Estymatory

Statystyka i analiza danych 2017/2018

Jurek Błaszczyński

25 marca 2018

Estymator

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 $\hat{\theta}$ jest więc (jako funkcja próby) **zmienną losową**.

Oznacza to, że ma swój rozkład, wartość oczekiwaną, wariancję, itp. (jak każda zmienna losowa).

θ nie jest zmienną losową.
Jest parametrem (nieznanym), który chcemy oszacować (estymować).

Estymator nieobciążony

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

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

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

Estymator nieobciążony

Przykład 1

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

Estymator nieobciążony

Przykład 1

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

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

Estymator nieobciążony

Przykład 2

Pokażmy, że estymator $\hat{\sigma}^2 = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest obciążony.

Estymator nieobciążony

Przykład 2

Pokażmy, że estymator $\hat{\sigma}^2 = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest obciążony.

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 - \overline{X}^2 \tag{1}$$

Estymator nieobciążony

Przykład 2

Pokażmy, że estymator $\hat{\sigma}^2 = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest obciążony.

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 - \overline{X}^2 \tag{1}$$

$$\begin{aligned} \mathbb{E}[S_n^2] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^2] - \mathbb{E}[\overline{X}^2] = & (2) \\ &= \frac{1}{n} n\mathbb{E}[X^2] - \mathbb{E}[\overline{X}^2] = \mathbb{E}[X^2] - \mathbb{E}[\overline{X}^2] & (3) \end{aligned}$$

Estymator nieobciążony

Przykład 2

Pokażmy, że estymator $\hat{\sigma}^2 = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest obciążony.

$$\mathbb{E}[S_n^2] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i^2] - \mathbb{E}[\overline{X}^2] = \tag{4}$$

$$= \frac{1}{n} n\mathbb{E}[X^2] - \mathbb{E}[\overline{X}^2] = \mathbb{E}[X^2] - \mathbb{E}[\overline{X}^2] \tag{5}$$

ponieważ $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$$\mathbb{E}[S_n^2] = Var(X) + (\mathbb{E}[X])^2 - Var(\overline{X}) - (\mathbb{E}[\overline{X}])^2 \tag{6}$$

Estymator nieobciążony

Przykład 2

Pokażmy, że estymator $\hat{\sigma}^2 = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest obciążony.

$$\mathbb{E}[S_n^2] = Var(X) + (\mathbb{E}[X])^2 - Var(\overline{X}) - (\mathbb{E}[\overline{X}])^2 = \tag{7}$$

$$= \sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2 = \tag{8}$$

$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 \tag{9}$$

Estymator nieobciążony

Przykład 3

Pokażmy, że estymator $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony.

Estymator nieobciążony

Przykład 3

Pokażmy, że estymator $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}^2 \qquad (10)$$

Estymator nieobciążony

Przykład 3

Pokażmy, że estymator $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony.

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}^2 \tag{10}$$

$$\begin{aligned}
\mathbb{E}[S_n^2] &= \mathbb{E}\Big[\frac{1}{n-1}\sum_{i=1}^{n} X_i^2 - \frac{n}{n-1}\overline{X}^2\Big] = && (11)\\
&= \frac{1}{n-1}\Big(\sum_{i=1}^{n}\mathbb{E}[X_i^2] - n\mathbb{E}[\overline{X}^2]\Big) = && (12)\\
&= \frac{1}{n-1}\Big(n\mathbb{E}[X_i^2] - n\mathbb{E}[\overline{X}^2]\Big) && (13)
\end{aligned}$$

Estymator nieobciążony

Przykład 3

Pokażmy, że estymator $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony.

$$\mathbb{E}[S_n^2] = \frac{1}{n-1}\Big(n\mathbb{E}[X_i^2] - n\mathbb{E}[\overline{X}^2]\Big) \tag{14}$$

ponieważ $Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

$$\mathbb{E}[S_n^2] = \frac{n}{n-1}\Big(Var(X) + (\mathbb{E}[X])^2 - Var(\overline{X}) - (\mathbb{E}[\overline{X}])^2\Big) \tag{15}$$

Estymator nieobciążony

Przykład 3

Pokażmy, że estymator $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ wariancji σ^2 jest nieobciążony.

