Estymacja parametrów

Krzysztof Topolski

Wykład 7

Wrocław, 28 listopada 2019

Niech $X_1, X_2, ..., X_n$ będzie próbą losową z rozkładu o gęstości $f(x \,|\, \theta)$ gdzie θ oznacza parametr rozkładu.

Rozkład łączny próby, czyli rozkład łączny wektora losowego $(X_1, X_2, ..., X_n)$ jest postaci

$$f(x_1 \,|\, \theta) \times ... \times f(x_n \,|\, \theta),$$

Niech $X_1, X_2, ..., X_n$ będzie próbą losową z rozkładu o gęstości $f(x \mid \theta)$ gdzie θ oznacza parametr rozkładu.

Rozkład łączny próby, czyli rozkład łączny wektora losowego $(X_1, X_2, ..., X_n)$ jest postaci

$$f(x_1 \mid \theta) \times ... \times f(x_n \mid \theta),$$

Błąd średniokwadratowy

Definicja

Błędem średniokwadratowym estymatora W parametru θ nazywamy funkcję θ, $E^2(\theta)$ zdefiniowana jako

$$E^2(\theta) = E_\theta(W-\theta)^2 =$$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} [W(x_1, \dots, x_n) - \theta]^2 f(x_1 \mid \theta) \times \dots \times f(x_n \mid \theta) dx_1 dx_2 \dots dx_n$$

$$E_\theta(W-\theta)^2 = Var_\theta W + (E_\theta W - \theta)^2$$

$$= Var_\theta W + (\text{Obciążenie } W)^2.$$

Błąd średniokwadratowy

Definicja

Błędem średniokwadratowym estymatora W parametru θ nazywamy funkcję θ, $E^2(\theta)$ zdefiniowana jako

$$E^2(\theta) = E_\theta(W - \theta)^2 =$$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} [W(x_1, \dots, x_n) - \theta]^2 f(x_1 \,|\, \theta) \times \dots \times f(x_n \,|\, \theta) dx_1 dx_2 \dots dx_n$$

$$E_\theta(W - \theta)^2 = Var_\theta W + (E_\theta W - \theta)^2$$

$$= Var_\theta W + (\text{Obciążenie } W)^2.$$

Obciążenie estymatora

Definicja.

Obciążeniem estymatora W parametru θ nazywamy funkcję $E(\theta)$ zdefiniowaną jako

$$E(\theta) = E_\theta W - \theta =$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} W(x_1, x_2, \dots, x_n) f(x_1 \mid \theta) \dots f(x_n \mid \theta) dx_1 dx_2 \dots dx_n - \theta$$

Estymatory nieobciążone

Estymator W parametru θ nazywamy **nieobciążonym** gdy dla wszystkich θ

$$E_\theta W = \theta.$$

Dla estymatora nieobciążonego zachodzi równość

$$E_\theta (W - \theta)^2 = Var_\theta W.$$

Estymatory nieobciążone

Estymator W parametru θ nazywamy **nieobciążonym** gdy dla wszystkich θ

$$E_\theta W = \theta.$$

Dla estymatora nieobciążonego zachodzi równość

$$E_\theta (W-\theta)^2 = Var_\theta W.$$

Estymatory nieobciążone

Przykład

Niech $(X_1, \ldots, X_n)$ będzie próbą losową z rozkładu $N(\mu, \sigma^2)$. Statystyki $\bar{X}$ oraz S^2 są nieobciążonymi estymatorami odpowiednio μ oraz σ^2.

$$E\,\bar{X} = E\frac{1}{n}\sum_{i=1}^{n} X_i = \mu,$$

$$E\,S^2 = E\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sigma^2.$$

Estymatory nieobciążone

Przykład

Niech $(X_1, ..., X_n)$ będzie próbą losową z rozkładu $N(\mu, \sigma^2)$. Statystyki $\bar{X}$ oraz S^2 są nieobciążonymi estymatorami odpowiednio μ oraz σ^2.

$$E\,\bar{X} = E\frac{1}{n}\sum_{i=1}^{n} X_i = \mu,$$

$$E\,S^2 = E\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sigma^2.$$

Przykład cd.

Błędy średniokwadratowe tych estymatorów są w tym przypadku dane przez

$$E(\bar{X} - \mu)^2 = Var(\bar{X}) = \frac{\sigma^2}{n},$$

$$E(S^2 - \sigma^2)^2 = Var(S^2) = \frac{2\sigma^4}{n-1}.$$

Przykład cd.

