Wykład 13

1 Estymacja przedziałowa.

DEFININICJA. *Estymatorem przedziałowym parametru rzeczywistego* θ *nazywamy parę funkcji* $L(x_1, x_2, ..., x_n)$ *i* $U(x_1, x_2, ..., x_n)$ *próby* $x_1, x_2, ..., x_n$ *spełniającą dla wszystkich* $x = (x_1, x_2, ..., x_n)$ *nierówność*

$$L(x) \leqslant U(x).$$

Dla x, realizacji próby losowej $(X_1, ..., X_n)$, oszacowanie przedziałowe parametru θ mówi, że

$$L(x) \leqslant \theta \leqslant U(x).$$

Przedział losowy $[L(X),\, U(X)]$ nazywamy estymatorem przedziałowym parametru θ.

1.1 Przedział ufności dla wariancji

Estymatorem punktowym dla wariancji jest wariancja probkowa

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

zgodny, nieobciążony estymator parametru σ^2 w rozkładzie normalnym.

Gdy $X_1,\, X_2, ..., X_n$ tworzą próbę prostą ze standardowego rozkładu normalnego $N(0,\, 1)$ to

$$\sum_{i=1}^{n} X_i^2,$$

ma rozkład χ^2 z n stopniami swobody.

Stąd dla próby losowej $(X_1,\, X_2, ..., X_n)$, z rozkładu normalnego $N(\mu, \sigma^2)$, unormowana zmienna losowa

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2,$$

ma rozkład χ^2 z n stopniami swobody.

W przypadku gdy nie znamy parametrów μ i σ^2 rozpatrujemy statystykę

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 = \frac{(n-1)S_n^2}{\sigma^2}.$$

Zmienna ta ma rozkład χ^2 z $n-1$ stopniami swobody.

Gdy ustalimy liczbę α, to oznaczając przez $\chi_{\frac{\alpha}{2};n-1}$ i $\chi_{1-\frac{\alpha}{2};n-1}$ odpowiednio kwantyle rzędu $\frac{\alpha}{2}$ i $1-\frac{\alpha}{2}$, czyli gdy

$$P\left(\chi^2 \leqslant \chi_{\frac{\alpha}{2};n-1}\right) = \frac{\alpha}{2},$$

$$P\left(\chi^2 \leqslant \chi_{1-\frac{\alpha}{2};n-1}\right) = 1 - \frac{\alpha}{2},$$

otrzymujemy

$$P\left(\chi_{\frac{\alpha}{2};n-1} \leqslant \frac{(n-1)S_n^2}{\sigma^2} \leqslant \chi_{1-\frac{\alpha}{2};n-1} \right) = 1 - \alpha.$$

Przekształcając otrzymujemy

$$P\left(\frac{(n-1)S_n^2}{\chi_{1-\frac{\alpha}{2};n-1}} \leqslant \sigma^2 \leqslant \frac{(n-1)S_n^2}{\chi_{\frac{\alpha}{2};n-1}} \right) = 1 - \alpha.$$

Z równości tej wynika, że

$$\left[\frac{(n-1)s_n^2}{\chi_{1-\frac{\alpha}{2};n-1}}, \frac{(n-1)s_n^2}{\chi_{\frac{\alpha}{2};n-1}} \right],$$

stanowi przedział ufności dla wariancji na poziomie ufności $1-\alpha$.

Dla odchylenia standardowego σ, przedział ufnosci ma postać

$$\left[\sqrt{\frac{(n-1)s_n^2}{\chi_{1-\frac{\alpha}{2};n-1}}}, \sqrt{\frac{(n-1)s_n^2}{\chi_{\frac{\alpha}{2};n-1}}} \right].$$

1.2 Przedział ufności dla p

Dla zmiennej losowej X o rozkładzie Bernoulliego $b(x; n, p)$ zmienna losowa

$$\frac{X - np}{\sqrt{np(1-p)}}, \tag{1}$$

ma asymptotyczny rozkład normalny $N(0, 1)$.

Dzieląc we wzorze (1) licznik i mianownik przez n otrzymujemy

$$\frac{\frac{X}{n} - p}{\sqrt{\frac{np(1-p)}{n^2}}} = \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}. \tag{2}$$

Dla dużych n

$$P\left(-z_{\frac{\alpha}{2}} \leqslant \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leqslant z_{\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

Zdarzenie

$$\left\{-z_{\frac{\alpha}{2}} \leqslant \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leqslant z_{\frac{\alpha}{2}}\right\}$$

jest takie samo jak zdarzenie

$$\left\{\left[\frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}}\right]^2 \leqslant z^2_{\frac{\alpha}{2}}\right\},$$

lub jak zdarzenie

$$\left\{\left(\frac{X}{n} - p\right)^2 - z^2_{\frac{\alpha}{2}}\frac{p(1-p)}{n} \leqslant 0\right\}. \tag{3}$$

W celu konstrukcji przedziału ufnosci dla p mozemy nierówność (3) zapisać w postaci

$$L(X) \leqslant p \leqslant U(X).$$

Zamiast to zrobić dla przypadku ogólnego zróbmy to dla konkretnych danych.

