Estymacja

Mgr inż. Szymon Łukasik

szymonl@pk.edu.pl

Wprowadzenie

Zapoznaliście się już Państwo z podstawowymi rozkładami statystycznymi i parametrami, które w nich występują. Dla przypomnienia:

E[X] – wartość oczekiwana

V(X) – wariancja

f(x) – gęstość rozkładu prawdopodobieństwa

F(x) – dystrybuanta

q=F(r) – kwantyl rzędu r

dodatkowo, każdy z typowych rozkładów ma swoje parametry – różnicujące zmienne które ten rozkład przyjmują, np. lambda w rozkładzie ekspotencjalnym:

$$f(x;\lambda) = \lambda e^{-\lambda x}; \quad x \geq 0; \lambda > 0$$

Problemem który będzie rozważany w trakcie niniejszych zajęć jest odpowiedź na pytanie: jak ustalić wartość tych parametrów/charakterystyk funkcyjnych na podstawie próby zmiennej losowej? Innymi słowy – jak je **estymować**?

Estymator i jego cechy

Estymator parametru *b* oznacza się $\hat{b}$. A jego wartość oczekiwaną wyznaczoną na podstawie próby $E(\hat{b})$. **Obciążeniem estymatora** nazywamy różnicę pomiędzy jego wartością oczekiwaną, a „prawdziwą" wartością estymowanego parametru tj. $E(\hat{b}) - b$.

Estymatorem nieobciążonym nazywamy oczywiście estymator o zerowym obciążeniu.

Obciążenie estymatora informuje nas o tym, wokół jakiej wartości będą oscylować uzyskiwane w praktyce wartości estymatora. Wariancja tego estymatora $V(\hat{b})$ z kolei mówi nam o rozrzucie, jakie wartości estymatora będą uzyskiwały – wokół wartości oczekiwanej.

Estymatorem asymptotycznie nieobciążonym nazywamy taki estymator dla którego wraz z wzrostem liczności próby obciążenie maleje do zera:

$$\lim_{m\to\infty}(E(\hat{b})-b)=0 \tag{1}$$

Estymator zgodny – to estymator, który zmierza do wartości „prawdziwej” parametru według prawdopodobieństwa, czyli dla każdego ε:

$$\lim_{m\to\infty} P(|E(\hat{b})-b|>\varepsilon)=0 \tag{2}$$

Podstawowe estymatory

- Estymator wartości oczekiwanej:

$$\hat{E}=\frac{1}{m}\sum_{i=1}^{m}x_i$$

(czyli średnia arytmetyczna)

Estymator ten jest nieobciążony i zgodny.

- Estymator wariancji

$$\hat{V}=\frac{1}{m}\sum_{i=1}^{m}(x_i-E)^2$$

lub gdy wartość oczekiwana nie jest znana (w większości przypadków tak jest) i *m ≥ 2:*

$$\hat{V}=\frac{1}{m-1}\sum_{i=1}^{m}(x_i-\hat{E})^2$$

Estymator ten jest również nieobciążony i zgodny.

Gdy chcemy wyznaczyć wariancję i średnią w jednej pętli warto zastosować przydatny wzór:

$$\hat{V} = \frac{1}{m-1}\sum_{i=1}^{m} {x_i}^2 - \frac{1}{m(m-1)}\left(\sum_{i=1}^{m} x_i\right)^2$$

Estymator odchylenia standardowego to bezpośrednio $\hat{\sigma} = \sqrt{\hat{V}}$

- Estymator kwantyla rzędu r

Niech rozważana próba $x_1, x_2, \ldots, x_m$ zostanie uporządkowana rosnąco $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_m$. Najprostszym klasycznym estymatorem kwantyla r-tego rzędu jest element próby o indeksie $[mr + 0{,}5]$. [] – oznacza część całkowitą z liczby.

Więc:

$$\hat{q} = \tilde{x}_{[mr+0,5]}.$$

Jest to estymator jednoskładnikowy kwantyla. W niektórych przypadkach lepiej jest stosować estymator postaci:

$$\hat{q} = (0{,}5 - mr + [mr + 0{,}5])\tilde{x}_{[mr+0,5]} + (0{,}5 + mr - [mr + 0{,}5])\tilde{x}_{[mr+1,5]}$$

Oczywiście, gdy $mr < 0{,}5$ to $\hat{q} = \tilde{x}_1$, a gdy $mr > m - 0{,}5$ to $\hat{q} = \tilde{x}_m$.

Metoda momentów dla estymacji parametrów rozkładu

Metoda ta pozwala na wyznaczanie wartości parametrów występujących w **założonej postaci** funkcji gęstości prawdopodobieństwa lub innej charakterystyce funkcyjnej (takiej jak np. dystrybuanta). Polega ona na przyjęciu owych parametrów na podstawie estymatorów pierwszych momentów rozkładu (wartości oczekiwanej i wariancji). Przykładowo parametry rozkładu jednostajnego wyznacza się otrzymując $\hat{a}$ i $\hat{b}$ z układu równań:

$$\begin{cases} \mu = \dfrac{\hat{a} + \hat{b}}{2} \\ \hat{V} = \dfrac{(\hat{b} - \hat{a})^2}{12} \end{cases}$$