Statystyka opisowa

Robert Pietrzykowski

email: robert_pietrzykowski@sggw.pl

www.ekonometria.info

Na dziś…

- Sprawy bieżące

Na dziś…

- Wykład 5:
  - Statystyka matematyczna
  - Estymatory punktowe i przedziałowe

Model probabilistyczny:

$$(\mathcal{X}, \mathcal{P}_\theta)$$

np.

$$\{0,1\}\,,\;\; P_\theta(x) = \theta^x(1-\theta)^{1-x}$$

gdzie $\theta = p$ jest znaną liczbą.

Model statystyczny:

$$(\mathcal{X}, \{P_\theta, \theta \in \Theta\})$$

np.

$$\left(\{0,1\}\,, \left\{P_\theta(x) = \theta^x(1-\theta)^{1-x}, \theta \in (0,1)\right\}\right)$$

gdzie $\theta = p$

Estymacja parametrów rozkładu cechy

Estymujemy parametr θ rozkładu cechy X
Próba: $X_1, X_2, \ldots, X_n$

Estymator (punktowy) jest funkcją próby

$$\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$$

przybliżającą wartość parametru θ

METODY WYZNACZANIA ESTYMATORÓW

Metoda największej wiarogodności.

Metoda najmniejszych kwadratów.

Przykład: zadania.pdf

Podstawowe charakterystyki estymatorów:

1. Obciążenie estymatora $\hat{\theta}(\boldsymbol{X})$ parametru θ:

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

2. Błąd średniokwadratowy estymatora $\hat{\theta}(\boldsymbol{X})$ parametru θ:

$$\begin{aligned} \mathrm{BSK}_\theta(\hat{\theta}(\boldsymbol{X})) &= E[(\hat{\theta}(\boldsymbol{X}) - \theta)^2] \\ &= D^2[\hat{\theta}(\boldsymbol{X})] + [E(\hat{\theta}(\boldsymbol{X})) - \theta]^2 \end{aligned}$$

Pożądane własności estymatora:

1. Nieobciążoność

$$E(\hat{\theta}(\boldsymbol{X})) = \theta, \text{ dla każdego } \theta \in \Theta$$

2. Zgodność

$$\lim_{n\to\infty} P\{|\hat{\theta}(\boldsymbol{X}) - \theta| > \varepsilon\} = 0, \text{ dla dowolnego } \varepsilon > 0.$$

3. Efektywność

Spośród zbioru wszystkich estymatorów nieobciążonych $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r$ najefektywniejszym nazywamy estymator o **najmniejszej wariancji**.

Estymator nazywamy estymatorem nieobciążonym o minimalnej wariancji **ENMW**, jeżeli wśród wszystkich estymatorów nieobciążonych parametru θ nie istnieje estymator, którego wariancja byłaby mniejsza dla dowolnej wartości parametru $\theta \in \Theta$.

ROZKŁAD DWUMIANOWY

p — frakcja, wskaźnik struktury

Próba: $X_1, \ldots, X_n$ ($X_i = 0$ lub $= 1$)

$k = \sum_{i=1}^{n} X_i$ — ilość jedynek (sukcesów)

Estymator punktowy:

$$\hat{p} = \frac{k}{n}$$

$$\hat{p} \sim AN\left(p, \frac{1}{n}p(1-p)\right)$$

ROZKŁAD POISSONA

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$n\hat{\lambda} \sim Poisson(n\lambda)$$

$$\hat{\lambda} \sim AN\left(\lambda, \frac{\lambda}{n}\right)$$

ROZKŁAD NORMALNY

Estymator średniej μ — średnia arytmetyczna

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n}$$

Estymator wariancji σ^2 — wariancja próbkowa

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

Estymator odchylenia standardowego σ

$$S = \sqrt{S^2}$$

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

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

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

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Estymator przedziałowy

Przedział ufności (estymator przedziałowy) jest przedziałem o końcach zależnych od próby, który z pewnym z góry zadanym prawdopodobieństwem pokrywa nieznaną wartość parametru θ

$$P\{\theta \in (\underline{\theta}(X_1, \ldots, X_n),\ \overline{\theta}(X_1, \ldots, X_n))\} = 1 - \alpha$$

Poziom ufności: prawdopodobieństwo $1 - \alpha$

Przykład:

przedziały.xls
wykład2_estymacja.xls

1. Rozkład dwupunktowy, przedział asymptotyczny:

$$p \in \left(\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} U_{(1-\alpha/2)}, \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} U_{(1-\alpha/2)} \right)$$

gdzie: $U_{(1-\alpha/2)}$ jest kwantylem standardowego rozkładu normalnego $N(0,1)$ rzędu $1-\alpha/2$.

2. Rozkład Poissona, przedział asymptotyczny:

$$\lambda \in \left(\hat{\lambda} - \sqrt{\frac{\hat{\lambda}}{n}} U_{(1-\alpha/2)}, \hat{\lambda} + \sqrt{\frac{\hat{\lambda}}{n}} U_{(1-\alpha/2)}\right)$$

3. Rozkład normalny:

$$\mu \in \left(\bar{X} - \frac{S}{\sqrt{n}} t_{(\alpha,n-1)}, \bar{X} + \frac{S}{\sqrt{n}} t_{(\alpha,n-1)} \right)$$

gdzie $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}$ oraz $t_{(\alpha,n-1)}$ to wartość krytyczna z rozkładu t-Studenta o $n-1$ stopniach swobody.

$$\sigma^2 \in \left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{(1-\alpha/2,n-1)}}, \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\chi^2_{(\alpha/2,n-1)}} \right)$$

gdzie: $\chi^2_{(\alpha,n-1)}$ to kwantyl z rozkładu Chi-kwadrat o $n-1$ stopniach swobody rzędu α.

Co wpływa na długość przedziału ufności?

[10.00 - 1.25 = 8.75; 10.00 + 1.25 = 11.2

d =

- Liczebność próby
- Poziom ufności
- Wariancja

PYTANIA

1. Co wpływa na długość przedziału ufności?
2. Podaj wzór na przedział ufności dla średniej w rozkładzie normalnym.
3. Podaj wzór na przedział ufności dla wariancji w rozkładzie Poissona.
4. Wyjaśnij pojęcie poziom ufności.
5. Czy można powiedzieć o informacji podawanej dla pewnego modelu samochodu, że średnie zużycie paliwa wynosi 6 litrów na 100 km. Mając informację o zużyciu paliwa dla tego modelu samochodu w postaci przedziału ufności (5,82; 7,48) na poziomie ufności 95% .
6. Wymień znane Ci sposoby uzyskiwania estymatorów.
7. Co jest estymatorem μ w rozkładzie normalnym. Jaki rozkład ma ten estymator.
8. W jakim rozkładzie prawdopodobieństwa estymator średniej jest taki sam jak estymator wariancji.
9. Z jakich elementów skład się model statystyczny.
10. Jaki rozkład ma asymptotyczny estymator p w rozkładzie dwumianowym.