4 Przedziały ufności

Niech $\theta \in \Theta$ oznacza parametr modelu statystycznego.

DEFINICJA

Przedział (L, R) określony parą statystyk L i R takich, że $\mathrm{P}_\theta(L \leq R) = 1$ dla każdego $\theta \in \Theta$, nazywamy **przedziałem ufności** dla parametru θ, na **poziomie ufności** $1 - \alpha$ ($0 < \alpha < 1$), gdy dla każdego $\theta \in \Theta$

$$\mathrm{P}_\theta(L < \theta < R) \geq 1 - \alpha.$$

Typowe wartości poziomu ufności to: 0,9; 0,95; 0,99 (zazwyczaj podawane w %).

Konstrukcja przedziałów ufności

DEFINICJA

Funkcję $Q(\boldsymbol{X}, \theta)$ nazywamy **funkcją centralną** dla parametru θ, gdy

1. rozkład prawdopodobieństwa zmiennej losowej Q jest absolutnie ciągły i nie zależy od parametru θ,
2. funkcja $Q(\boldsymbol{X}, \theta)$ jest ciągła i ściśle monotoniczna względem θ.

Uwaga! Warunek pierwszy można osłabić, żądają tylko aby rozkład graniczny zmiennej losowej Q był absolutnie ciągły. Wtedy uzyskany przedział można stosować jedynie dla dużych prób.

Konstrukcja

1. Obieramy funkcję centralną $Q(\boldsymbol{X}, \theta)$.
2. Wyznaczamy stałe a i b tak, aby

$$P(a < Q < b) = 1 - \alpha.$$

3. Rozwiązujemy nierówność

$$a < Q(\boldsymbol{X}, \theta) < b$$

względem θ otrzymując szukany przedział

$$(L(\boldsymbol{X}), R(\boldsymbol{X})).$$

Uwaga! Stałe a i b można dobrać na wiele sposobów. Zazwyczaj dobieramy je tak, aby

$$P(Q \le a) = P(Q \ge b) = \frac{\alpha}{2}.$$

DEFINICJA

Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0,1)$.\

Mówimy, że zmienna losowa

$$X_1^2 + X_2^2 + \cdots + X_n^2$$

ma rozkład **chi-kwadrat** z n stopniami swobody (ozn. $\chi^2(n)$).

FAKT

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-(x/2)}, \quad x > 0.$$

FAKT

Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $Ex(\lambda)$, $\lambda > 0$.

Wtedy

$$2n\lambda\bar{X} \sim \chi^2(2n).$$

Przykład 1.

W modelu jednej próby prostej z rozkładu wykładniczego $(1-\alpha)\cdot 100\%$ przedział ufności dla parametry λ ma postać:

$$\left(\frac{\chi^2\left(\frac{\alpha}{2}, 2n\right)}{2n\bar{X}}; \frac{\chi^2\left(1-\frac{\alpha}{2}, 2n\right)}{2n\bar{X}} \right),$$

gdzie $\chi^2(p,n)$ oznacza kwantyl rzędu p z rozkładu $\chi^2(n)$.

DEFINICJA

Niech $X \sim N(0,1)$ oraz $Y \sim \chi^2(n)$ będą niezależnymi zmiennymi losowymi.

Mówimy, że zmienna losowa

$$\frac{X}{\sqrt{\frac{1}{n}Y}}$$

ma rozkład z n stopniami swobody (ozn. $t(n)$).

FAKT

$$f(x)=\frac{1}{\sqrt{n\pi}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x\in\mathbf{R}.$$

FAKT

Niech $X_1, X_2, \ldots, X_n$, $n>1$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(\mu, \sigma^2)$.

Wtedy

$$\frac{\bar{X}-\mu}{S}\sqrt{n}\sim t(n-1).$$

Przykład 2.

W modelu jednej próby prostej z rozkładu normalnego $(1-\alpha)\cdot 100\%$ przedział ufności dla parametru μ ma postać:

$$\left(\bar{X}-\frac{S}{\sqrt{n}}t(1-\frac{\alpha}{2},n-1),\bar{X}+\frac{S}{\sqrt{n}}t(1-\frac{\alpha}{2},n-1)\right)$$

gdzie $t(p,n)=F_t^{-1}(p)$ oznacza kwantyl rzędu p z rozkładu $t(n)$.

FAKT

Niech $X_1, X_2, \ldots, X_n$, $n>1$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(\mu, \sigma^2)$.

Wtedy

$$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1).$$

Przykład 3.

W modelu jednej próby prostej z rozkładu normalnego $(1-\alpha)\cdot 100\%$ przedział ufności dla parametru σ^2 ma postać:

$$\left(\frac{(n-1)S^2}{\chi^2\left(1-\frac{\alpha}{2},n-1\right)};\frac{(n-1)S^2}{\chi^2\left(\frac{\alpha}{2},n-1\right)}\right),$$

gdzie $\chi^2(p,n)$ oznacza kwantyl rzędu p z rozkładu $\chi^2(n)$.

Funkcje związane z przedziałami ufności:

e… (EnvStats) - pozwalają wyznaczyć przedziały ufności dla parametrów wybranego modelu.

Przykładowo:

eexp (EnvStats) - pozwala wyznaczyć przedział ufności dla parametru w modelu wykładniczym,

enorm (EnvStats) - pozwala wyznaczyć przedziały ufności dla parametrów w modelu normalnym.

Bootstrapowe przedziały ufności

Przyjmujemy

$$Q(\boldsymbol{X}, \theta) = \hat{\theta} - \theta.$$

Wtedy

$$P(a < \hat{\theta} - \theta < b) = 1 - \alpha$$

.

$$P(\hat{\theta} - b < \theta < \hat{\theta} - a) = 1 - \alpha.$$

Ponieważ rozkład $\hat{\theta} - \theta$ jest, przy ustalonych wartościach $x_1, x_2, \ldots, x_n$, bliski rozkładowi $\hat{\theta}^\star - \hat{\theta}$, zatem

$$P(a < \hat{\theta}^\star - \hat{\theta} < b) \approx 1 - \alpha.$$

Stąd

$$a = q(\alpha/2), \quad b = q(1 - \alpha/2),$$

gdzie $q(p)$ jest kwantylem rzędu p z rozkładu $\hat{\theta}^\star - \hat{\theta}$.

Uwaga: Nieznane wartości kwantyli szacujemy za pomocą percentyli uzyskanych z k realizacji próby bootstrapowej $\boldsymbol{X}^\star$.

Funkcje związane z bootstrapowymi przedziałami ufności:

boot(boot) - próba bootstrapowa,

boot.ci(boot) - bootstrapowy przedział ufności.