ZAGADNIENIE V.3

Estymacja punktowa wartości oczekiwanej i wariancji. Przedziały ufności dla wartości oczekiwanej.

Estymacja punktowa – grupa metod statystycznych, służąca do punktowego oszacowania wartości szukanego parametru rozkładu. *Punktowe oszacowanie* oznacza tutaj, że uzyskujemy konkretną wartość liczbową, nie zaś przedział liczbowy, jak dzieje się to w przypadku estymacji przedziałowej.

Metody estymacji punktowej sprowadzają się do wyznaczenia odpowiednią metodą estymatora szacowanego parametru.

Estymacja wartości oczekiwanej
Zdefiniujmy estymator zwany *średnia z próby*

$$\bar{x} : \mathbb{R}^n \to \mathbb{R}, \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Fakt Jeżeli X jest zmienna losowa posiadającą wartość oczekiwana, to x (z daszkiem) jest nieobciążonym i mocno zgodnym estymatorem wartości oczekiwanej.

Estymacja wariancji
Zdefiniujemy estymator zwany *wariancja z próby*.

1. Niech *X* będzie zmienną losową o znanej wartości oczekiwanej, równej *a*. Wówczas wariancję z próby definiujemy

$$^{*}s^2 : \mathbb{R}^n \to \mathbb{R}, \qquad ^{*}s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2.$$

   Fakt Jeżeli X jest zmienną losową posiadającą wariancję, to $*s^2$ jest nieobciążonym i mocno zgodnym estymatorem wariancji.

2. Niech *X* będzie zmienną losową o nieznanej wartości oczekiwanej. Wówczas wariancję z próby definiujemy

$$s^2 : \mathbb{R}^n \to \mathbb{R}, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad n \geqslant 2,$$

$$\hat{s}^2 : \mathbb{R}^n \to \mathbb{R}, \qquad \hat{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

   Fakt Jeżeli X jest zmienną losową posiadającą wariancję, to
   - s^2 jest nieobciążonym i mocno zgodnym estymatorem wariancji,
   - s^2 (z daszkiem) jest asymptotycznie nieobciążonym i mocno zgodnym estymatorem wariancji.

Przedziały ufności dla wartości oczekiwanej
(z notatek szczelesa, mictyda, welo i newtona)
Zadajemy poziom ufności (=prawdopodobieństwo tego, ze tworzony przedział zawiera nieznany parametr).

$$P(\theta \in (T_1(x), T_2(x))) = \alpha$$

$$P(\theta \in (\theta_-(x), \theta_+(x))) = \alpha$$

Oba przedziały powyżej to to samo, $\theta_-(x)$ i $\theta_+(x)$ to inne oznaczenie krańców tego przedziału (a wzięło się ono stad, ze szacują one θ).
Standardowo *alfa = 0.95*
Szacujemy θ - nieznana wartość oczekiwana.

1. Rozkład x_i jest normalny, σ^2 znana
2. Rozkład x_i jest normalny, σ^2 nieznana
3. Rozkład x_i nie jest normalny

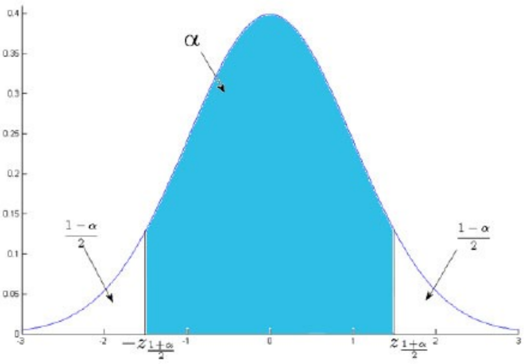
Weźmy pewną zmienną losową, której rozkład jest określony, i która zależny od θ.

Przypadek 1, $\{x_i\}$ ma rozkład N(0, 1), σ^2 znana

Definicja. *Kwantylem rzędu alfa (oznaczenie: z_{alfa})rozkładu normalnego standardowego nazywamy taka liczbę, dla której prawdopodobieństwo trafienia zmiennej losowej o tym rozkładzie do przedziału $(-\infty; z_{alfa})$ wynosi alfa. Innymi słowy, pole pod wykresem na lewo od tego kwantyla wynosi alfa.*

Uwaga. *Na wykresie obok zaznaczone są kwantyle rzędu (1−alfa)/2 oraz alfa+((1-alfa)/2). Liczby te dobrane są tak, aby zaciemnione pole wynosiło alfa. Korzystamy również z tego, ze oba kwantyle są położone symetrycznie względem osi OY.*

