

Metody estymacji

Metoda momentów

Metodę momentów pochodzącą od K. Pearsona można krótko opisać w następujący sposób:

Niech $X=(X_1, X_2, \dots, X_n)$ będzie próbą prostą z rozkładu P_θ zależnego od wektorowego parametru $\theta=(\theta_1, \dots, \theta_k)$.

- wyznaczamy k pierwszych momentów zwykłych: $m_r = E_{(\theta_1, \dots, \theta_k)} X_1^r = h_r(\theta_1, \dots, \theta_k)$, $r=1, \dots, k$.

- estymujemy momenty teoretyczne m_r momentami empirycznymi $\hat{m}_r = \frac{1}{n} \sum_{i=1}^n X_i^r$

- rozwiązujemy układ równań
$$\begin{cases} \hat{m}_1 = h_1(\theta_1, \dots, \theta_k) \\ \vdots \\ \hat{m}_k = h_k(\theta_1, \dots, \theta_k) \end{cases}$$
 i otrzymujemy estymatory

$$\begin{cases} \hat{\theta}_1(X) = g_1(\hat{m}_1, \dots, \hat{m}_k) \\ \vdots \\ \hat{\theta}_k(X) = g_k(\hat{m}_1, \dots, \hat{m}_k) \end{cases} \quad \text{będące zwykle ciągłymi funkcjami momentów empirycznych, a więc}$$

mocno zgodne.

Przykład. Niech $X=(X_1, X_2, \dots, X_n)$ będzie próbą prostą z rozkładu o skończonych momentach do rzędu 2 włącznie. Wiadomo, że $E(X_i)=m$ i $E(X_i^2)=m^2 + \sigma^2$. Z układu

$$\begin{cases} \hat{m}_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} = m \\ \hat{m}_2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \sigma^2 + m^2 \end{cases} \quad \text{otrzymujemy} \quad \begin{cases} \hat{m} = \bar{X} \\ \hat{\sigma}^2 = \hat{m}_2 - \hat{m}_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = S_0^2 \end{cases}$$

Estymatory empiryczne $\hat{m}_1$ i $\hat{m}_2$ są mocno zgodnymi i nieobciążonymi estymatorami momentów m_1 i

m_2 , natomiast estymator wariancji $S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ jest obciążony. Rzeczywiście

oznaczając $\theta = (m, \sigma^2)$ mamy

$$\begin{aligned} E_\theta \left(\sum_{i=1}^n (X_i - \bar{X})^2 \right) &= E_\theta \left(\sum_{i=1}^n (X_i - m + m - \bar{X})^2 \right) = E_\theta \left(\sum_{i=1}^n (X_i - m)^2 + 2(X_i - m)(m - \bar{X}) + (m - \bar{X})^2 \right) = \\ &= \sum_{i=1}^n E_\theta (X_i - m)^2 + 2E_\theta \left((m - \bar{X}) \sum_{i=1}^n (X_i - m) \right) + nE_\theta (\bar{X} - m)^2 = n\sigma^2 - nE_\theta (\bar{X} - m)^2 = (n-1)\sigma^2 \end{aligned}$$

gdyż

$$E_\theta (\bar{X} - m)^2 = E_\theta \left(\frac{1}{n} \sum_{i=1}^n (X_i - m) \right)^2 = \frac{1}{n^2} \left(\sum_{i,j=1}^n E_\theta (X_i - m)(X_j - m) \right) = \frac{1}{n^2} E_\theta \sum_{i=1}^n E((X_i - m)^2) = \frac{1}{n} \sigma^2.$$

Wobec tego $E_\theta(S_0^2) = \frac{n-1}{n} \sigma^2$ a zmodyfikowany estymator $S^2 = \frac{n}{n-1} S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ jest już zgodnym i nieobciążonym estymatorem wariancji.

