

Estymacja punktowa

Podstawowe pojęcia estymacji punktowej

Niech $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\theta; \theta \in \Theta\})$, będzie przestrzenią statystyczną. Na podstawie obserwacji $X \in \mathcal{X}$ oszacować $g(\theta) \in \mathcal{Y}$, gdzie $g: \Theta \rightarrow \mathcal{Y}$ jest znaną funkcją. Wartość $g(\theta)$ jest nieznana, gdyż nie znamy θ . Rozwiązaniem tego problemu będzie pewna funkcja $\hat{g}: \mathcal{X} \rightarrow \mathcal{Y}$ zwana estymatorem. Estymator może być uznany za dobry estymator, jeżeli funkcja $\hat{g}$ przyjmuje wartości bliskie wartościom $g(\theta) \forall \theta \in \Theta$. Oczywiście nie każdy estymator jest dobrym estymatorem. Wprowadza się więc pewne pojęcia umożliwiające porównywanie estymatorów i w konsekwencji wybór najlepszego z nich.

Założmy, że dana jest pewna funkcja $L: \mathcal{Y} \times \mathcal{Y} \rightarrow R$ zwana **funkcją strat**, której wartość $L(g(\theta), \hat{g}(X)) =$ określa stratę jaką ponosi statystyk przyjmując $\hat{g}(X)$ za oszacowanie nieznanej wielkości $g(\theta)$. Wobec tego $L(g(\theta), \hat{g}(X))$ jest, dla dowolnego ustalonego θ , zmienną losową określoną na przestrzeni prób $\mathcal{X}$. Określamy więc (o ile to możliwe) średnią (oczekiwaną) stratę będącą funkcją deterministyczną.

Def. Funkcję $R_L(\theta, \hat{g}) = E_\theta L(g(\theta), \hat{g}(X))$ parametru θ dla dowolnego ustalonego estymatora $\hat{g}$ nazywamy **funkcją ryzyka** estymatora $\hat{g}$ indukowaną przez funkcję strat L .

Estymatory będziemy porównywać ze sobą porównując ich funkcje ryzyka przy czym estymator jest tym lepszy im jego funkcja ryzyka przyjmuje mniejsze wartości.

Def. Estymator $\hat{g}_1$ **jest nie gorszy** od estymatora $\hat{g}_2$ w sensie ryzyka R_L indukowanego przez funkcję straty L jeżeli $R_L(\theta, \hat{g}_1) \leq R_L(\theta, \hat{g}_2) \forall \theta \in \Theta$.

Def. Estymator $\hat{g}_1$ **jest lepszy** niż $\hat{g}_2$ w sensie ryzyka R_L indukowanego przez funkcję straty L jeżeli $\hat{g}_1$ jest nie gorszy od estymatora $\hat{g}_2$ i $\exists \theta \in \Theta : R_L(\theta, \hat{g}_1) < R_L(\theta, \hat{g}_2)$.

Niech $\mathcal{D}$ oznacza wyspecyfikowany zbiór estymatorów. Zakładając, że każdemu estymatorowi z klasy $\mathcal{D}$ odpowiada funkcja ryzyka, możemy określić w zbiorze $\mathcal{D}$ relację quasi porządkującą (zwrotną i przechodnią)

$$\hat{g}_1 \leq \hat{g}_2 \Leftrightarrow \hat{g}_1 \text{ jest nie gorszy od } \hat{g}_2$$

Relacja taka w naturalny sposób generuje relację równoważnościową w $\mathcal{D}$

$$\hat{g}_1 = \hat{g}_2 \Leftrightarrow (\hat{g}_1 \leq \hat{g}_2 \text{ i } \hat{g}_2 \leq \hat{g}_1)$$

Ponadto w zbiorze ilorazowym (klas równoważności) określona jest w naturalny sposób relacja porządku częściowego

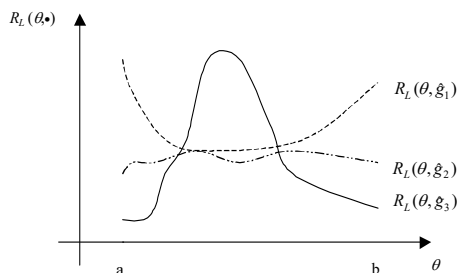
$$[\hat{g}_1] \leq [\hat{g}_2] \Leftrightarrow \hat{g}_1 \leq \hat{g}_2$$

Nie wszystkie estymatory (właściwie ich klasy równoważności) są porównywalne w powyższym sensie. Jeśli funkcje ryzyka danych estymatorów przecinają się (dla pewnych θ niższe wartości przyjmuje jedna z nich a dla innych θ druga), to estymatory są nieporównywalne w powyższym sensie.