PRZYKŁAD. *Przepytano 125 głosujacych i okazało się, że 78 z nich poparło kandydata X. Jaki jest przedział ufności na poziomie 95% dla proporcji głosujacych popierajacych kandydata X w całej populacji?*

Realizacja zmiennej losowej X o rozkładzie Bernouliego $b(x; p, 125)$ jest $x = 78$. Ponieważ

$$z_{\frac{\alpha}{2}} = z_{0,025} = 1,96$$

wieć nierówność 3 przyjmuje postać

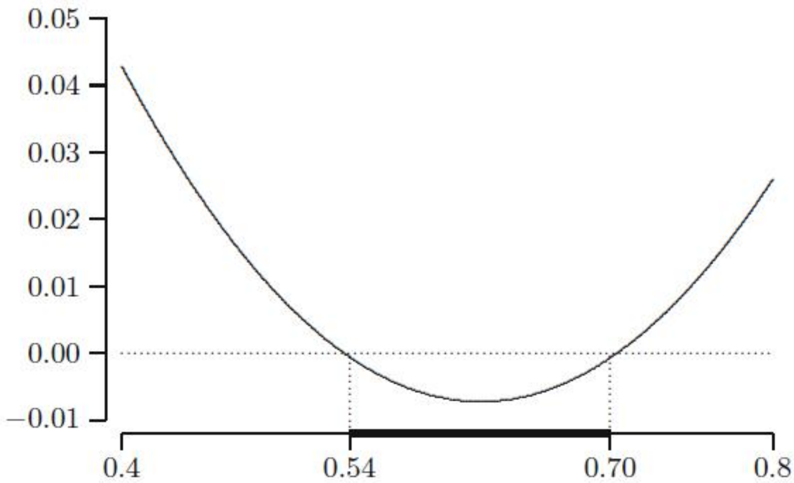
$$\left(\frac{78}{125} - p\right)^2 - \frac{1,96}{125} p(1-p) \leqslant 0,$$

co po przekształcenich daje nam

$$1,0307p^2 - 1,2787p + 0,3894 \leqslant 0. \tag{4}$$

Pierwiatkami równaia kwadratowego (4) są

$$\begin{aligned} p_{1,2} &= \frac{-(-1,2787) \pm \sqrt{(-1,2787)^2 - 4 \times 1,0307 \times 0,3894}}{2 \times 1,0307} \\ &= 0,6203 \pm 0,0835 \end{aligned}$$



Rysunek 1: Parabola $1,0307p^2 - 1,2787p + 0,3894$ i wyznaczony przy jej pomocy przedział ufności.

Stąd $l = 0,54$ i $u = 0,70$ i otrzymujemy przedział ufności dla p na poziomie 95%

$$[0,54,\ 0,70].$$

1.3 Przedział ufności dla proporcji

Przy konstrukcji przedziału ufności dla prawdopodobieństwa sukcesu w rozkładzie 0-1

$$P\left(X_i = 1\right) = 1 - P\left(X_i = 0\right) = p, \quad i = 1, 2, ..., n,$$

możemy skorzystać z nastepujacego z faktu, że zmienna losowa

$$\frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}},$$

ma rozkład asymptotyczny $N(0,1)$.

TWIERDZENIE. *Dla próby prostej $X_1, X_2, ..., X_n$ z rozkładu o średniej μ i skończonej wariancji $0 < \sigma^2 < \infty$ zachodzi zbieżność*

$$\begin{aligned}\lim_{n\to\infty} P\left(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \in [a,b]\right) &= P(N(o,1) \in [a,b]) \\ &= \Phi(b) - \Phi(a),\end{aligned}$$

gdzie

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\, dx.$$

Inaczej mówiąc zmienna losowa $\bar{X}$ ma rozkład asymptotyczny $N\left(\mu,\, \frac{\sigma}{\sqrt{n}}\right)$.

W przypadku zmiennych losowych o rozkładzie 0-1

$$E\bar{X} = p, \quad Var(\bar{X}) = \frac{p(1-p)}{n},$$

i twierdzenie ma postać

TWIERDZENIE. *Dla próby prostej $X_1, X_2, ..., X_n$ z rozkładu 0-1 zachodzi zbieżność*

$$\lim_{n\to\infty} P\left(\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}} \in [a,b]\right) = \Phi(b) - \Phi(a),$$

gdzie

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}}\, dx.$$

UWAGA. *Z oszacowanie mozemy skorzystać gdy $n \geqslant 30$ oraz $np \geqslant 5$ i $n(1-p) \geqslant 5$ lub gdy p jest nieznane $n\bar{X} \geqslant 5$ i $n(1-\bar{X}) \geqslant 5$.*

Teraz możemy skonstruować przedział ufności dla parametru p.

$$P\left(-z_{\frac{\alpha}{2}} \leqslant \frac{\bar{X}-p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \leqslant z_{\frac{\alpha}{2}}\right) \approx 1-\alpha.$$

Stąd

$$P\left(\bar{X} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leqslant p \leqslant \bar{X} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) \approx 1-\alpha.$$

Przedział ufności dla proporcji p na poziomie ufności $(1-\alpha)$ ma postać

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}},\ \bar{X} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right].$$