$$\mathbb{E}[S_n^2] = \frac{n}{n-1}\Big(Var(X) + (\mathbb{E}[X])^2 - Var(\overline{X}) - (\mathbb{E}[\overline{X}])^2\Big) = \tag{16}$$

$$= \frac{n}{n-1}\Big(\sigma^2 + \mu^2 - \frac{\sigma^2}{n} - \mu^2\Big) = \frac{n}{n-1}\Big(\sigma^2 - \frac{\sigma^2}{n}\Big) = \tag{17}$$

$$= \frac{n\sigma^2}{n-1} - \frac{\sigma^2}{n-1} = \frac{n\sigma^2 - \sigma^2}{n-1} = \frac{\sigma^2(n-1)}{(n-1)} = \sigma^2 \tag{18}$$

Estymator przedziałowy

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

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

Co tu jest losowe?

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

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

Estymator przedziałowy

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

$$X \sim N(\mu, \sigma), \overline{X} \sim N\Big(\mu, \frac{\sigma}{\sqrt{n}}\Big)$$

Estymator przedziałowy

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

$$X \sim N(\mu, \sigma), \overline{X} \sim N\Big(\mu, \frac{\sigma}{\sqrt{n}}\Big)$$

Tworzymy przedział ufności $[\overline{X} - \Delta, \overline{X} + \Delta]$ taki, że:

$$P(\overline{X} - \Delta \leq \mu \leq \overline{X} + \Delta) = 1 - \alpha$$

Estymator przedziałowy

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

$$P(\overline{X} - \Delta \leq \mu \leq \overline{X} + \Delta) = P(-\Delta \leq \overline{X} - \mu \leq \Delta)$$

Estymator przedziałowy

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

$$P(\overline{X} - \Delta \leq \mu \leq \overline{X} + \Delta) = P(-\Delta \leq \overline{X} - \mu \leq \Delta)$$

Doprowadźmy do tego aby wyrażenie “w środku” nierówności miało rozkład N(0,1)

$$P\Big(- \frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\Delta}{\sigma}\sqrt{n}\Big)$$

Estymator przedziałowy

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

$$P(\overline{X} - \Delta \leq \mu \leq \overline{X} + \Delta) = P(-\Delta \leq \overline{X} - \mu \leq \Delta)$$

Doprowadźmy do tego aby wyrażenie "w środku" nierówności miało rozkład N(0,1)

$$P\Big(- \frac{\Delta}{\sigma}\sqrt{n} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\Delta}{\sigma}\sqrt{n}\Big)$$

Wprowadźmy $z_\alpha = \frac{\Delta}{\sigma}\sqrt{n}$

$$P\Big(- z_\alpha \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_\alpha\Big)$$

Estymator przedziałowy

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

$$P\Big(-z_\alpha \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq z_\alpha\Big) = \Phi(z_\alpha) - (1-\Phi(z_\alpha)) = 2\Phi(z_\alpha) - 1$$

Estymator przedziałowy

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

$$P\Big(- z_\alpha \leq \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \leq z_\alpha \Big) = \Phi(z_\alpha) - (1 - \Phi(z_\alpha)) = 2\Phi(z_\alpha) - 1$$

Możemy ustalić ile wynosi z_α

$$2\Phi(z_\alpha) - 1 = 1 - \alpha$$
$$\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$$

Estymator przedziałowy

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

$$P\Big(-z_\alpha \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le z_\alpha\Big) = \Phi(z_\alpha) - (1 - \Phi(z_\alpha)) = 2\Phi(z_\alpha) - 1$$

Możemy ustalić ile wynosi z_α

$$2\Phi(z_\alpha) - 1 = 1 - \alpha$$
$$\Phi(z_\alpha) = 1 - \frac{\alpha}{2}$$

Stąd $z_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$. $z_\alpha = \frac{\Delta}{\sigma}\sqrt{n}$, czyli $\Delta = z_\alpha \frac{\sigma}{\sqrt{n}}$.

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 $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$ przy każdej wartości parametru θ.

Estymator nieobciążony o minimalnej wariancji, czyli 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 gdy $n \to \infty$, to $\hat{\theta} \to \theta$, czyli estymator zbiega[2] do prawdziwej wartości parametru θ.

[2] Jeśli zbieżność jest prawie na pewno (z prawdoopodobieństwem 1), to estymator nazywa się silnie zgodnym, gdy zbieżność jest według prawdopodobieństwa, to estymator nazywa się słabo zgodnym.