Def. Estymator $\hat{g}_1$ nazywamy **niedopuszczalnym** w $\mathcal{D}$ w sensie ryzyka R_L indukowanego przez funkcję strat L jeżeli istnieje w zbiorze $\mathcal{D}$ estymator $\hat{g}_2$ lepszy od $\hat{g}_1$.

Ze zbioru $\mathcal{D}$ rozważanych estymatorów można usunąć estymatory niedopuszczalne i ograniczyć rozważania jedynie do zbioru estymatorów dopuszczalnych $\mathcal{D}_{\text{dop}}$. Niestety, zwykle nie udaje się dla rozważanego problemu scharakteryzować klasy estymatorów dopuszczalnych. Czasami udaje się udowodnić dopuszczalność konkretnego estymatora uzyskanego z rozważań optymalizacyjnych lub heurystycznych.

Niech $\Theta=[a,b]$. Rozważmy trójelementowy zbiór estymatorów $\mathcal{D}=\{\hat{g}_1, \hat{g}_2, \hat{g}_3\}$ pewnej wielkości $g(\theta)$ o funkcjach ryzyka przedstawionych na rysunku.



Widać, że estymator $\hat{g}_1$ jest niedopuszczalny, gdyż lepszym estymatorem jest $\hat{g}_2$.

Porównując estymatory poprzez porównywanie ich funkcji ryzyka możemy odrzucić pewne estymatory (niedopuszczalne). Pozostałe estymatory (dopuszczalne) są nieporównywalne w powyższym sensie, gdyż ich funkcje ryzyka wzajemnie się przecinają. Ponadto, praktyk wolałby otrzymać jakiś jeden estymator (najlepiej optymalny) zamiast zbioru dopuszczalnych estymatorów z którego i tak w końcu musi wybrać pewien konkretny estymator. Pokonać te trudności można na różne sposoby. Wymienić tu należy :

- podejście polegające na utrzymaniu kryterium porównywania estymatorów poprzez porównywanie ich funkcji ryzyka i ograniczaniu klasy rozważanych estymatorów np. estymatorów nieobciążonych. Ograniczanie klasy estymatorów jest konieczne, gdyż w klasie wszystkich estymatorów posiadających funkcje ryzyka przy założonej funkcji straty nie istnieje estymator o jednostajnie minimalnym ryzyku. Rzeczywiście, jeśli jako zbiór $\mathcal{D}$ estymatorów rozważymy zbiór wszystkich estymatorów dla których potrafimy wyznaczyć funkcje ryzyka, to taki optymalny estymator nie istnieje, gdyż wybierając estymator stały $\hat{g}_0(X)=g(\theta_0)$ dla pewnego $\theta_0 \in \Theta$ uzyskujemy $R_L(\theta, \hat{g}_0)=0$ w punkcie θ_0 . Tak więc dla estymatora $\hat{g}$ o jednostajnie

najmniejszym ryzyku musi być $R_L(\theta_0, \hat{g}) = 0$. Z dowolności θ_0 wynika, że optymalny estymator musiałby mieć ryzyko stale równe 0, co jest oczywiście niemożliwe, gdyż wymaga znajomości θ . Ponieważ rozważana klasa wszystkich estymatorów zawiera tak "bezsensowne" estymatory jak estymatory stałe, nie wykorzystujące obserwacji $X \in \mathcal{X}$, wobec tego całkiem naturalne jest ograniczenie klasy rozważanych estymatorów. Interesującą klasę stanowią tzw. estymatory nieobciążone.

Def. Estymator $\hat{g}$ wielkości $g(\theta)$ nazywamy **nieobciążonym** jeżeli spełnia warunek

$$E_{\theta} \hat{g}(X) = g(\theta) \quad \forall \theta \in \Theta.$$

Wielkość $b_{\theta}(\hat{g}) = b(\theta, \hat{g}) = E_{\theta} \hat{g}(X) - g(\theta)$ nazywamy **obciążeniem (bias)** estymatora $\hat{g}$.

