1 Przedziały ufności

W tym rozdziale będziemy zajmować się przede wszystkim zadaniami związanymi z przedziałami ufności. Będą nas również interesować statystki pozycyjne oraz estymatory największej wiarygodności.

1. (Eg 48/10) Niech $X_1, X_2, ..., X_n$, $n > 5$ będą niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na przedziale $(0, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Wyznaczamy przedział ufności dla parametru θ postaci
$$[2X_{3,n}, 2X_{n-2,n}],$$
gdzie $X_{k,n}$ oznacza k-tą statystykę pozycyjną z próby $X_1, X_2, ..., X_n$. Dla jakiej najmniejszej liczebności próby losowej n zachodzi
$$\mathbf{P}_\theta(\theta \in [2X_{2,n}, 2X_{n-2,n}]) \geqslant 0,9.$$
Odp: D-> 11.

Rozwiązanie. Należy pamiętać, że dla ustalonej wartości t zdarzenie $X_{k,n} \leqslant t$ jest osiągnięciem co najmniej k sukcesów w doświadczeniu Bernoulliego z prawdopodobieństwem sukcesu $F(t) = t/\theta$ dla $0 < t < \theta$, gdzie F jest dystrybuantą rozkładu jednostajnego na $(0, \theta)$. Zatem z prawdopodobieństwem sukcesu $F(t) = t/\theta$ dla $0 < t < \theta$, gdzie F jest dystrybuantą rozkładu jednostajnego na $(0, \theta)$. Zatem
$$\mathbf{P}_\theta(X_{2,n} \leqslant \frac{\theta}{2} \leqslant X_{n-2,n}) = \mathbf{P}(2 \leqslant S \leqslant n-2),$$
gdzie S pochodzi z rozkładu $\mathcal{B}(n, \frac{1}{2})$. Obliczamy
$$\mathbf{P}(2 \leqslant S \leqslant n-2) = 1 - \mathbf{P}(S \leqslant 2) - \mathbf{P}(S \geqslant n-2) = 1 - 2(\binom{n}{2}2^{-n} + n2^{-n} + 2^{-n}) = 1 - (n^2+n+2)2^{-n}.$$
Dla $n = 11$ otrzymujemy po raz pierwszy wartość $(n^2 + n + 2)2^{-n}$ mniejszą od $0,1$.

■

333333

2. (Eg 49/10) Niech $X_1, X_2, ..., X_{10}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie o gęstości
$$f_\theta(x) = \theta x^{\theta-1} 1_{x \in (0,1)},$$
gdzie $\theta > 0$ jest nieznanym parametrem. Wyznaczamy przedział ufności dla parametru θ postaci
$$[c\hat{\theta}, d\hat{\theta}],$$
gdzie $\bar{\theta} = \bar{\theta}(X_1, X_2..., X_{10})$ jest estymatorem największej wiarogodności, a stałe c i d są dobrane tak, aby $\mathbf{P}_\theta(\theta < c\hat{\theta}) = \mathbf{P}_\theta(\theta > d\hat{\theta}) = 0,05$. Wyznaczyć c i d. *Odp:* A-> $c = 0,54$ i $d = 1,57$.

Rozwiązanie. Obliczamy wiarygodność
$$L(\theta, x) = \theta^{10} \prod_{i=1}^{10} x_i^{\theta-1} 1_{x_i \in (0,1)}.$$
Zatem dla ustalonego x funkcja $f(\theta) = \ln L(\theta, x)$ osiąga maksimum dla θ spełniającego $f'(\theta) = 0$, czyli
$$10\frac{1}{\theta} + \sum_{i=1}^{10} \ln x_i = 0.$$
Stąd
$$\bar{\theta} = -\frac{10}{\sum_{i=1}^{10} \ln x_i}.$$

Pozostaje znaleźć c i d takie, że

$$\mathbf{P}_\theta(\theta < c\bar{\theta}) = P_\theta(\theta > d\bar{\theta}) = 0,05.$$

