Estymatory

Na podstawie materiałów WK

Definicje

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

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

Estymator nieobciążony

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

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

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

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

Estymator nieobciążony

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

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

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

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

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

Estymator obciążony

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

Estymator obciążony

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

$$\sigma^2 = E\left[X^2\right] - E^2\left[X\right]$$
$$E\left[X^2\right] = \sigma^2 + E^2\left[X\right]$$

Estymator obciążony

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

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

Estymator obciążony

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

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E\left[X_i^2\right]}_{\sigma^2+E^2[X_i]=\sigma^2+\mu^2} - \underbrace{E\left[\overline{X}^2\right]}_{\frac{\sigma^2}{n}+E^2\left[\overline{X}\right]=\frac{\sigma^2}{n}+\mu^2}$$

$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

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

Efektywność estymatorów

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

Mówimy, że $\hat{\theta}_1$ jest efektywniejszy[1] $\hat{\theta}_2$, jeśli $D^2[\hat{\theta}_1] \leq D^2[\hat{\theta}_2]$ przy każdej wartości parametru θ.

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

[1] ściślej, powinno być: „co najmniej tak efektywny jak".

Zgodność estymatorów

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

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

Estymator przedziałowy

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

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

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

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

Estymator przedziałowy

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

Estymator przedziałowy

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

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

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

Estymator przedziałowy

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

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

Estymator przedziałowy

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

Zatem estymator przedziałowy ma postać:

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

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