3 Estymacja punktowa

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą z populacji o rozkładzie P_θ, gdzie $\theta \in \Theta$ jest parametrem.

DEFINICJA

Estymatorem parametru θ nazywamy statystykę $T(\boldsymbol{X})$, o wartościach w zbiorze Θ, której wartość, dla konkretnej realizacji $\boldsymbol{x}$ próby $\boldsymbol{X}$, przyjmujemy za ocenę nieznanej wartości parametru θ (ozn. $\hat{\theta}(\boldsymbol{X})$ lub $$).

Popularne Metody wyznaczania estymatorów (punktowych):

1. metoda momentów,
2. metoda największej wiarogodności,
3. metoda najmniejszych kwadratów.

Metoda momentów

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą z populacji o rozkładzie P_θ, gdzie $\theta \in \Theta \subset \mathbf{R}^d$.

Ponadto, niech rozkłady P_θ posiadają skończone momenty do rzędu d włącznie.

Metoda momentów polega na przyrównaniu kolejnych d momentów z próby $m_i = \frac{1}{n}\sum_{k=1}^{n} X_k^i$, $i = 1, \ldots, d$ do odpowiednich momentów rozkładu populacji $\mathrm{E}(X^i)$, $i = 1, \ldots, d$. Rozwiązują otrzymany w ten sposób układ równań uzyskujemy **estymatory metody momentów (EMM)**.

Uwaga: W metodzie momentów możemy zamiast momentów zwykłych, wykorzystać momenty centralne.

Metoda największej wiarogodności

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą z populacji o rozkładzie P_θ, gdzie $\theta \in \Theta \subset \mathbf{R}^d$.

Ponadto, niech rozkłady P_θ opisane będą za pomocą funkcji prawdopodobieństwa (gęstości) p_θ.

DEFINICJA

Funkcję L określoną wzorem

$$L(\theta; \boldsymbol{x}) = p_{\theta}(\boldsymbol{x})$$

nazywamy **funkcją wiarogodności**.

Uwaga! Funkcją wiarogodności nazywamy czasem funkcję $\ln p_{\theta}(\boldsymbol{x})$.

DEFINICJA

Estymatorem największej wiarogodności (ENW) parametru θ nazywamy statystykę $\hat{\theta}(\boldsymbol{X})$, której wartości $\hat{\theta}(\boldsymbol{x})$ spełniają warunek:

$$\forall \boldsymbol{x} \in \mathbf{X}\colon\ L(\hat{\theta}(\boldsymbol{x}); \boldsymbol{x}) = \sup_{\theta \in \Theta} L(\theta; \boldsymbol{x}).$$

Uwagi!

1. Dla danego parametru θ, ENW może nie istnieć lub może być wyznaczony niejednoznacznie.
2. Zazwyczaj, podczas wyznaczanie ENW, wygodniej jest operować funkcją $\ln L$ niż funkcją L.

Przykład 1. Estymacja parametru λ w modelu wykładniczym

Estymatorem metody momentów (EMM) oraz estymatorem największej wiarogodności (ENW) parametru λ, w modelu jednej próby prostej z rozkładu wykładniczego, jest statystyka

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

Przykład 2. Estymacja parametrów μ i σ^2 w modelu normalnym}

Estymatorami metody momentów (EMM) oraz estymatorami największej wiarogodności (ENW) parametrów μ i σ^2, w modelu jednej próby prostej z rozkładu normalnego, są statystyki

$$\hat{\mu} = \bar{X}$$

oraz

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2.$$

Estymatory nieobciążone

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

DEFINICJA

Statystykę $\hat{\theta}$ nazywamy **estymatorem nieobciążonym** parametru θ, gdy dla każdego $\theta \in \Theta$:

$$\mathrm{E}(\hat{\theta}) = \theta.$$

Uwaga! Klasa estymatorów nieobciążonych danego parametru może być pusta. Zazwyczaj jednak, dla danego parametru istnieje wiele różnych estymatorów nieobciążonych. Najlepszym z nich jest ten, który ma minimalną wariancję - estymator nieobciążony o minimalnej wariancji (ENMW).

TWIERDZENIE

Jeżeli dla parametru θ istnieje estymator nieobciążony o minimalnej wariancji, to jest on wyznaczony jednoznacznie (z dokładnością do zbioru miary zero).

Przykład 3.

W modelu jednej próby prostej z rozkładu wykładniczego, EMM i ENW parametru λ postaci

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

jest obciążonym estymatorem tego parametru.

Estymator nieobciążony (o minimalnej wariancji) parametru λ ma postać:

$$\hat{\lambda} = \frac{n-1}{n\bar{X}}$$

Przykład 4.