Nietrudno sprawdzić, że $-\ln X_i$ ma rozkład wykładniczy z parametrem θ. Stąd $-\sum_{i=1}^{10} \ln X_i$ ma rozkład $Gamma(10, \theta)$. Zatem $\theta/\bar{\theta}$ ma rozkład $Gamma(10, 10)$, który dalej jest tym samym co $\frac{1}{20}Gamma(\frac{20}{2}, \frac{1}{2})$ czyli $\frac{1}{20}\chi^2(20)$. To oznacza, że $20c$ będzie dolnym a $20d$ górnym kwantylem dla rozkładu $\chi^2(20)$ dla wartości $0,05$. Korzystając z tablica znajdujemy $c \simeq 0,54$, $d \simeq 1,57$.

■

3. (Eg 50/5) Niech $X_1, ..., X_n$, $n > 1$ będzie próbką z rozkładu jednostajnego o gęstości danej wzorem:

$$f_\theta(x) = \frac{1}{\theta}1_{x\in(0,\theta)},$$

gdzie θ jest nieznanym parametrem. Zmienne losowe $X_1, ..., X_n$ nie są w pełni obserwowalne. Obserwujemy zmienne losowe $Y_i = \min(X_i, M)$, gdzie M jest ustaloną liczbą dodatnią. Oblicz estymator największej wiarogodności $\hat{\theta}$ parametru θ jeśli wiadomo, że w próbce $Y_1, ..., Y_n$, jest K obserwacji o wartościach mniejszych niż M i $K \in \{1, ..., n-1\}$ *Odp:* B-> Mn/K.

Rozwiązanie. Z danych zadania dostajemy, że $\theta > M$. Szansa, że dokładnie K zmiennych będzie mniejszych niż M wynosi

$$\binom{n}{K}(\frac{M}{\theta})^K(\frac{\theta - M}{\theta})^{n-K}.$$

Warto zauważyć, że przy okazji jest to funkcja wiarygodności. Poszukujemy wartości θ dla której funkcja $f(\theta) = \ln L(\theta, x)$ przyjmuje maksimum czyli punktu θ takiego, że $f'(\theta) = 0$. Zachodzi

$$-n\frac{1}{\theta} + (n-K)\frac{1}{\theta - M} = 0.$$

Zatem $\bar{\theta} = Mn/K$.

■

4. (Eg 51/2) Załóżmy, że $X_1, X_2, ..., X_n$, $n > 2$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie wykładniczym. Niech $S = \sum_{i=1}^n X_i$. Oblicz

$$p = \mathbf{P}(X_1 \leqslant S/2, ..., X_n \leqslant S/2).$$

Odp: D-> $1 - \frac{n}{2^{n-1}}$.

Rozwiązanie. Kluczowe to aby zauważyć, że zdarzeniem przeciwnym do $X_1 \leqslant S/2, ..., X_n \leqslant S/2$ jest $\bigcup_{i=1}^n \{X_i > S/2\}$. Zatem

$$\mathbf{P}(X_1 \leqslant S/2, ..., X_n \leqslant S/2) = 1 - n\mathbf{P}(X_1 > S/2).$$

Rozkład X_1/S jest postaci $Beta(1, n-1)$. Stąd

$$\mathbf{P}(X_1 \leqslant S/2) = \int_{\frac{1}{2}}^1 (n-1)(1-y)^{n-2}dy = \frac{1}{2^{n-1}}.$$

Stąd $p = 1 - n2^{-n+1}$.

■

5. (Eg 52/9) Załóżmy, że $X_1, X_2, ..., X_n, ...$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie wykładniczym o gęstości

$$f(x) = \frac{1}{\mu}\exp(-x/\mu) \text{ dla } x > 0.$$

Zmienna losowa N jest niezależna od $X_1, X_2, ..., X_n, ...$ i ma rozkład geometryczny dany wzorem:

$$\mathbf{P}(N = n) = p(1-p)^n \text{ dla } n = 0, 1, 2, ...$$

Niech $S_N = \sum_{i=1}^{N}$ (przy tym $S_0 = 0$, zgodnie z konwencją). Oblicz $\mathbf{E}(N|S_N = s)$, dla $s > 0$.
Odp: D-> $s(1-p)/\mu + 1$.