Metoda podstawienia dystrybuanty empirycznej

Przypuśćmy, że interesujący nas parametr η rozkładu jest znaną funkcją dystrybuanty F , czyli

$$\eta = g(F). \text{ Np. } r\text{-ty moment zwykły } m_r = \int_{\Omega} X^r(\omega) dP_F = \int_R x^r dF \text{ jest funkcją (funkcjonałem)}$$

dystrybuanty rozkładu. Zwykle wprowadzając odpowiednie pojęcia topologiczne (np. metrykę, normę) w przestrzeni funkcyjnej zawierającej zbiór dystrybuant można pokazać, że interesujący nas funkcjonal jest ciągły. Wobec tego jego wartość „niewiele” się zmieni, gdy dystrybuantę zastąpimy jej „dobrym przybliżeniem”, czyli dystrybuantą empiryczną. Uzyskujemy w ten sposób naturalny estymator $\hat{\eta} = g(\hat{F})$ będący funkcją (funkcjonałem) dystrybuanty empirycznej np.

$$\hat{m}_r = \int_R x^r d\hat{F}_n = \frac{1}{n} \sum_{i=1}^n X_i^r = M_{r,n}. \text{ Intuicyjne widać, że ciągłość funkcjonalu i twierdzenie Gliwienki}$$

Cantellego implikują mocną zgodność estymatorów uzyskiwanych metodą podstawienia dystrybuanty.

Ponadto $E_F(\hat{m}_r) = \int_R x^r E_F(d\hat{F}_n) = \int_R x^r dF = m_r$, więc rozważany estymator momentu zwykłego rzędu r jest nieobciążony. Zaprezentowane powyżej intuicje wymagają oczywiście stosownych uściśleń.

Metoda Markowa

Metoda Markowa pochodząca z przełomu XIX i XX wieku polega na wyznaczaniu estymatorów

- nieobciążonych
- liniowych względem obserwowanych zmiennych losowych
- posiadających najmniejszą wariancję w klasie wszystkich estymatorów liniowych

Metodę Markowa można więc potraktować jako szczególny przypadek estymacji nieobciążonej o minimalnej wariancji, który ze względu na postulowaną liniowość jest szczególnie prosty i wygodny obliczeniowo. Wyjaśnimy to na przykładzie.

Przykład. Niech $X_1, \dots, X_n$ będzie ciągiem niezależnych zmiennych losowych o tych samych wartościach oczekiwanych $E(X_i) = m$ i znanych wariancjach $V(X_i) = \sigma_i^2$ (Uwaga $X = (X_1, X_2, \dots, X_n)$ nie jest próbą prostą). Znaleźć nieobciążony estymator liniowy dla m o najmniejszej wariancji.

Z uwagi na postulowaną liniowość estymator wartości oczekiwanej m ma postać $\hat{m} = \sum_{i=1}^n a_i X_i$.

Warunek nieobciążoności $m = E(\hat{m}) = \sum_{i=1}^n a_i E(X_i) = m \sum_{i=1}^n a_i$ prowadzi do równości $\sum_{i=1}^n a_i = 1$. Pisząc

dla prostoty E zamiast E_θ i V zamiast V_θ wariancja estymatora

$$V(\hat{m}) = E\left(\sum_{i=1}^n a_i X_i - m\right)^2 = E\left(\sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i m\right)^2 = E\left(\sum_{i=1}^n a_i (X_i - m)\right)^2 = \sum_{i=1}^n a_i^2 \sigma_i^2.$$

Aby znaleźć estymator typu Markowa należy więc znaleźć współczynniki $a_1, \dots, a_n$ minimalizujące

$\sum_{i=1}^n a_i^2 \sigma_i^2$ przy warunku $\sum_{i=1}^n a_i = 1$. Rozwiązując powyższy problem ekstremum warunkowego poprzez

rozwikłanie ograniczeń i redukcję problemu do problemu ekstremum bezwarunkowego $n-1$

zmiennych lub stosując metodę mnożników Lagrange'a otrzymujemy $a_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}$.

Metoda najmniejszych kwadratów MNK

Obserwujemy zmienne losowe $Y_1, \dots, Y_n$ o których wiemy, że

$$E(Y_i) = g_i(\theta), \quad i=1, \dots, n,$$

gdzie $g_i: R^k \supset \Theta \ni \theta \rightarrow \mathbb{R}$ są znanymi funkcjami.

Jeżeli parametr θ przebiega zbiór Θ , to punkt $(g_1(\theta), \dots, g_n(\theta))$ przebiega pewien zbiór $\Gamma \subset R^n$.