W modelu jednej próby prostej z rozkładu normalnego, EMM i ENW parametru μ postaci

$$\hat{\mu} = \bar{X}$$

jest nieobciążonym (o minimalnej wariancji) estymatorem tego parametru.

Ponadto, statystyka

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

jest nieobciążonym (o minimalnej wariancji) estymatorem parametru σ^2.

Rozkłady estymatorów

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

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

Model wykładniczy

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą z populacji o rozkładzie wykładniczym $Ex(\lambda)$, gdzie $\lambda > 0$ jest parametrem. Estymator nieobciążony (o minimalnej wariancji) parametru λ ma postać:

$$\hat{\lambda} = \frac{n-1}{n\bar{X}}$$

Jaki rozkład ma estymator $\hat{\lambda}$?

FAKT

W modelu jednej próby prostej z rozkładu wykładniczego, funkcja

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

Model normalny

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$, $n > 1$ będzie próbą z populacji o rozkładzie normalnym $N(\mu, \sigma^2)$, gdzie μ i σ^2 są parametrami. Estymatory nieobciążone (o minimalnej wariancji) parametrów μ i σ^2 mają postać:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = S^2.$$

Jakie rozkłady mają estymatory $\hat{\mu}$ i $\hat{\sigma}^2$?

TWIERDZENIE (Fishera)

W modelu jednej próby prostej z rozkładu normalnego

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \quad \text{i} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Ponadto estymatory

$$\bar{X} \quad \text{i} \quad S^2$$

są niezależnymi zmiennymi losowymi.

Metoda Monte Carlo

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą z populacji o rozkładzie P_θ, gdzie θ jest parametrem.

Ponadto niech

$$\hat{\theta} = T(\boldsymbol{X})$$

będzie estymatorem parametru θ.

Załóżmy, że dysponujemy k niezależnymi realizacjami próby $\boldsymbol{X}$: $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ oraz że $\hat{\theta}_i = T(\boldsymbol{x}_i)$, $i = 1, 2, \ldots, k$.

FAKT

Histogram wartości $\hat{\theta}_1 = T(\boldsymbol{x}_1), \hat{\theta}_2 = T(\boldsymbol{x}_2), \ldots, \hat{\theta}_k = T(\boldsymbol{x}_k)$ jest dla dużych k, dobrym przybliżeniem rozkładu $\hat{\theta}$.

Metoda bootstrapowa

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą z populacji o rozkładzie P_θ, gdzie θ jest parametrem.

Ponadto niech

$$\hat{\theta} = T(\boldsymbol{X})$$

będzie estymatorem parametru θ oraz F oznacza dystrybuantę rozkładu P_θ.

Dystrybuantą empiryczną nazywamy statystykę:

$$\hat{F}(x) = \frac{\#\{k : X_k \leq x\}}{n}.$$

TWIERDZENIE (Gliwenki-Cantelliego)

Niech $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ będzie próbą prostą z populacji o rozkładzie opisanym dystrybuantą F.

Wtedy

$$\sup_{-\infty < x < \infty} |\hat{F}(x) - F(x)| \xrightarrow{1} 0.$$

Próbą bootstrapową nazywamy próbę losową z rozkładu $\hat{F}$, ozn: $\boldsymbol{X}^{\star} = (X_1^{\star}, X_2^{\star}, \ldots, X_n^{\star})'$.

Uwaga: W celu otrzymania realizacji próby bootstrapowej dokonujemy n-krotnego losowania ze zwracaniem spośród wartości oryginalnej próby.

FAKT (Zasada bootstrap)

Rozkład statystyki $T(\boldsymbol{X}^{\star}) - \hat{\theta}$, przy ustalonych wartościach $x_1, x_2, \ldots, x_n$ jest bliski rozkładowi $T(\boldsymbol{X}) - \theta$.

Załóżmy, że dysponujemy k realizacjami próby bootstrapowej $\boldsymbol{X}^{\star}$: $\boldsymbol{x}_1^{\star}, \boldsymbol{x}_2^{\star}, \ldots, \boldsymbol{x}_k^{\star}$ oraz że $\hat{\theta}_i^{\star} = T(\boldsymbol{x}_i^{\star})$, $i = 1, 2, \ldots, k$.

FAKT

Histogram wartości $\hat{\theta}_1^{\star} - \hat{\theta}, \hat{\theta}_2^{\star} - \hat{\theta}, \ldots, \hat{\theta}_k^{\star} - \hat{\theta}$ jest dla dużych k, dobrym przybliżeniem rozkładu $\hat{\theta} - \theta$.