Estymatory

Wojciech Kotłowski

Statystyka i analiza danych 2019/2020

31.03.2020

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.

Parametr θ **nie jest** zmienną losową! (niby oczywiste, ale. . .)

Estymator nieobciążony

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

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

czyli obciążenie $\mathbb{E}[\hat{\theta}] - \theta$ 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

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

czyli obciążenie $\mathbb{E}[\hat{\theta}] - \theta$ 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.

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

Estymator nieobciążony

Zadanie: Dla jakiego c estymator $\hat{\sigma}^2 = c\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony?

Estymator nieobciążony

Zadanie: Dla jakiego c estymator $\hat{\sigma}^2 = c\sum_{i=1}^n (X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony?
Najpierw policzymy:

$$\begin{aligned}
\mathbb{E}\big[(X_i - \overline{X})^2\big] &= \mathbb{E}\big[(X_i - \mu + \mu - \overline{X})^2\big] \\
&= \mathbb{E}\big[(X_i - \mu)^2 + 2(X_i - \mu)(\mu - \overline{X}) + (\mu - \overline{X})^2\big] \\
&= \underbrace{\mathbb{E}\big[(X_i - \mu)^2\big]}_{=\sigma^2} + 2\mathbb{E}\big[(X_i - \mu)\underbrace{(\mu - \overline{X})}_{=\frac{1}{n}\sum_{j=1}^n(\mu - X_j)}\big] + \underbrace{\mathbb{E}\big[(\overline{X} - \mu)^2\big]}_{=\frac{\sigma^2}{n}} \\
&= \sigma^2 + \frac{2}{n}\sum_{j=1}^n \mathbb{E}\big[(X_i - \mu)(\mu - X_j)\big] + \frac{\sigma^2}{n} \\
&= \sigma^2 - \frac{2}{n}\underbrace{\mathbb{E}\big[(\mu - X_i)^2\big]}_{=\sigma^2} + \frac{2}{n}\sum_{j \neq i}\underbrace{\mathbb{E}\big[(X_i - \mu)\big]\mathbb{E}\big[(\mu - X_j)\big]}_{=0} + \frac{\sigma^2}{n} \\
&= \sigma^2 - \frac{2}{n}\sigma^2 + \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.
\end{aligned}$$

Estymator nieobciążony

Zadanie: Dla jakiego c estymator $\hat{\sigma}^2 = c\sum_{i=1}^n (X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony?

$$\mathbb{E}\big[(X_i - \overline{X})^2\big] \;=\; \frac{n-1}{n}\sigma^2.$$

Stąd:

$$\mathbb{E}[\hat{\sigma}^2] \;=\; c\sum_{i=1}^n \mathbb{E}\big[(X_i - \overline{X})^2\big] \;=\; cn\frac{n-1}{n}\sigma^2 \;=\; c(n-1)\sigma^2.$$

Przyjmując $c = \frac{1}{n-1}$, $\hat{\sigma}^2$ jest nieobciążonym estymatorem σ^2.

Efektywność estymatorów

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

Mówimy, że $\hat{\theta}_1$ jest efektywniejszy[1] niż $\hat{\theta}_2$, 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:

$$\hat{\theta} \overset{n\to\infty}{\longrightarrow} \theta$$

(słabo zgodny, gdy zbieżność według prawdopodobieństwa; silnie zgodny, gdy zbieżność z prawdopodobieństwem jeden).

Estymator przedziałowy

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

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

Pytanie: co tu jest losowe?

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 $N(\mu, \sigma^2)$ ze znaną wariancją σ^2.

Estymator przedziałowy

Zadanie: Znajdź estymator przedziałowy dla wartości oczekiwanej μ gdy dane pochodzą z rozkładu normalnego $N(\mu, \sigma^2)$ ze znaną wariancją σ^2.

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

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

Estymator przedziałowy

Zadanie: Znajdź estymator przedziałowy dla wartości oczekiwanej μ gdy dane pochodzą z rozkładu normalnego $N(\mu, \sigma^2)$ ze znaną wariancją σ^2.

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

Czyli estymator przedziałowy ma postać:

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

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