Zaobserwowany punkt $Y = (Y_1, \dots, Y_n)$ również leży w R^n . Idea MNK polega na tym, żeby w zbiorze Γ

znaleźć punkt $\gamma(Y_1, \dots, Y_n)$ najbliższy zaobserwowanemu punktowi Y a następnie za oszacowanie

parametru θ przyjąć taki punkt $\hat{\theta} \in \Theta$, któremu odpowiada wyznaczony punkt γ , tzn. taki $\hat{\theta}$, że

$$(g_1(\hat{\theta}), \dots, g_n(\hat{\theta})) = \gamma(Y_1, \dots, Y_n).$$

Zwykle oba etapy łączy się w jeden i za estymator MNK przyjmuje się θ minimalizujące wielkość

$$\sum_{i=1}^n (Y_i - g_i(\theta))^2, \text{ czyli } \hat{\theta}(Y_1, \dots, Y_n) = \arg \min_{\theta \in \Theta} \sum_{i=1}^n (Y_i - g_i(\theta))^2$$

MNK znajduje szczególne zastosowanie w tzw. liniowych modelach statystyki matematycznej i zostanie omówiona później.

Metoda największej wiarygodności

Niech $(\mathcal{X}, \mathcal{B}, \mathcal{P} = \{P_\theta: \theta \in \Theta \subset R^k\})$, będzie dominowaną przez σ -skończoną miarę μ przestrzenią

statystyczną. Oznaczmy przez $p(x; \theta) = p_\theta(x) = \frac{dP_\theta}{d\mu}$ odpowiednie gęstości.

Def. Dla ustalonego wyniku eksperymentu $X \in \mathcal{X}$ wielkość $L(\theta, X) = p(X, \theta)$ nazywamy wiarygodnością parametru θ , a funkcję $L(\cdot, X): \Theta \ni \theta \rightarrow L(\theta, X)$ określoną na przestrzeni parametrów Θ nazywamy **funkcją wiarygodności**.

Z definicji widać, że funkcja $L(\theta, X)$ dla każdego ustalonego θ jest funkcją gęstości $p(X, \theta)$ rozkładu prawdopodobieństwa P_θ na przestrzeni prób $\mathcal{X}$. Interpretacja funkcji wiarygodności jest szczególnie prosta, gdy rozważymy dyskretną przestrzeń statystyczną. Funkcja wiarygodności przypisuje parametrowi θ prawdopodobieństwo (wyznaczone z rozkładu P_θ) zaobserwowania danego wyniku eksperymentu $X \in \mathcal{X}$. W przypadku ciągłym interpretacja funkcji wiarygodności jest podobna. Prawdopodobieństwo zaobserwowania danego wyniku eksperymentu zastępujemy gęstością prawdopodobieństwa uzyskania danego wyniku eksperymentu. Ten sam wynik eksperymentu może mieć przypisane różne prawdopodobieństwa w zależności od wyboru parametru θ (czyli rozkładu P_θ). **Zasada największej wiarygodności sugeruje taki wybór parametru θ , przy którym zaobserwowany wynik eksperymentu $x \in \mathcal{X}$ jest najbardziej prawdopodobny.**

W przypadku przestrzeni produktowej $(\mathcal{R}, \mathcal{B}(\mathcal{R}), \{p(x, \theta): \theta \in \Theta\})^n$ gdy wynik eksperymentu jest ciągiem $X = (X_1, \dots, X_n)$, funkcja wiarygodności wyraża się wzorem $L(\theta, X_1, \dots, X_n) = \prod_{i=1}^n p(X_i, \theta)$.

Def. Estymatorem największej wiarygodności parametru θ (ENW(θ)) (o ile istnieje) nazywamy estymator $\hat{\theta}: \mathcal{X} \ni X \rightarrow \hat{\theta}(X) = \arg \max_{\theta \in \Theta} L(\theta, X)$

Uwaga techniczna. Ze względu na monotoniczność funkcji logarytmicznej maksymalizacja funkcji wiarygodności $L(\theta, X)$ jest równoważna maksymalizacji jej logarytmu $l(\theta, X) = \ln L(\theta, X)$.

