\theta}] - \theta$ ) jest zatem mniejsze od 0.



## Efektywność estymatorów

Weźmy dwa **nieobciążone** estymatory  $\hat{\theta}_1$  i  $\hat{\theta}_2$  parametru  $\theta$ .

Mówimy, że  $\hat{\theta}_1$  jest efektywniejszy  $\hat{\theta}_2$ , jeśli  $D^2[\hat{\theta}_1] \leq D^2[\hat{\theta}_2]$  przy każdej wartości parametru  $\theta$ .

Estymator, który jest efektywniejszy od wszystkich nieobciążonych estymatorów nazywamy estymatorem **efektywnym**.

# Zgodność estymatorów

Estymator  $\hat{\theta}$  parametru  $\theta$  jest **zgodny**, jeśli:

$$\hat{\theta} \xrightarrow{n \rightarrow \infty} \theta$$

## Estymator przedziałowy

Zamiast pojedynczego estymatora  $\hat{\theta}$  tworzy się **przedział ufności**  $[\hat{\theta}_L, \hat{\theta}_R]$  tak, aby z prawdopodobieństwem  $1 - \alpha$  (zwanym **poziomem ufności**) pokrył prawdziwą wartość parametru  $\theta$ :

$$\Pr(\hat{\theta}_L \leq \theta \leq \hat{\theta}_R) = 1 - \alpha.$$

Przedział ufności tworzy się zwykle poprzez rozpięcie przedziału wokół estymatora punktowego  $\hat{\theta}$ :

$$\hat{\theta}_L = \hat{\theta} - \Delta, \quad \hat{\theta}_R = \hat{\theta} + \Delta.$$

## Estymator przedziałowy

**Zadanie:** Znajdź estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znanym odchyleniem standardowym  $\sigma$ .

## Estymator przedziałowy

**Zadanie:** Znajdź estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znanym odchyleniem standardowym  $\sigma$ .

Bierzemy estymator punktowy  $\hat{\mu} = \bar{X}$  i rozpinamy przedział  $[\bar{X} - \Delta, \bar{X} + \Delta]$ . Wyznaczamy  $\Delta$  z zależności:

$$\begin{aligned}1 - \alpha &= \Pr(\bar{X} - \Delta \leq \mu \leq \bar{X} + \Delta) \\&= \Pr(-\Delta \leq \mu - \bar{X} \leq \Delta) = \Pr(-\Delta \leq \bar{X} - \mu \leq \Delta) \\&= \Pr\left(-\frac{\Delta}{\sigma}\sqrt{n} \leq \underbrace{\frac{\bar{X} - \mu}{\sigma}\sqrt{n}}_{Z \sim N(0,1)} \leq \frac{\Delta}{\sigma}\sqrt{n}\right) \\&= \Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{\Delta}{\sigma}\sqrt{n}\right) = 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1.\end{aligned}$$

## Estymator przedziałowy

**Zadanie:** Znajdź estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znanym odchyleniem standardowym  $\sigma$ .

$$\begin{aligned}1 - \alpha &= 2\Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) - 1 \iff \Phi\left(\frac{\Delta}{\sigma}\sqrt{n}\right) = 1 - \frac{\alpha}{2} \\ &\iff \frac{\Delta}{\sigma}\sqrt{n} = \underbrace{\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}_{z_{\frac{\alpha}{2}}} \\ &\iff \Delta = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.\end{aligned}$$

## Estymator przedziałowy

**Zadanie:** Znajdź estymator przedziałowy dla wartości oczekiwanej  $\mu$  gdy dane pochodzą z rozkładu normalnego  $N(\mu, \sigma)$  ze znanym odchyleniem standardowym  $\sigma$ .

Zatem estymator przedziałowy ma postać:

$$\left[ \bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right],$$

gdzie  $z_{\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$  jest **kwantylem** rzędu  $1 - \frac{\alpha}{2}$ .