Alternatywnym estymatorem σ^2 jest estymator NW postaci

$$\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n-1}{n}S^2.$$

Stąd

$$E\hat{\sigma^2} = E\left(\frac{n-1}{n}S^2\right) = \frac{n-1}{n}\sigma^2.$$

Estymator $\hat{\sigma^2}$ jest estymatorem obciążonym parametru σ^2. Wariancja tego estymatora jest równa

$$Var(\hat{\sigma^2}) = Var\left(\frac{n-1}{n}S^2\right) = \left(\frac{n-1}{n}\right)^2 Var\, S^2$$

$$= \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} = \frac{2(n-1)\sigma^4}{n^2}$$

Przykład cd.

Błąd średniokwadratowy jest równy

$$E\left[\hat{\sigma^2} - \sigma^2\right]^2 = \frac{2(n-1)}{n^2} < \left(\frac{2}{n-1}\right)\sigma^4 = E\left[S^2 - \sigma^2\right]^2$$

Przykład cd.

Estymator $\hat{\sigma}^2$ ma mniejszy błąd średniokwadratowy niż estymator S^2. Zgadzając się na obciążenie estymatora otrzymujemy estymator o mniejszym błędzie średniokwadratowym.

Estymatory nieobciążone o jednostajnie najmniejszej wariancji

Definicja.

Estymator W^* jest najlepszym nieobciążonym estymatorem parametru θ jeśli

1. $EW^* = \theta$
2. dla dowolnego estymatora W, dla którego $EW = \theta$, spełniona jest nierówność
$$Var(W^*) \leq Var(W).$$

Jeśli taki estymator W^* istnieje to nazywamy go estymatorem nieobciążonym o jednostajnie najmniejszej wariancji.

Estymatory nieobciążone o jednostajnie najmniejszej wariancji

Definicja.

Estymator W^* jest najlepszym nieobciążonym estymatorem parametru θ jeśli

1. $EW^* = \theta$
2. dla dowolnego estymatora W, dla którego $EW = \theta$, spełniona jest nierówność
$$Var(W^*) \leq Var(W).$$

Jeśli taki estymator W^* istnieje to nazywamy go estymatorem nieobciążonym o jednostajnie najmniejszej wariancji.

Przykład.

Niech $(X_1, \dots, X_n)$ będzie próbą losową z rozkładu Poissona $\pi(\lambda)$. Statystyki $\bar{X}$ oraz S^2 są nieobciążonymi estymatorami średniej i wariancji. W przypadku rozkładu Poissona zarówno średnia jak i wariancją są równe λ, a więc

$$E\,\bar{X} = \lambda,$$

$$E\,S^2 = \lambda.$$

Oba estymatory są więc nieobciążonymi estymatorami parametru λ.

Który z nich jest lepszy?

Przykład.

Niech $(X_1, ..., X_n)$ będzie próbą losową z rozkładu Poissona $\pi(\lambda)$. Statystyki $\bar{X}$ oraz S^2 są nieobciążonymi estymatorami średniej i wariancji. W przypadku rozkładu Poissona zarówno średnia jak i wariancją są równe λ, a więc

$$E\,\bar{X} = \lambda,$$

$$E\,S^2 = \lambda.$$

Oba estymatory są więc nieobciążonymi estymatorami parametru λ.

Który z nich jest lepszy?

Przykład cd.

Możemy porównać ich wariancje. Łatwo pokazać, że

$$Var(\bar{X}) = \frac{\lambda}{n}.$$

Obliczenie $Var(S^2)$ wymaga całkiem sporo rachunków w celu pokazania, że

$$Var(\bar{X}) \leq Var(S^2).$$

Możemy rozpatrywać inne nieobciążone estymatory parametru λ. Na przykład estymatory postaci

$$W_a = a\bar{X} + (1-a)S^2,$$

które dla dowolnego a są nieobciążonymi estymatorami parametru λ gdyż

$$E\,W_a = E(a\bar{X} + (1-a)S^2) = aE\bar{X} + (1-a)ES^2 = \lambda.$$

Przykład cd.

Możemy porównać ich wariancje. Łatwo pokazać, że

$$Var(\bar{X}) = \frac{\lambda}{n}.$$

Obliczenie $Var(S^2)$ wymaga całkiem sporo rachunków w celu pokazania, że

$$Var(\bar{X}) \leq Var(S^2).$$

Możemy rozpatrywać inne nieobciążone estymatory parametru λ. Na przykład estymatory postaci

$$W_a = a\bar{X} + (1-a)S^2,$$

które dla dowolnego a są nieobciążonymi estymatorami parametru λ gdyż

$$E\,W_a = E(a\bar{X} + (1-a)S^2) = aE\bar{X} + (1-a)ES^2 = \lambda.$$