Kłopoty z metodą największej wiarygodności

- estymator największej wiarygodności może nie istnieć,
- estymator największej wiarygodności może nie być określony jednoznacznie,
- efektywne wyznaczenie estymatora największej wiarygodności może być bardzo trudne.

Estymatory NW mają wiele cennych własności w próbach skończonych (szczegóły będą podane w wykładzie Statystyki Matematycznej II). Przy pewnych założeniach regularności (tzw. **warunki regularności Cramera**) (zapewniających różniczkowalność całek niewłaściwych) estymatory największej wiarygodności są:

- mocno zgodne
- asymptotycznie nieobciążone
- asymptotycznie najefektywniejsze (asymptotycznie osiągają dolne ograniczenie CR)

- niezmiennicze tzn. jeśli $\hat{\theta}(X) = ENW(\theta)$ i $\eta = h(\theta)$ to $\hat{\eta}(X) = h(\hat{\theta}(X)) = ENW(\eta)$
- asymptotycznie normalne tzn. $\sqrt{n}(\hat{\theta} - \theta)$ ma asymptotycznie rozkład $N(0, \sigma_{as}^2)$, gdzie asymptotyczna wariancja $\sigma_{as}^2 = i_{\theta}^{-1}$, a $i_{\theta} = E_{\theta} \left(\frac{\partial}{\partial \theta} \ln(p(X_1, \theta)) \right)^2 = -E_{\theta} \left(\frac{\partial^2}{\partial \theta^2} \ln(p(X_1, \theta)) \right)$ jest informacją w sensie Fishera pojedynczej (np. pierwszej) obserwacji. Gdy parametr $\theta = (\theta_1, \dots, \theta_k)$ jest parametrem wektorowym, to

$$\sqrt{n} \left(\begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_k \end{bmatrix} - \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_k \end{bmatrix} \right) \sim as N_k(\mathbf{0}, \mathbf{V}_{as}), \text{ gdzie } \mathbf{V}_{as} = \mathbf{i}^{-1}(\theta) \text{ a}$$

$$\mathbf{i}(\theta) = \left[E_{\theta} \left\{ \left(\frac{\partial}{\partial \theta_i} \ln p(X, \theta_1, \dots, \theta_k) \right) \left(\frac{\partial}{\partial \theta_j} \ln p(X, \theta_1, \dots, \theta_k) \right) \right\} \right]_{(k,k)} = - \left[E_{\theta} \frac{\partial^2}{\partial \theta_j \partial \theta_i} \ln p(X, \theta_1, \dots, \theta_k) \right]_{(k,k)} \text{ jest}$$

macierzą informacji w sensie Fishera.

Szkic dowodu asymptotycznej normalności. W regularnych przypadkach ENW otrzymuje się jako rozwiązanie równania wiarygodności $\frac{\partial}{\partial \theta} l(\theta, X) = 0$. Korzystając z mocnej zgodności dla dostatecznie dużych n estymator przyjmuje (z prawdopodobieństwem bliskim 1) wartości w pobliżu (nieznanej) wartości θ . Wobec tego rozwijając w szereg Taylora (war. regularności) i pomijając wyrazy wyższego rzędu mamy $0 = \frac{\partial}{\partial \theta} l(\hat{\theta}, X) \approx \frac{\partial}{\partial \theta} l(\theta, X) + \frac{\partial^2}{\partial \theta^2} l(\theta, X) (\hat{\theta} - \theta)$.

$$\text{Stąd } \hat{\theta} - \theta = - \frac{\frac{\partial}{\partial \theta} l(\theta, X)}{\frac{\partial^2}{\partial \theta^2} l(\theta, X)} \text{ i w konsekwencji } \sqrt{n}(\hat{\theta} - \theta) = \frac{\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l(\theta, X)}{-\frac{1}{n} \frac{\partial^2}{\partial \theta^2} l(\theta, X)}$$

Zauważmy, że $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l(\theta, X) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \ln p(X_i, \theta)$, przy czym każda ze zmiennych losowych w powyższej sumie ma wartość oczekiwaną 0 i wariancję $i(\theta)$ (informacja Fishera w pojedynczej obserwacji). Z CTG otrzymujemy, że $\frac{1}{\sqrt{n}} \frac{\partial}{\partial \theta} l(\theta, X) \Rightarrow N(0, i(\theta))$.