Nieobciążoność estymatora, która wyraża jego bezstronność (neutralność) wyrażającą się w braku skłonności do przeszacowywania bądź niedoszacowywania estymowanej wielkości przez estymator, jest pozytywną cechą estymatora, której nie należy jednak demonizować. Nieobciążoność jest szczególnie cenną własnością dopiero w przypadku gdy estymator ma niewielką wariancję.

Rozważmy pewien szczególny przypadek problemu estymacji punktowej. Niech $g: \Theta \ni \theta \rightarrow R$ będzie daną funkcją rzeczywistą, której wartość $g(\theta) \in R$ należy oszacować na podstawie obserwacji $X = (X_1, \dots, X_n)$. Przyjmijmy **kwadratową funkcję strat**

$$L(u, v) = (v - u)^2.$$

Wobec tego $L(g(\theta), \hat{g}(X)) = (\hat{g}(X) - g(\theta))^2$ jest kwadratem błędu oszacowania $g(\theta)$ poprzez $\hat{g}(X)$ i jest wielkością losową. Funkcja ryzyka estymatora $\hat{g}$ jest równa

$$R_L(\theta, \hat{g}) = E_{\theta} (\hat{g}(X) - g(\theta))^2$$

jest nazywana **błędem średniokwadratowym BSK** (ang. **MSE** mean square error) estymatora $\hat{g}$.

Łatwo zauważyć, że dla kwadratowej funkcji straty

$$R_L(\theta, \hat{g}) = E_{\theta} (\hat{g}(X) - g(\theta))^2 = E_{\theta} (\hat{g}(X) - E_{\theta}(\hat{g}) + E_{\theta}(\hat{g}) - g(\theta))^2 = V_{\theta}(\hat{g}) + b_{\theta}^2(\hat{g})$$

Ryzyko estymatora jest sumą jego wariancji i kwadratu obciążenia. Ta dekompozycja pokazuje, że czasami warto poszerzyć klasę estymatorów nieobciążonych o estymatory obciążone, gdyż niewielkie obciążenie może zostać zrekomensowane obniżką wariancji tak, że BSK estymatora obciążonego może być niższy od BSK najlepszego estymatora nieobciążonego.

Uwaga. Estymator nieobciążony o minimalnej wariancji może być niedopuszczalny

Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu $N(m, \sigma^2)$. Można pokazać, że nieobciążonym estymatorem σ^2 jest estymator $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. (Uwaga: S^2 jest funkcją statystyki dostatecznej zupełnej $(\bar{X}, S^2)$, więc jest to estymator nieobciążony o jednostajnie (względem σ^2) minimalnej wariancji ENJMW[σ^2]).

Wiadomo z tw. Fishera, że zmienna losowa $Y = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$ ma rozkład χ^2_{n-1} . Stąd

$$E_{(m, \sigma)}(Y) = n-1, \quad V_{(m, \sigma)}(Y) = 2(n-1), \quad E_{(m, \sigma)}(Y^2) = V(Y) + E^2(Y) = n^2 - 1.$$

Rozważmy klasę estymatorów $\mathcal{K} = \{\hat{\sigma}_c^2 = c \sum_{i=1}^n (X_i - \bar{X})^2, c > 0\}$. Oczywiście $S^2 \in \mathcal{K}$. (w tym przypadku $c = \frac{1}{n-1}$). Wyznamy funkcję ryzyka estymatora $\hat{\sigma}_c^2$ przy kwadratowej funkcji straty.

$$\begin{aligned} R(\sigma^2, \hat{\sigma}_c^2) &= E_{(m, \sigma)} \left(c \sum_{i=1}^n (X_i - \bar{X})^2 - \sigma^2 \right)^2 = \sigma^4 E_{(m, \sigma)} \left(c \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} - 1 \right)^2 = \sigma^4 E_{(m, \sigma)} (cY - 1)^2 = \\ &= \sigma^4 \{c^2 E_{(m, \sigma)}(Y^2) - 2c E_{(m, \sigma)}(Y) + 1\} = \sigma^4 \{c^2 E_{(m, \sigma)}(Y^2) - 2c E_{(m, \sigma)}(Y) + 1\} = \\ &= \sigma^4 \{c^2 (n^2 - 1) - 2c(n-1) + 1\}. \end{aligned}$$