Rozwiązanie. Stosujemy ogólny wzór Bayesa. Najpierw wyznaczamy gęstość (N, S) względem $\mu \otimes \lambda$, gdzie μ jest miarą liczącą na $\mathbb{Z}$, a λ miarą Lebesgue'a na $\mathbb{R}$. Przypomnijmy, że S_n ma rozkład Gamma$(n, \frac{1}{\mu})$

$$f_{N,S_N}(n, s) = p(1-p)^n \frac{s^{n-1}}{(n-1)!\mu^n} e^{-\frac{s}{\mu}} 1_{s>0} 1_{N=0}.$$

Rozkład (N, S_N) ma też atom w $(0, 0)$ który osiąga z prawdopodobieństwem p. Obliczamy gęstość warunkową dla $s > 0$

$$f_{N,S_N}(n|s) = \frac{[(1-p)s]^{n-1}}{(n-1)!\mu^{n-1}} e^{\frac{(1-p)s}{\mu}} 1_{n>0}.$$

Zatem

$$\mathbf{E}(N|S_n = s) = \sum_{n=1}^{\infty} \frac{n[(1-p)s]^{n-1}}{(n-1)!\mu^{n-1}} e^{-\frac{(1-p)s}{\mu}} = 1 + \frac{s(1-p)}{\mu}.$$

■

6. (Eg 53/6) Rozważmy zmienne losowe N, X, Y. Wiadomo, że rozkład warunkowy zmiennej losowej N, gdy $X = x$ i $Y = y$ jest rozkładem Poissona o wartości oczekiwanej x. Rozkład warunkowy zmiennej losowej X, gdy $Y = y$ jest rozkładem Gamma$(2, y)$, a rozkład zmiennej Y jest rozkładem Gamma$(4, 3)$, gdzie rozkład Gamma(α, β) ma gęstość

$$p_{\alpha,\beta} = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} 1_{x>0}.$$

Wtedy wariancja $\mathbf{Var}N$ jest równa
Odp: B-> 7.

Rozwiązanie. Obliczamy

$$\mathbf{Var}N = \mathbf{E}N^2 - (\mathbf{E}N)^2 = \mathbf{EE}(N^2|X, Y) - (\mathbf{EE}(N, X, Y))^2 = \mathbf{E}(X^2 + X) - (\mathbf{E}X)^2 =$$
$$= \mathbf{EE}(X^2 + X|Y) - (\mathbf{EE}(X|Y))^2 = \mathbf{E}\frac{6}{Y^2} + \frac{2}{Y} - (\mathbf{E}\frac{2}{Y})^2.$$

Dla zmiennej Z z rozkładu Gamma(α, β) jeśli $\alpha > 2$, to $EZ^{-2} = \frac{\beta^2}{(\alpha-1)(\alpha-2)}$ oraz jeśli $\alpha > 1$, to $\mathbf{E}Z^{-1} = \frac{\beta}{\alpha-1}$. Zatem $\mathbf{E}Y^{-2} = \frac{3}{2}$, $\mathbf{E}Y^{-1} = \frac{3}{3} = 1$. Czyli

$$\mathbf{Var}N = 9 + 2 - 4 = 7.$$

■

7. (Eg 54/8) Niech $X_1, X_2, ..., X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie o gęstości

$$p_{a,b} = be^{-b(x-a)} 1_{x \geqslant a}$$

gdzie $a \in \mathbb{R}$ i $b > 0$ są nieznanymi parametrami. Rozważmy estymator największej wiarygodności (T_a, T_b) wektora parametrów (a, b). Wartości oczekiwane ET_a i $\mathbf{E}T_b$ są równe
Odp: D-> $\mathbf{E}T_a = a + \frac{1}{nb}$, $\mathbf{E}T_b = \frac{n}{n-2}b$.