Z kolei z MPWL $-\frac{1}{n} \frac{\partial^2}{\partial \theta^2} l(\theta, X) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \ln p(X_i, \theta) \xrightarrow{zP_{\theta}1} -E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \ln p(X_1, \theta) \right] = i(\theta)$. Z twierdzenia Slutskiego otrzymujemy że $\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, i^{-1}(\theta))$.

Przykład Skonstruować estymator największej wiarygodności parametru θ i asymptotyczny przedział ufności na poziomie $1-\alpha=0,95$ oparty na n niezależnych obserwacjach $X_1, X_2, \dots, X_n$ z rozkładu o gęstości $p(x, \theta) = \theta^2 x \exp(-\theta x)$, $\theta > 0, x > 0$.

$$L(\theta; X_1, X_2, \dots, X_n) = \theta^{2n} \left(\prod_{i=1}^n X_i \right) \exp(-\theta \sum_{i=1}^n X_i)$$

$$l(\theta; X_1, X_2, \dots, X_n) = \ln L(\theta; X_1, X_2, \dots, X_n) = 2n \ln \theta + \sum_{i=1}^n \ln X_i - \theta \sum_{i=1}^n X_i$$

$$WK: \frac{2n}{\theta} - \sum_{i=1}^n X_i = 0 \text{ i } WW \Rightarrow \hat{\theta} = \frac{2}{\bar{X}} \text{ ENW}[\theta] \quad , \quad i_{\theta} = \frac{2}{\theta^2} \Rightarrow \sigma_{as}^2 = \frac{\theta^2}{2}$$

Korzystając z faktu, że asymptotyczny rozkład statystyki $\sqrt{n} \frac{\hat{\theta} - \theta}{\sigma_{as}}$ jest rozkładem $N(0,1)$ możemy z

tablic tego rozkładu odczytać dla danego α wartość $u_{1-\alpha/2}$ taką, że

$$P\left(\left|\sqrt{n} \frac{\hat{\theta} - \theta}{\sigma_{as}}\right| < u_{1-\alpha/2}\right) = 1 - \alpha \quad \text{a stąd otrzymujemy}$$

$$P\left(\hat{\theta} - \frac{\sigma_{as} u_{1-\alpha/2}}{\sqrt{n}} < \theta < \hat{\theta} + \frac{\sigma_{as} u_{1-\alpha/2}}{\sqrt{n}}\right) = 1 - \alpha.$$

Po podstawieniu za σ_{as}^2 przybliżonej wartości $\frac{\hat{\theta}^2}{2}$ (wniosek z mocnej zbieżności ENW) otrzymujemy asymptotyczny przedział ufności na poziomie $1 - \alpha$ postaci:

$$P\left(\hat{\theta}\left(1 - \frac{u_{1-\alpha/2}}{\sqrt{2n}}\right) < \theta < \hat{\theta}\left(1 + \frac{u_{1-\alpha/2}}{\sqrt{2n}}\right)\right) \approx 1 - \alpha.$$

Przykład. Skonstruować estymator największej wiarygodności parametru θ oparty na n niezależnych obserwacjach $X_1, X_2, \dots, X_n$ z rozkładu jednostajnego $U[0, \theta]$.

W tym przypadku $L(\theta; X_1, X_2, \dots, X_n) = \begin{cases} \frac{1}{\theta^n}, & \forall i: X_i \leq \theta \\ 0, & \exists i: X_i > \theta \end{cases} = \begin{cases} \frac{1}{\theta^n}, & X_{\max} \leq \theta \\ 0, & X_{\max} > \theta \end{cases}$ skąd natychmiast otrzymujemy,

że $X_{\max} = \max(X_1, X_2, \dots, X_n)$ jest ENW parametru θ . Jest to przypadek, gdzie warunki regularności nie są spełnione i nie można wykorzystać własności asymptotycznej normalności ENW do konstrukcji przedziału ufności dla θ .

Ciekawy przykład estymacji NW dla danych cenzurowanych.