Minimalizując $R(c)$ otrzymujemy $c_{\min} = \frac{1}{n+1}$. Ryzyko (przy kwadratowej funkcji strat) obciążonego estymatora $\hat{\sigma}_{c_{\min}}^2 = \frac{1}{n+1} \sum_{i=1}^n (X_i - \bar{X})^2$ równe

$$R(\sigma^2, \hat{\sigma}_{c_{\min}}^2) = \frac{2\sigma^4}{n+1} < \frac{2\sigma^4}{n-1} = R(\sigma^2, S^2)$$

jest jednostajnie mniejsze od ryzyka estymatora S^2 , który jest w tej sytuacji niedopuszczalny w klasie $\mathcal{K}$ przy kwadratowej funkcji strat.

Estymatory nieobciążone o jednostajnie minimalnej wariancji, nazywane także estymatorami **najefektywniejszymi** (przy kwadratowej funkcji strat), potrafimy efektywnie konstruować tylko w pewnych szczególnych przypadkach. Jeżeli nie potrafimy ustalić, czy istnieje estymator nieobciążony o jednostajnie minimalnej wariancji, to otwiera się inna możliwość. Okazuje się, że przy pewnych technicznych założeniach dotyczących regularności estymatora można podać oszacowanie od dołu (dokładnie kres dolny) wariancji estymatorów nieobciążonych. Możemy wobec tego porównywać wariancję rozważanego estymatora nieobciążonego, uzyskanego na innej drodze, z wartością kresu dolnego wariancji, czyli szacować jego efektywność. Może okazać się, że badany estymator ma wariancję niewiele większą od kresu dolnego wariancji wszystkich (regularnych) estymatorów nieobciążonych i wobec tego jest zadowalający z praktycznego punktu widzenia.

Twierdzenie (Cramer-Rao). Niech $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ będzie rodziną rozkładów na przestrzeni prób $\mathcal{X}$ mających gęstości $p(\theta, x)$ względem pewnej ustalonej miary, i niech Θ będzie zbiorem

otwartym w R . Jeżeli są spełnione pewne warunki regularności, to wariancja każdego estymatora nieobciążonego $\hat{g}(X)$ wielkości $g(\theta)$ spełnia nierówność (Cramera-Rao)

$$Var_{\theta}[\hat{g}(X)] \geq \frac{\left[\frac{dg(\theta)}{d\theta} \right]^2}{E_{\theta} \left[\frac{\partial \ln p(\theta, X)}{\partial \theta} \right]^2}.$$

Wielkość $I_{\theta} = Var_{\theta} \left[\frac{\partial \ln p(\theta, X)}{\partial \theta} \right] = E_{\theta} \left[\frac{\partial \ln p(\theta, X)}{\partial \theta} \right]^2$ (mianownik w nierówności Cramera-Rao)

nazywamy informacją w sensie Fishera o parametrze θ zawartą w próbie (obserwowanej zmiennej losowej, zwykle wektorowej) X . Jeżeli (regularny) estymator nieobciążony ma wariancję

równą dolnemu ograniczeniu Cramera-Rao $D_{CR} = \frac{\left[\frac{dg(\theta)}{d\theta} \right]^2}{E_{\theta} \left[\frac{\partial \ln p(\theta, X)}{\partial \theta} \right]^2}$, to jest on estymatorem

najefektywniejszym w klasie estymatorów regularnych. Efektywnością w sensie Cramera-Rao estymatora nieobciążonego $\hat{g}$ o wariancji $Var_{\theta}(\hat{g})$ nazywamy wielkość

$$eff_{CR}(\hat{g}) = \frac{D_{CR}}{Var_{\theta}(\hat{g})}.$$

Przykład. Niech $X_1, \dots, X_n$ będzie ciągiem niezależnych zmiennych losowych o tym samym

rozkładzie (iid) $N(\theta, 1)$. Wówczas $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ jest nieobciążonym estymatorem parametru θ (w

tym przypadku $g(\theta) = \theta$), gdyż $E_{\theta}(\bar{X}) = \frac{1}{n} \sum_{i=1}^n E_{\theta}(X_i) = \frac{1}{n} n\theta = \theta$. Wariancja estymatora $\bar{X}$ jest