Rozwiązanie. Obliczamy wiarygodność

$$L((a,b),x) = b^n \exp(-b\sum_{i=1}^{n}(x_i - a))1_{\min(x_i)\geqslant a}.$$

Szukamy punktu (a,b) maksymalizującego wiarygodność. Oczywiście $T_a = \min(x_1,...,x_n)$ natomiast b wyznaczamy korzystając z funkcji $f(b) = \ln L((T_a,b),x)$, rozwiązując równanie $f'(b) = 0$ czyli

$$\frac{n}{b} = \sum_{i=1}^{n}(x_i - T_a), \quad \text{zatem } T_b = \frac{n}{\sum_{i=1}^{n}(x_i - T_a)}.$$

Obliczamy

$$\mathbf{E}T_a - a = \int_0^\infty \mathbf{P}(X_i - a > t)^n dt = \int_0^\infty e^{-nbx}dx = \frac{1}{nb},$$

czyli $\mathbf{E}T_a = a + \frac{1}{nb}$. Z powodu braku pamięci rozkładu wykładniczego rozkład $\sum_{i=1}^{n}(x_i - T_a)$ będzie rozkładem $(n-1)$ zmiennych niezależnych o rozkładzie wykładniczym z parametrem b czyli $Gamma(n-1,b)$. Obliczamy

$$\mathbf{E}T_b = n\int_0^\infty \frac{b^{n-1}}{(n-2)!}x^{n-2}e^{-bx}dx = \frac{n}{n-2}b.$$

■

8. (Eg 55/6) Niech N oraz $X_1, X_2, ...$ będą niezależnymi zmiennymi losowymi, przy czym N ma rozkład Poissona z wartością oczekiwaną $\lambda = 1$, zaś rozkład każdej ze zmiennych X_n podaje następująca tabelka:

$$\begin{bmatrix} x & 1 & 2 & 3 \\ \mathbf{P}(X_n = x) & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

Niech $S = \sum_{i=1}^{N} X_i$ dla $N > 0$ i $S = 0$ dla $N = 0$. Oblicz warunkową wartość oczekiwaną $\mathbf{E}(N|S = 3)$.
Odp: A-> $\frac{27}{19}$.

Rozwiązanie. Mamy klasyczny wzór Bayesa

$$\mathbf{E}(N|S=3) = \sum_{n=0}^{\infty} n\mathbf{P}(N = n|S = 3) = \sum_{n=0}^{\infty} n\frac{\mathbf{P}(S = 3|N = n)\mathbf{P}(N = n)}{\sum_{m=0}^{\infty}\mathbf{P}(S = 3|N = m)\mathbf{P}(N = m)}.$$

Nadto

$$\sum_{n=0}^{\infty} n\mathbf{P}(S = 3|N = n)\mathbf{P}(N = n) = \mathbf{P}(X_1 = 3)\mathbf{P}(N = 1) + 2(\mathbf{P}(X_1 = 2, X_2 = 1)+$$
$$+ \mathbf{P}(X_1 = 1, X_2 = 2))\mathbf{P}(N = 2) + 2\mathbf{P}(X_1 = X_2 = X_3 = 1)P(N = 3) =$$
$$= e^{-1}(\frac{1}{4} + \frac{1}{4} + \frac{1}{16})$$

oraz

$$\sum_{m=0}^{\infty} \mathbf{P}(S = 3|N = m)\mathbf{P}(N = m) = \mathbf{P}(X_1 = 3)\mathbf{P}(N = 1)+$$
$$+ (\mathbf{P}(X_1 = 2, X_2 = 1) + \mathbf{P}(X_1 = 1, X_2 = 2))P(N = 2) + \mathbf{P}(X_1 = X_2 = X_3 = 1) =$$
$$= e^{-1}(\frac{1}{4} + \frac{1}{8} + \frac{1}{48})$$

Stąd $\mathbf{E}(N|S = 3) = \frac{27}{19}$.