Niech $X_1, \dots, X_{100}$ będzie próba losową z rozkładu wykładniczego o nieznanej wartości oczekiwanej a (tzn. $f(x) = \frac{1}{a} e^{-\frac{x}{a}} \mathbf{1}_{[0, \infty)}(x)$). Estymujemy a na podstawie częściowej informacji o próbce, a mianowicie na podstawie tego, iż

- 80 zmiennych (spośród wszystkich 100 z próbki) przybrało wartości poniżej 3,
- średnia arytmetyczna z tych wartości wynosi 2. Znaleźć ENW[a].

Rozwiązanie. Niech Y będzie zmienną losową zdefiniowaną wzorem

$$Y = X \cdot \mathbf{1}_{[0,3)}(X) + b \cdot \mathbf{1}_{[3,\infty)}(X) = \begin{cases} X, & \text{gdy } X < 3 \\ b, & \text{gdy } X \geq 3 \end{cases}, \text{ gdzie liczba } b > 3 \text{ jest etykietą zdarzenia } X \geq 3.$$

Zmienna Y ma dystrybuantę

$$F_Y(y) = \begin{cases} 1 - e^{-\frac{y}{a}}, & y \leq 3 \\ 1 - e^{-\frac{3}{a}}, & 3 < y \leq b \\ 1, & y > b \end{cases}.$$

Niech $\mu = \lambda + \delta_b$ gdzie λ oznacza miarę Lebesgue'a i $\forall A \in \mathcal{B}$ mamy

$$P(A) = \int_A \frac{1}{a} e^{-\frac{y}{a}} \mathbf{1}_{[0,3)} d\lambda(y) + e^{-\frac{3}{a}} \delta_b(A) = \int_A \left(\frac{1}{a} e^{-\frac{y}{a}} \mathbf{1}_{[0,3)} + e^{-\frac{3}{a}} \mathbf{1}_{\{b\}} \right) d\mu.$$

Stąd $f_Y(y) = \frac{1}{a} e^{-\frac{y}{a}} \mathbf{1}_{[0,3)}(y) + e^{-\frac{3}{a}} \mathbf{1}_{\{b\}}(y)$ jest gęstością rozkładu zmiennej losowej Y względem μ .

Wobec tego funkcja wiarygodności jest postaci

$$L(a; Y_1, \dots, Y_{100}) = \prod_{i=1}^{100} \left(\frac{1}{a} e^{-\frac{Y_i}{a}} \mathbf{1}_{[0,3)}(Y_i) + e^{-\frac{3}{a}} \mathbf{1}_{\{b\}}(Y_i) \right) = \frac{1}{a^{80}} e^{-\frac{1}{a} \sum_{i=1}^{100} \mathbf{1}_{[0,3)}(Y_i)} e^{-\frac{3}{a} 20} = \frac{1}{a^{80}} e^{-\frac{160}{a}} e^{-\frac{60}{a}}$$

$$l(a; Y_1, \dots, Y_{100}) = \ln L(a; Y_1, \dots, Y_{100}) = -80 \ln a - \frac{220}{a}$$

$$\text{WK: } \frac{dl}{da} = -\frac{80}{a} + \frac{220}{a^2} = 0 \Rightarrow \hat{a} = \frac{11}{4} = \text{ENW}[a] \quad (\text{WW oczywisty})$$

Przykład. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu normalnego $N(m, \sigma^2)$ o funkcji

gęstości $p(x; m, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$. Funkcja wiarygodności jest postaci

$$L(m, \sigma^2; \mathbf{X}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - m)^2} \quad \text{a jej logarytm } l(m, \sigma^2; \mathbf{X}) = -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - m)^2.$$

$$\text{WK istnienia ekstremum } \begin{cases} \frac{\partial l}{\partial m} = 0 \\ \frac{\partial l}{\partial \sigma} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^n (X_i - m) = 0 \\ -\frac{n}{\sigma} + \frac{2}{2\sigma^3} \sum_{i=1}^n (X_i - m)^2 = 0 \end{cases}, \text{ stąd } \begin{cases} m = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \\ \sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \end{cases}. \text{ Pokażemy,}$$

że $(\hat{m}, \hat{\sigma}^2) = (\bar{X}, \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2) = \text{ENW}[(m, \sigma^2)]$, wykazując, że $l(\hat{m}, \hat{\sigma}^2; \mathbf{X}) - l(m, \sigma^2; \mathbf{X}) \geq 0 \quad \forall (m, \sigma^2)$ i równość zachodzi tylko dla $m = \hat{m}$ i $\sigma^2 = \hat{\sigma}^2$.