równa

$$\begin{aligned} V_{\theta}(\bar{X}) &= E_{\theta} \left(\frac{1}{n} \sum_{i=1}^n X_i - \theta \right)^2 = E_{\theta} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \theta) \right)^2 = \frac{1}{n^2} \sum_{i,j=1}^n E_{\theta} (X_i - \theta)(X_j - \theta) = \\ &= \frac{1}{n^2} \sum_{i,j=1}^n Cov(X_i, X_j) = \frac{1}{n} \end{aligned}$$

i jest równa dolnemu ograniczeniu Cramera-Rao. Rzeczywiście

$$p(\theta, X_1, \dots, X_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2}, \quad \ln p(\theta, X_1, \dots, X_n) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2$$

$$\frac{\partial}{\partial \theta} \ln p(\theta, X_1, \dots, X_n) = \sum_{i=1}^n (X_i - \theta), \quad \left[\frac{\partial}{\partial \theta} \ln p(\theta, X_1, \dots, X_n) \right]^2 = \sum_{i,j=1}^n (X_i - \theta)(X_j - \theta)$$

wiec $I_{\theta} = \sum_{i,j=1}^n E_{\theta} (X_i - \theta)(X_j - \theta) = \sum_{i,j=1}^n Cov(X_i, X_j) = n$, a stąd $Var_{\theta}(\bar{X}) = I_{\theta}^{-1} = \frac{1}{n}$.

- **podejście minimaksowe** polegające na porównywaniu estymatorów poprzez porównywanie maksimum globalnych ich funkcji ryzyka $\max_{\theta \in \Theta} R_L(\theta, \hat{g})$, przy czym estymator jest tym lepszy im ma mniejsze maksimum funkcji ryzyka.

Def. Estymator $\hat{g}_m$ jest estymatorem minimaksowym wielkości $g(\theta)$ w rozważanej klasie estymatorów $\mathcal{D}$, $\forall \hat{g} \in \mathcal{D}$ $\max_{\theta \in \Theta} R_L(\theta, \hat{g}_m) \leq \max_{\theta \in \Theta} R_L(\theta, \hat{g})$.

Z uwagi na to, że nie ma generalnej potrzeby zakładania złośliwości natury, podejście minimaksowe nie cieszy się zbyt dużym powodzeniem wśród praktyków.

- **podejście bayesowskie** polegające na porównywaniu pewnych średnich wartości funkcji ryzyka dla poszczególnych estymatorów. Zakłada się tu, że statystyk posiada pewną wiedzę *a priori* o parametrze θ w postaci tak zwanego rozkładu *a priori* ν określonego na mierzalnej przestrzeni parametrów. Każdemu estymatorowi $\hat{g}$ przypisujemy wartość (liczbową) ryzyka bayesowskiego

$$r_\nu(\hat{g}) = E_\nu[R_L(\theta, \hat{g})]$$

względem rozkładu *a priori*, które jest średnią względem rozkładu *a priori* wartością funkcji ryzyka R_L i estymator jest tym lepszy im ma mniejsze ryzyko bayesowskie

Zgodność i mocna zgodność estymatorów

Niech $X=(X_1, \dots, X_n)$ będzie n elementową próbą prostą z rozkładu $P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$. Niech

$\hat{g}_n(X_1, \dots, X_n)$ będzie estymatorem funkcji $g(\theta)$ opartym na n elementowej próbie. Intuicyjnie można się spodziewać, że powiększanie rozmiaru próby n powinno skutkować tym, że $\hat{g}_n(X_1, \dots, X_n)$ jest „coraz lepszą oceną” $g(\theta)$. Interesujące jest graniczne zachowanie się estymatora $\hat{g}_n$, gdy $n \rightarrow \infty$

Def. Ciąg $\hat{g}_n$ estymatorów wielkości $g(\theta)$ nazywamy

$$\begin{array}{ll} \text{zgodnym} & \text{gdy } \forall \theta \in \Theta \quad \hat{g}_n(X) \xrightarrow{wg P_\theta} g(\theta), \quad n \rightarrow \infty \\ \text{mocno zgodnym} & \text{gdy } \forall \theta \in \Theta \quad \hat{g}_n(X) \xrightarrow{z pr. 1(P_\theta)} g(\theta), \quad n \rightarrow \infty. \end{array}$$

Momenty empiryczne

Niech $X=(X_1, \dots, X_n)$ będzie n elementową próbą prostą z rozkładu $P_\theta \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$ o skończonych momentach zwykłych $m_r = E_\theta(X_1^r)$ do rzędu k (tzn. $r \leq k$).

