Poprzednim razem: Prawo przenoszenia błędów – uwaga na zapis macierzowy

$$cov(Y_k, Y_q) = E\left[(Y_k - E(Y_k))(Y_q - E(Y_q))\right] =$$

$$= E\left[\sum_l \left(\frac{\partial Y_k}{\partial X_l}\right)\Bigg|_{\vec{X}=E(\vec{X})} (X_l - E(X_l)) \sum_m \left(\frac{\partial Y_q}{\partial X_m}\right)\Bigg|_{\vec{X}=E(\vec{X})} (X_m - E(X_m))\right] =$$

$$= \sum_{l,m} \left(\frac{\partial Y_k}{\partial X_l}\right)\Bigg|_{\vec{X}=E(\vec{X})} \left(\frac{\partial Y_q}{\partial X_m}\right)\Bigg|_{\vec{X}=E(\vec{X})} E\left[(X_l - E(X_l))(X_m - E(X_m))\right] =$$

$$= \boxed{\sum_{l,m} \left(\frac{\partial Y_k}{\partial X_l}\right)\Bigg|_{\vec{X}=E(\vec{X})} \left(\frac{\partial Y_q}{\partial X_m}\right)\Bigg|_{\vec{X}=E(\vec{X})} cov(X_l, X_m)}$$

$$[C(X)]_{l,m} \equiv cov(X_l, X_m); \quad [C(Y)]_{k,q} \equiv cov(Y_k, Y_q); \quad T_{i,j} \equiv \left(\frac{\partial Y_i}{\partial X_j}\right)\Bigg|_{\vec{X}=E(\vec{X})}$$

$$[C(Y)]_{k,q} = \sum_{l,m} T_{k,l} T_{q,m} [C(X)]_{l,m} = \sum_{l,m} T_{k,l} [C(X)]_{l,m} T^T_{m,q}$$

$\boxed{C(Y) = TC(X)T^T}$ ← macierzowo

Statystyka – podstawowe pojęcia

- **Populacja generalna** – zbiór wszystkich możliwych wyników (skończony lub nieskończony).
- **Próba** – skończony zbiór doświadczeń.
- **Estymacja** – zajmuje się wnioskowaniem o własnościach populacji na podstawie próby.
  Najczęściej interesują nas wartość oczekiwana i odchylenie standardowe (lub wariancja).
- Próba prosta to ciąg niezależnych doświadczeń odnoszących się do tej samej populacji generalnej. Wynikiem n-elementowej próby jest n-wymiarowa zmienna losowa.

- **Statystyka** – funkcja zmiennych losowych obserwowanych w próbie, sama też jest zmienną losową.

Estymatory - podstawowe definicje

- Syt. $X_1,\ldots,X_n$ – wyniki pomiarów w próbie, mają one pewien rozkład prawdopodobieństwa (funkcję gęstości prawdopodobieństwa). Zakładamy, że ten rozkład jest taki sam dla każdego X_i, czyli wylosowanie pojedynczego X_i nie zmienia rozkładu (np. przedziałów rozkładu jednorodnego) i próba jest odpowiednio liczna. Rozkład ten zależy od jakiegoś parametru Θ (lub kilku parametrów).

- **Estymatorem** parametru Θ nazywamy statystykę o rozkładzie prawdopodobieństwa zależnym od Θ i oznaczamy $T_n(\Theta)$ lub $T_n(x_1,x_2,\ldots,x_n;\Theta)$.
  Zatem dla danego parametru może istnieć wiele estymatorów o różnych własnościach. Pożądanymi własnościami są:

- **1. Zgodność** – estymator $T_n(\Theta)$ jest zgodny gdy
  $$\forall \varepsilon > 0: \quad \lim_{n\to\infty} P(|T_n(\Theta) - \Theta| < \varepsilon) = 1$$
  wartość estymatora (statystyki)

Estymatory - podstawowe definicje

- **2. Obciążenie (jego brak)** – estymator $T_n(\Theta)$ jest zmienną losową (różne próby dają różne wartości estymatora), zatem ma wartość oczekiwaną i wariancję. Obciążenie estymatora to
  $$B_n \equiv E(T_n(\Theta)) - \Theta$$
- **Estymator jest nieobciążony**, gdy niezależnie od wielkości próby
  $$\boxed{E(T_n(\Theta)) = \Theta \quad \text{czyli} \quad B_n = 0}$$
- Estymator jest obciążony gdy $B_n \neq 0$.
- **Estymator jest asymptotycznie nieobciążony** gdy
  $$\lim_{n\to\infty} B_n = 0$$

- **3. Efektywność** – estymator jest **najbardziej efektywny** gdy ma najmniejszą wariancję.