Z własności niezmienniczości ENW widać, że $(\hat{m}, \hat{\sigma}) = (\bar{X}, \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}) = \text{ENW}[(m, \sigma)]$.

$$l(\hat{m}, \hat{\sigma}^2; \mathbf{X}) - l(m, \sigma^2; \mathbf{X}) = -\frac{n}{2} \ln \hat{\sigma}^2 - \frac{n}{2} + \frac{n}{2} \ln \sigma^2 + \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - m)^2 \stackrel{(1)}{=} -\frac{n}{2} \left(\ln \frac{\hat{\sigma}^2}{\sigma^2} + 1 - \frac{\hat{\sigma}^2}{\sigma^2} \right) + \frac{n}{2\sigma^2} (\bar{X} - m)^2 \geq 0$$

Ostatnie wyrażenie jest sumą dwóch nieujemnych składników (bo $\ln x + 1 - x \leq 0$ i równość ma miejsce tylko dla $x = 1$) które jednocześnie się zerują tylko dla $m = \hat{m}$ i $\sigma^2 = \hat{\sigma}^2$.

$$\text{Uwaga: } \sum_{i=1}^n (X_i - m)^2 \stackrel{(1)}{=} \sum_{i=1}^n (X_i - \bar{X} + \bar{X} - m)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - m)^2 = n\hat{\sigma}^2 + n(\bar{X} - m)^2.$$

$$\text{Macierz informacji } \mathbf{i}_{(m, \sigma)} = -E_{(m, \sigma)} \begin{bmatrix} \frac{\partial^2}{\partial m^2} \ln p(X, m, \sigma) & \frac{\partial^2}{\partial \sigma \partial m} \ln p(X, m, \sigma) \\ \frac{\partial^2}{\partial m \partial \sigma} \ln p(X, m, \sigma) & \frac{\partial^2}{\partial \sigma^2} \ln p(X, m, \sigma) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}. \text{ Stąd}$$

$$\sqrt{n} \left(\begin{bmatrix} \hat{m} \\ \hat{\sigma} \end{bmatrix} - \begin{bmatrix} m \\ \sigma \end{bmatrix} \right) \sim as N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{bmatrix} \right) \quad \sqrt{n} \left(\begin{bmatrix} \hat{m} \\ \hat{\sigma}^2 \end{bmatrix} - \begin{bmatrix} m \\ \sigma^2 \end{bmatrix} \right) \sim as N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \right).$$

Przykład. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Poissona $\mathcal{P}(\lambda)$ o funkcji gęstości

$p(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, $x = 0, 1, \dots$ względem miary liczącej $\mu(\cdot) = \sum_{k=0}^{\infty} \delta_k(\cdot)$. Funkcja wiarygodności jest postaci

$$L(\lambda; \mathbf{X}) = \frac{\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!} e^{-n\lambda} \text{ a jej logarytm } l(\lambda; \mathbf{X}) = \sum_{i=1}^n X_i \ln \lambda - n\lambda - \ln \prod_{i=1}^n X_i!.$$

WK : $\frac{\partial}{\partial \lambda} l(\lambda; \mathbf{X}) = 0 \Leftrightarrow \frac{\sum_{i=1}^n X_i}{\lambda} - n = 0 \Rightarrow \lambda = \bar{X}$. Badając monotoniczność funkcji $l(\lambda; \mathbf{X})$ stwierdzamy, że

$\hat{\lambda} = \bar{X} = ENW[\lambda]$. Widać, że $\frac{\partial^2}{\partial \lambda^2} \ln p(X, \lambda) = -\frac{X}{\lambda^2}$, stąd $i_\lambda = -E_\lambda(-\frac{X}{\lambda^2}) = \frac{1}{\lambda}$, więc $\sqrt{n}(\bar{X} - \lambda) \sim as N(0, \lambda)$.