Definicja. Statystykę $M_{r,n} = \frac{1}{n} \sum_{i=1}^n X_i^r$ nazywamy momentem empirycznym (próbkowym) zwykłym rzędu r opartym na n elementowej próbie prostej.

Fakty

- $E_\theta(M_{r,n}) = E_\theta(\frac{1}{n} \sum_{i=1}^n X_i^r) = \frac{1}{n} \sum_{i=1}^n E_\theta(X_i^r) = \frac{1}{n} \sum_{i=1}^n m_r = m_r$, czyli momenty empiryczne (zwykle)

są nieobciążonymi estymatorami momentów teoretycznych.

- Z MPWL Kołmogorowa zastosowanego do ciągu $X_1^r, \dots, X_n^r, \dots$ wynika, że jeżeli istnieje m_r , to

$$\forall \theta \in \Theta \quad M_{r,n} \xrightarrow{\text{z pr}(P_\theta) \rightarrow 1} m_r$$

- Podobnie z CTG Lindeberga Levy'ego zastosowanego do ciągu $X_1^r, \dots, X_n^r, \dots$ wynika, że jeżeli istnieje m_{2r} , to $\forall \theta \in \Theta$ „statystyka” $\sqrt{n} \frac{M_{r,n} - m_r}{\sqrt{m_{2r} - m_r^2}}$ ma asymptotycznie ($n \rightarrow \infty$) rozkład

normalny $N(0,1)$

- Analogiczne fakty prawdziwe są dla p wymiarowych wektorów losowych $\mathbf{X}_1, \dots, \mathbf{X}_n, \dots$, gdzie

$$\mathbf{X}_i = \begin{bmatrix} X_{1,i} \\ \vdots \\ X_{p,i} \end{bmatrix}, \quad \mathbf{X}_i^r = \begin{bmatrix} X_{1,i}^r \\ \vdots \\ X_{p,i}^r \end{bmatrix}, \quad \mathbf{m}_r = \begin{bmatrix} E_\theta(X_1^r) \\ \vdots \\ E_\theta(X_p^r) \end{bmatrix}, \quad \mathbf{M}_{r,n} = \begin{bmatrix} \frac{1}{n} \sum_{i=1}^n X_{1,i}^r \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n X_{p,i}^r \end{bmatrix}$$

Jeżeli istnieje $\mathbf{m}_r$, to statystyka wektorowa $\mathbf{M}_{r,n} \xrightarrow{\text{z pr}(P_\theta) \rightarrow 1} \mathbf{m}_r \quad \forall \theta \in \Theta$.

Jeżeli istnieje $\mathbf{m}_{2r}$, to statystyka wektorowa $\sqrt{n}(\mathbf{M}_{r,n} - \mathbf{m}_r)$ ma asymptotycznie rozkład

$$N_p(\mathbf{0}, \mathbf{V}(\mathbf{X}^r)), \text{ gdzie } \mathbf{V}(\mathbf{X}^r) = [Cov(X_i^r, X_j^r)]_{(p,p)}$$

- Niech $g: R^k \rightarrow R^q$ będzie funkcją borelowsko mierzalną i prawie wszędzie ciągłą. Jeżeli istnieją momenty $\mathbf{m}_1, \dots, \mathbf{m}_k$, to $g(\mathbf{M}_{1,n}, \dots, \mathbf{M}_{k,n}) \xrightarrow{\text{z prawd. } P_\theta \rightarrow 1} g(\mathbf{m}_1, \dots, \mathbf{m}_k) \quad (\forall \theta \in \Theta)$.