- Optymalnie estymator powinien być: zgodny, nieobciążony i najbardziej efektywny.

Estymatory – własności ogólne

- **Estymacja punktowa** – oszacowanie wartości parametru Θ przez podanie wartości estymatora $T_n(\Theta)$ tego parametru.
- **Estymacja przedziałowa** – podanie przedziału liczbowego, wewnątrz którego, z założonym prawdopodobieństwem, leży prawdziwa wartość parametru Θ.

Estymacja punktowa:

- Estymatorem wartości oczekiwanej jest średnia z próby:

$$\boxed{T_n(E(X)) \equiv \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

Estymacja punktowa

- Nieobciążony
  Dow: niech wszystkie X_i pochodzą z populacji o $E(X)=X_0$

$$E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} E(X) = \frac{1}{n}\sum_{i=1}^{n} X_0 = \frac{1}{n} n X_0 = X_0$$

- Zgodny

$$P\left(\left|\overline{X} - E(X)\right| > \varepsilon\right) = P\left(\left|\overline{X} - E(\overline{X})\right| > \varepsilon\right) \le var(\overline{X})/\varepsilon^2 = var(X)/(n\varepsilon^2)$$

$$\lim_{n\to\infty} P\left(\left|\overline{X} - E(X)\right| > \varepsilon\right) = \lim_{n\to\infty}\left(var(X)/(n\varepsilon^2)\right) = 0$$

- Najbardziej efektywny

$$\forall a > 0: \quad P(|X - \mu| \ge a) \le \frac{\sigma^2}{a^2}$$

Nierówność Czebyszewa-Bienayme

Estymacja punktowa

$$var(\overline{X}) = E\left[\left(\overline{X} - E(\overline{X})\right)^2\right] = E\left(\overline{X}^2\right) - \left(E\left(\overline{X}\right)\right)^2 = \frac{1}{n^2}E\left(\left(\sum_{i=1}^{n} x_i\right)^2\right) - X_0^2 =$$

$$= \frac{1}{n^2}E\left(\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} x_j\right)\right) - X_0^2 = \frac{1}{n^2}E\left(\sum_{i=1,\dots,n\ j\neq i} x_i x_j + \sum_{i=1}^{n} x_i^2\right) - X_0^2 =$$

$$= \frac{1}{n^2}E\left(\sum_{i=1,\dots,n\ j\neq i} x_i x_j\right) + \frac{1}{n^2}E\left(\sum_{i=1}^{n} x_i^2\right) - X_0^2 = \frac{1}{n^2}\sum_{i=1,\dots,n\ j\neq i} E(x_i)E(x_j) + \frac{n}{n^2}E\left(X^2\right) - X_0^2 =$$

$$= \frac{n(n-1)}{n^2}(E(X))^2 + \frac{1}{n}E\left(X^2\right) - X_0^2 = \frac{1}{n}\left[(n-1)(E(X))^2 + E\left(X^2\right) - n(E(X))^2\right] =$$

$$= \frac{1}{n}\left[E\left(X^2\right) - (E(X))^2\right] = \frac{1}{n}var(X)$$

- Czyli $\boxed{var\left(\overline{X}\right) = \frac{1}{n}var(X) \quad \Rightarrow \quad \sigma\left(\overline{X}\right) = \frac{1}{\sqrt{n}}\sigma(X)}$
- Ostatecznie przyjmujemy $X_0 = \overline{X} \pm \sigma\left(\overline{X}\right)$

RPiS 2023/2024 7

Estymatory odchylenia standardowego

- Estymator zgodny, asymptotycznie nieobciążony

$$\boxed{T_n(\sigma(X)) \equiv S(X) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}}$$