Zadania

- Niech $X_1, \dots, X_n$ będzie próbą prostą z rozkładu równomiernego (dyskretnego jednostajnego) na $\{1, \dots, k\}$ i $k \in \mathbb{N}$. Wyznaczyć estymator największej wiarygodności parametru k . Wykazać, że estymator ten jest zgodny. Odp. $\hat{k} = X_{(n)}$
- Niech $X_1, \dots, X_n$ będzie próbą prostą z rozkładu jednostajnego $U[\theta_1, \theta_2]$ na przedziale na $[\theta_1, \theta_2]$. Wyznaczyć estymator największej wiarygodności parametru (θ_1, θ_2) . (Odp. $(\hat{\theta}_1, \hat{\theta}_2) = (X_{(1)}, X_{(n)})$).
- Czas pracy elementu jest zmienną losową X o gęstości $f(x) = abx^{a-1} \exp(-bx^a) \mathbf{1}_{(0, \infty)}(x)$, gdzie a jest znanym dodatnim parametrem zaś b jest nieznaną dodatnią stałą (rozkład Weibulla). Wyznaczyć estymator największej wiarygodności parametru b oparty na n elementowej próbie prostej. Wyznaczyć asymptotyczny przedział ufności dla parametru b na poziomie $1-\alpha$.
- Zmienna losowa N ma rozkład Poissona z parametrem intensywności λ , który chcemy oszacować. Niestety możemy obserwować jedynie zmienną losową M , która przyjmuje wartość 0 jeśli N jest równa 0, a wartość 1 jeśli N jest większa od 0. Wyznaczyć estymator największej wiarygodności parametru λ i asymptotyczny przedział ufności dla λ na poziomie $1-\alpha$. (Odp. $\hat{\lambda} = -\ln(1 - \bar{M})$)
- Skonstruować estymator największej wiarygodności parametru θ i asymptotyczny przedział ufności na poziomie $1-\alpha=0.95$ oparty na n elementowej próbie prostej $X_1, X_2, \dots, X_n$ z rozkładu
 - geometrycznego $p(x, \theta) = \theta(1-\theta)^{x-1}$, $\theta \in (0, 1)$, $x=1, 2, \dots$,
 - geometrycznego $p(x, \theta) = (1-\theta)\theta^{x-1}$, $\theta \in (0, 1)$, $x=1, 2, \dots$,
 - wykładniczego $p(x, \theta) = \theta \exp(-\theta x)$, $\theta > 0$, $x > 0$,
 - wykładniczego $p(x, \theta) = \theta^{-1} \exp(-x/\theta)$, $\theta > 0$, $x > 0$.
 - o gęstości $p(x, \theta) = \theta^2 x \exp(-\theta x)$, $\theta > 0$, $x > 0$,
 - normalnego $p(x, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-1)^2}{2\theta^2}}$, $\theta > 0$.
 - normalnego $p(x, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\theta)^2}{2\theta}}$, $\theta > 0$.
 - Poissona $p(x, \theta) = \frac{\theta^x}{x!} e^{-\theta}$, $\theta > 0$, $x=0, 1, 2, \dots$.
 - z rozkładu Bernoulliego $p(x, \theta) = \theta^x (1-\theta)^{1-x}$, $\theta \in (0, 1)$, $x=0, 1$.
 - z rozkładu Pareto (ozn. $Pa(1, a)$) o funkcji gęstości $f(x) = a x^{-a-1} \mathbf{1}_{(1, \infty)}(x)$, gdzie $a > 1$
- Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu o gęstości zadanej wzorem $f(x, \theta) = \theta^{-1} x^{\theta-1} \mathbf{1}_{(0, 1)}(x)$. Wyznaczyć estymator największej wiarygodności parametru θ i wyznaczyć błąd średniokwadratowy (ryzyko) tego estymatora. (wskazówka - aby wyznaczyć ryzyko warto wyznaczyć rozkład zmiennej $Y=-\ln X$)
- Skonstruować estymator największej wiarygodności parametru θ i asymptotyczny przedział ufności na poziomie $1-\alpha=0.95$ oparty na n elementowej próbie prostej $X_1, X_2, \dots, X_n$ z rozkładu Laplace'a o gęstości $p(x, \theta) = \frac{\theta}{2} e^{-\theta|x|}$, $\theta > 0$.