Metoda delta. ($\Rightarrow$ oznacza słabą zbieżność). Niech

- $v_n(X_n(\omega) - \mathbf{x}_0) \Rightarrow X(\omega)$, gdzie wektory losowe $X_n(\omega)$, $X(\omega)$ oraz nielosowy wektor $\mathbf{x}_0$ (zwykle $\mathbf{x}_0 = E(X)$) przyjmują wartości w przestrzeni R^k , a ciąg liczbowy $v_n \rightarrow \infty$ (zwykle $v_n = \sqrt{n}$)
- funkcja $g: R^k \supset \text{Ot}(x_0) \rightarrow R^q$ jest różniczkowalna w punkcie x_0

Wówczas $v_n(g(X_n(\omega)) - g(\mathbf{x}_0)) \Rightarrow [g'(\mathbf{x}_0)]X(\omega)$

Zadania

1. Niech $\mathbf{X} = (X_1, \dots, X_n)$ i $\mathbf{Y} = (Y_1, \dots, Y_n)$ będą niezależnymi próbami prostymi z rozkładów odpowiednio $N(m_x, \sigma^2)$ i $N(m_y, \sigma^2)$. Który z dwóch następujących estymatorów:

$$T_1(\mathbf{X}, \mathbf{Y}) = \overline{X} \overline{Y}, \quad T_2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n} \sum_{i=1}^n X_i Y_i \quad \text{należy przyjąć za ocenę } m_x m_y \text{ biorąc BSK jako miarę dobroci}$$

estymatora. (Odp. $V(T_1) = \frac{1}{n^2} \sigma^4 + \frac{1}{n} \sigma^2 (m_x^2 + m_y^2)$, $V(T_2) = \frac{1}{n} \sigma^4 + \frac{1}{n} \sigma^2 (m_x^2 + m_y^2)$) - lepszy jest T_1

2. Niech $X_1, \dots, X_k, X_{k+1}, \dots, X_n$ będzie próbą prostą z rozkładu $N(m, \sigma^2)$. Obserwujemy zmienne $X_1, \dots, X_k$ i ponadto znamy średnią $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Dobrać tak stałą c_{nk} aby estymator $T = c_{nk} \sum_{i=1}^k (X_i - \bar{X}_n)^2$ był nieobciążonym estymatorem wariancji σ^2 . Odp $c_{nk} = \frac{n}{k(n-1)}$
3. Wykonano 10 pomiarów pewnej nieznannej wielkości m jednym przyrządem pomiarowym, a następnie 5 pomiarów innym przyrządem. Zakładamy, że wyniki pomiarów są $X_1, \dots, X_{10}, X_{11}, \dots, X_{15}$ są niezależnymi zmiennymi losowymi przy czym każda ze zmiennych $X_1, \dots, X_{10}$ ma rozkład normalny $N(m, 0.1^2)$, podczas, gdy każda ze zmiennych $X_{11}, \dots, X_{15}$ ma rozkład normalny $N(m, 0.2^2)$. Dobrać tak współczynniki $c_1, \dots, c_{15}$ aby estymator $\hat{m} = \sum_{i=1}^{15} c_i X_i$ był estymatorem nieobciążonym o minimalnej wariancji. (Odp. $c_1, \dots, c_{10} = \frac{4}{45}$, $c_{11}, \dots, c_{15} = \frac{1}{45}$)
4. Pobrano 100 niezależnych obserwacji z rozkładu normalnego $N(m, \sigma^2)$. Obliczono 10 sum po 10 kolejnych obserwacji a następnie zgubiono dane źródłowe. Zamiast pierwotnych obserwacji $(X_1, \dots, X_{100})$ mamy obserwacje $(Y_1, \dots, Y_{10})$ gdzie $Y_i = \sum_{j=0}^9 X_{10i-j}$. Szacujemy wariancję σ^2 używając estymatora postaci $c \sum_{i=1}^{10} (Y_i - \bar{Y})^2$. Dobrać tak stałą c aby estymator ten był nieobciążony. (Odp. $c = \frac{1}{90}$)
5. Niech $X_1, X_2, \dots, X_n$ będzie próbą prostą z rozkładu jednostajnego $U(0, \theta)$. Rozważmy dwa estymatory $\hat{g}_1(X_1, \dots, X_n) = \frac{n+1}{n} X_{(n)}$ i $\hat{g}_2(X_1, \dots, X_n) = 2\bar{X}$. Który z tych estymatorów jest lepszy?.
- Odp. $R(\hat{g}_1, \theta) = \frac{\theta^2}{n(n+2)}$, $R(\hat{g}_2, \theta) = \frac{\theta^2}{3n}$ lepszy jest $\hat{g}_1$.