■

9. (Eg 56/9) Niech $X_1, X_2, ..., X_n$ będą zmiennymi losowymi o rozkładzie $Pareto(1, a_1)$ a $Y_1, Y_2, ..., Y_m$ będą zmiennymi losowymi o rozkładzie $Pareto(1, a_2)$, gdzie $a_1, a_2 > 0$ są nieznanymi parametrami. Wszystkie zmienne są niezależne. Na poziomie ufności $1 - \alpha$ budujemy przedział ufności $[dT, cT]$ dla parametru $\frac{a_1}{a_2}$ na podstawie estymatora największej wiarogodności T tegoż parametru w ten sposób, że

$$\mathbf{P}_{a_1,a_2}(cT < \frac{a_1}{a_2}) = \mathbf{P}_{a_1,a_2}(dT > \frac{a_1}{a_2}) = \frac{\alpha}{2}.$$

Jeśli $\alpha = 0,1$ i $m = 4$ i $n = 5$, to przedział ufności ma długość
Odp: E-> $3,02T$.

Rozwiązanie. Jeśli $m = 4$ i $n = 5$, to funkcja wiarygodności ma postać

$$L((a_1, a_2), (x, y)) = \alpha_1^5 \alpha_2^4 \prod_{i=1}^{5} (x_i)^{-\alpha_1 - 1} 1_{x_i \geqslant 1} \prod_{j=1}^{4} (y_j)^{-\alpha_2 - 1} 1_{y_j \geqslant 1}.$$

Zatem estymatorem $ENW(\alpha_1, \alpha_2)$ jest punkt minimum funkcji $f(\alpha_1, \alpha_2) = \ln L((a_1, a_2), (x, y))$, czyli rozwiązanie równania $\frac{\partial f}{\partial \alpha_1} = \frac{\partial f}{\partial \alpha_2} = 0$. Obliczamy

$$\frac{5}{\alpha_1} = \sum_{i=1}^{5} \ln x_i, \quad \frac{4}{\alpha_2} = \sum_{i=1}^{4} \ln y_i.$$

Stąd $\hat{\alpha}_1 = \frac{5}{\sum_{i=1}^5 \ln X_i}$, $\hat{\alpha}_2 = \frac{4}{\sum_{i=1}^4 \ln Y_i}$. Zatem

$$T = \frac{\hat{\alpha}_1}{\hat{\alpha}_2} = \frac{5}{4} \frac{\sum_{i=1}^4 \ln Y_i}{\sum_{i=1}^5 \ln X_i}.$$

Oczywiście $\sum_{i=1}^5 \ln X_i$ ma rozkład $Gamma(5, \alpha_1)$, a $\sum_{i=1}^4 \ln Y_i$ rozkład $Gamma(4, \alpha_2)$ nadto te zmienne są niezależne. Stąd $2\alpha_1 \sum_{i=1}^5 \ln X_i$ ma rozkład $Gamma(\frac{10}{2}, \frac{1}{2}) = \chi^2(10)$, a $2\alpha_2 \sum_{i=1}^4 \ln Y_i$ ma rozkład $Gamma(\frac{8}{2}, \frac{1}{2}) = \chi^2(8)$. Zatem $\frac{a_1}{a_2} T$ ma rozkład Fishera-Snedecora $F(10, 8)$. Odczytujemy z tablic $c = 3,347$, $d = 0,326$. Stąd długość $[dT, cT]$ wynosi w przybliżeniu $3,021T$.

■

10. (Eg 57/7) Niech $X_1, X_2, ..., X_{10}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie o nieznanej medianie m. Budujemy przedział ufności dla parametru m postaci $[X_{3:10}, X_{7:10}]$, gdzie $X_{k:10}$ oznacza k-tą statystykę pozycyjną z próby $X_1, X_2, ..., X_{10}$. Prawdopodobieństwo $\mathbf{P}(m \in [X_{3:10}, X_{7:10}])$ jest równe
Odp: A-> $\frac{114}{128}$.