Uwaga $S^2(X)$ jest estymatorem nieobciążonym wariancji

- Estymator zgodny, asymptotycznie nieobciążony

$$T_n(\sigma(X)) \equiv \widehat{S}(X) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

- Estymator zgodny, nieobciążony, najbardziej efektywny

$$T_n(\sigma(X)) \equiv \tilde{S}(X) = \sqrt{\sqrt{\frac{n-1}{2}}\,\frac{\Gamma(0.5(n-1))}{\Gamma(0.5n)}\,S^2(X)}$$

RPiS 2023/2024 8

Estymatory współczynnika korelacji

$$T_n(\rho_{X,Y}) = \frac{\sum_{i=1}^{n}(x_i - \overline{X})\cdot(y_i - \overline{Y})}{\sqrt{\left(\sum_{i=1}^{n}(x_i - \overline{X})^2\right)\left(\sum_{i=1}^{n}(y_i - \overline{Y})^2\right)}}$$

Ten estymator jest zgodny i obciążony.

RPiS 2023/2024 9

Estymacja przedziałowa

- **Estymacja przedziałowa** – podanie przedziału liczbowego, $\left[T_n^L(\theta), T_n^P(\theta)\right]$ wewnątrz którego, z założonym prawdopodobieństwem γ =1-α, leży prawdziwa wartość parametru Θ.
- Przedział ten nazywamy **przedziałem ufności** dla parametru Θ na poziomie ufności 1-α = γ.
- Własności przedziału ufności:

  1). $P\left(T_n^L(\theta) \le \theta \le T_n^P(\theta)\right) = 1 - \alpha \equiv \gamma$

  2). Końce przedziału zależą od próby i od γ, a nie zależą funkcyjnie od Θ.

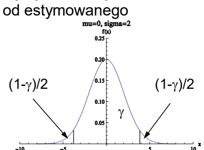
  3). Zwykle używa się 1-α ≈ 0.9 (zwiększając 1-α tracimy na dokładności, bo zwiększa się długość przedziału ufności).

RPiS 2023/2024 10

Estymacja przedziałowa wartości oczekiwanej

- Aby znaleźć przedział ufności szukamy statystyki o znanym rozkładzie prawdopodobieństwa zależnej od estymowanego parametru Θ i próby.
- **Przypadek 1** – znamy σ(X)
- Twierdzenie: Statystyka

$$Z \equiv \frac{\overline{X} - E(X)}{\sigma\left(\overline{X}\right)} = \frac{\left(\overline{X} - E(X)\right)\sqrt{n}}{\sigma(X)}$$



ma rozkład normalny N(0,1). (Średnia ma rozkład N(E(X),σ^2(X)/n), standaryzacja przekształca rozkład normalny w normalny).

Chcemy znaleźć przedział ufności taki, że

$$P\left(Z^L \le Z \le Z^P\right) = \gamma$$

Kwantyle rozkładu N(0,1)

Jest to spełnione przez $Z^L = Z_{\frac{1-\gamma}{2}}$ i $Z^P = Z_{\frac{1-\gamma}{2}+\gamma=\frac{1+\gamma}{2}}$

RPiS 2023/2024 11

Estymacja przedziałowa wartości oczekiwanej

- Dla rozkładu N(0,1) (z parzystości) $Z_{\frac{1-\gamma}{2}} = -Z_{\frac{1+\gamma}{2}}$, a stąd

$$P\left(-Z_{\frac{1+\gamma}{2}} \le Z \le Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(-Z_{\frac{1+\gamma}{2}} \le \frac{\left(\overline{X} - E(X)\right)\sqrt{n}}{\sigma(X)} \le Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(-\frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \le \left(\overline{X} - E(X)\right) \le \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(-\overline{X} - \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \le -E(X) \le -\overline{X} + \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

$$P\left(\overline{X} - \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \le E(X) \le \overline{X} + \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}\right) = \gamma$$

- Stąd

$$\boxed{T_n^L(E(X)) = \overline{X} - \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}} \qquad T_n^P(E(X)) = \overline{X} + \frac{1}{\sqrt{n}}\sigma(X)Z_{\frac{1+\gamma}{2}}}$$

RPiS 2023/2024 12