Tworzymy przedział tak, aby prawdopodobieństwo, że trafi do niego zmienna losowa o N(0,1) było równe *alfa*. Innymi słowy: tworzymy przedział o losowych końcach (końce przedziału - statystyki), który z prawdopodobieństwem pokrywa parametr θ.



$$P\left(\frac{\overline{x}-\theta}{\frac{\sigma}{\sqrt{n}}} \in \left(-z_{\frac{1+\alpha}{2}}, z_{\frac{1+\alpha}{2}}\right)\right) = \alpha$$

$$\overline{x} - z_{\frac{1+\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \theta < \overline{x} + z_{\frac{1+\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$P\left(\theta \in \left(\overline{x} - z_{\frac{1+\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\frac{1+\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)\right) = \alpha$$

Długość przedziału jest równa 2*$z_{(1+alfa)/2}$*(σ/sqrt(n)). Czynnik σ/sqrt(n) pojawia się, bo w ogólności nie mamy rozkładu normalnego standardowego (N(0,1)) - standaryzowaliśmy nasz rozkład przy pomocy tego właśnie czynnika.
Co możemy zrobić, żeby przedział był jak najkrótszy?
Jeśli wykres gęstości jest symetryczny, to najkrótszy przedział również jest symetryczny.
Zachodzą tez zależności:

większe *alfa* => większa długość przedziału
większe n => mniejsza długość przedziału

Przypadek 2, $\{x_i\}$ ma rozkład N(0, 1), σ^2 nie jest znana

W tym przypadku postępujemy analogicznie, z ta różnicą, że ma $\frac{\overline{x}-\theta}{\frac{s_p}{\sqrt{n}}}$ rozkład Studenta o (n-1) stopniach swobody, gdzie

$$s_p^2 = \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \overline{x})^2$$

Dalej jak poprzednio:

$$P\left(\frac{\overline{x}-\theta}{\frac{s_p}{\sqrt{n}}} \in \left(-t_{n-1,\frac{1+\alpha}{2}}, t_{n-1,\frac{1+\alpha}{2}}\right)\right) = \alpha$$

$$(\theta_-(x_1, x_2, ..., x_n), \theta_+(x_1, x_2, ..., x_n)) = \left(\overline{x} - t_{n-1,\frac{1+\alpha}{2}}\frac{s_p}{\sqrt{n}}, \overline{x} + t_{n-1,\frac{1+\alpha}{2}}\frac{s_p}{\sqrt{n}}\right)$$

Długość przedziału jest równa $2t_{n-1,\frac{1+\alpha}{2}}\frac{s_p}{\sqrt{n}}$.

Uwaga $t_{n,alfa}$ *to oznaczenie analogiczne do użytego wcześniej z_{alfa} - pole na lewo od tej liczby wynosi alfa, zaś n-1 oznacza ilość stopni swobody w rozkładzie.*

Przypadek 3, $\{x_i\}$ ma rozkład dowolny, ale n jest duże (n >= 100)

Twierdzenie (Centralne Twierdzenie Graniczne). *x_1, x_2, ..., x_n - niezależne zmienne losowe o jednakowym rozkładzie*
θ - wartość oczekiwana
σ^2 - wariancja
Przy takich założeniach prawda jest:

$$\frac{\overline{x}-\theta}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{} N(0,1)$$

Uwaga. *Stwierdzenie, iż też jest prawdziwe* $\frac{\overline{x}-\theta}{\frac{s_p}{\sqrt{n}}} \underset{n\to\infty}{\longrightarrow} N(0,1)$

Będziemy korzystać z uwagi, czyli przyjmiemy, że $\frac{\overline{x}-\theta}{\frac{s_p}{\sqrt{n}}} \approx N(0,1).$

$$P\left(\frac{\overline{x}-\theta}{\frac{s_p}{\sqrt{n}}} \in (-z_{\frac{1+\alpha}{2}}, z_{\frac{1+\alpha}{2}})\right) \approx \alpha$$

$$(\theta_-(x_1, x_2, ..., x_n), \theta_+(x_1, x_2, ..., x_n)) = \left(\overline{x} - z_{\frac{1+\alpha}{2}} \frac{s_p}{\sqrt{n}}, \overline{x} + z_{\frac{1+\alpha}{2}} \frac{s_p}{\sqrt{n}}\right)$$