Rozwiązanie. Należy zauważyć, że dla ustalonego t szansa, że $X_{k:10} \leqslant t$ jest równe prawdopodobieństwu uzyskania co najmniej k sukcesów w doświadczeniu Bernoulliego z prawdopodobieństwem $F(t)$, gdzie F jest dystrybuantą wspólnego rozkładu zmiennych $X_1, ..., X_{10}$. Przyjmujemy, że rozkład jest ciągły (ściślej, że $F(m) = \frac{1}{2}$), w przeciwnym razie zadanie nie daje się rozwiązać. Wówczas przyjmując, że zmienna S ma rozkład $\mathcal{B}(10, \frac{1}{2})$ zachodzi równość

$$\mathbf{P}(m \in [X_{3:10}, X_{7:10}]) = \mathbf{P}(3 \leqslant S < 7) = 1 - \frac{2(1 + 10 + 45) + 120}{2^{10}} = 1 - \frac{2^3(14 + 15)}{2^9} = \frac{99}{128}.$$

■

11. (Eg 58/2) Niech $X_1, X_2, ..., X_n$ będą niezależnymi zmiennymi losowymi z rozkładu Pareto o gęstości

$$f_\theta(x) = \frac{\theta \lambda^\theta}{(\lambda + x)^{\theta + 1}} 1_{x > 0},$$

gdzie $\theta, \lambda > 0$ są nieznanymi parametrami. Rozważmy estymatory największej wiarygodności $\bar{\theta}_n, \bar{\lambda}_n$ parametrów θ i λ. Chcemy dobrać stałą t tak aby przy n dążącym do nieskończoności

prawdopodobieństwo zdarzenia $|\bar{\theta}_n - \theta|\sqrt{n} > t$ było równe $0,1$. Jeżeli $\theta = 3$ i $\lambda = 1$, to stała t jest równa
Odp: A -> $19,7$.

Rozwiązanie. W tym zadaniu najprościej skorzystać z twierdzeń o asymptotycznej zbieżności. Twierdzenia te bazują na analizie wiarygodności

$$L((\theta,\lambda),x) = \prod_{i=1}^{n} \frac{\theta\lambda^{\theta}}{(\lambda+x_i)^{\theta+1}}, \quad \ln L((\theta,\lambda),x) = n(\ln\theta + \theta\ln\lambda) - (\theta+1)\sum_{i=1}^{n}\ln(\lambda+x_i).$$

Aby zachodziła asymptotyczna normalność po pierwsze musi być spełniony warunek zgodności. To znaczy, że $(\bar{\theta}_n, \bar{\lambda}_n) \to (\theta,\lambda)$ przynajmniej według prawdopodobieństwa. Okazuje się, że przy różniczkowalnej funkcji wiarygodności wystarczy aby $\partial \ln L = 0$, czyli równanie

$$\frac{\partial \ln L((\theta,\lambda),x)}{\partial\theta} = 0, \quad \frac{\partial \ln L((\theta,\lambda),x)}{\partial\lambda} = 0$$

miało dokładnie jedno rozwiązanie dla dowolnego n. Sprawdzamy warunek

$$\frac{\partial \ln L((\theta,\lambda),x)}{\partial\theta} = n(\frac{1}{\theta} + \ln\lambda) - \sum_{i=1}^{n}\ln(\lambda+x_i) = 0,$$
$$\frac{\partial \ln L((\theta,\lambda),x)}{\partial\lambda} = n\frac{\theta}{\lambda} - (\theta+1)\sum_{i=1}^{n}\frac{1}{\lambda+x_i} = 0.$$

Zatem w rozwiązaniu tego równania parametr θ można wyrazić jako funkcję λ to znaczy

$$\theta = \frac{\sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}}{n - \sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}}.$$

Natomiast powstała po podstawieniu wyliczonego θ funkcja od λ czyli

$$n(\frac{n}{\sum_{i=1}^{n}\frac{\lambda}{\lambda+x_i}} - 1) + n\sum_{i=1}^{n}\ln\frac{\lambda}{\lambda+x_i}$$

jest malejąca i zmienia się od $+\infty$ dla $\lambda = 0$ do $-n$ dla $\lambda = \infty$ zatem ma tylko jedno miejsce zerowe.

Jeśli teraz skończone są również drugie pochodne funkcji $\ln L$ to rozkłady $(\bar{\lambda}_n, \bar{\theta}_n)$ mają asymptotycznie rozkład normalny $\mathcal{N}((\lambda,\theta), C_n)$, gdzie macierz kowariancji $C_n = -(\mathbf{E}\partial^2 \ln L)^{-1}$. Obliczamy zatem drugie pochodne

$$-\partial^2 \ln L(\theta,\lambda) = \begin{bmatrix} \frac{n}{\theta^2} & -\frac{n}{\lambda} + \sum_{i=1}^{n}\frac{1}{\lambda+x_i} \\ -\frac{n}{\lambda} + \sum_{i=1}^{n}\frac{1}{\lambda+x_i} & \frac{n\theta}{\lambda^2} - (\theta+1)\sum_{i=1}^{n}\frac{1}{(\lambda+x_i)^2} \end{bmatrix}.$$

Pozostaje obliczyć odpowiednie parametry dla $\theta = 3$ i $\lambda = 1$

$$\mathbf{E}\frac{n}{\theta^2} = \frac{n}{9},$$
$$\mathbf{E} - \frac{n}{\lambda} + \sum_{i=1}^{n}\frac{1}{\lambda+X_i} = -n + \frac{3}{4}n = -\frac{n}{4}, \quad \mathbf{E}\frac{n\theta}{\lambda^2} - (\theta+1)\sum_{i=1}^{n}\frac{1}{(\lambda+X_i)^2} = 3n - 4n\frac{3}{5} = \frac{3n}{5},$$

gdzie korzystamy z faktu $\mathbf{E}(\lambda+X_i)^k = \frac{3}{3+k}$, dla $k \geqslant 0$. Stąd wynika, że

$$C_n = \begin{bmatrix} \frac{n}{9} & -\frac{n}{4} \\ -\frac{n}{4} & \frac{3n}{5} \end{bmatrix}^{-1} = \frac{15\cdot 16}{n^2}\begin{bmatrix} \frac{3n}{5} & \frac{n}{4} \\ \frac{n}{4} & \frac{n}{9} \end{bmatrix}.$$

W szczególności oznacza to, że $\frac{(\bar{\theta}_n-3)\sqrt{n}}{12}$ zbiega według rozkładu do Z z rozkładu $\mathcal{N}(0,1)$. Obliczamy

$$\lim_{n\to\infty}\mathbf{P}(|\bar{\theta}_n-\theta|\sqrt{n}>t)=\lim_{n\to\infty}\mathbf{P}(\frac{|\bar{\theta}_n-3|\sqrt{n}}{12}>\frac{t}{12})=\mathbf{P}(|Z|>\frac{t}{12})=0,1.$$

Z tablic dostajemy $\frac{t}{12}\simeq 1,64$. Stąd $t\simeq 19,68$.

■

12. (Eg 59/10) Niech $X_1,X_2,...,X_{10}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie Pareto o gęstości

$$f_\theta(x)=\frac{2^\theta\theta}{(2+x)^{\theta+1}}1_{x>0},$$

gdzie $\theta>0$jest nieznanym parametrem. W oparciu o estymator największej wiarogodności T parametru θ zbudowano przedział ufności dla θ na poziomie ufności $0,95$ postaci $[cT,dT]$, gdzie liczby c i d dobrano tak, aby $\mathbf{P}_\theta(\theta<cT)=\mathbf{P}_\theta(\theta>dT)=0,025$. Liczby c i d są równe *Odp:* A-> $c=0,48$ i $d=1,71$.

Rozwiązanie. Obliczamy estymator $ENW(\theta)$, wiarygodność ma postać

$$L(\theta,x)=\prod_{i=1}^{10}\frac{2^\theta\theta}{(2+x_i)^{\theta+1}},$$

stąd

$$\ln L(\theta,x)=10\theta\ln 2+10\ln\theta-(\theta+1)\sum_{i=1}^{10}\ln(2+x_i).$$

Należy znaleźć rozwiązanie równania $\frac{\partial\ln L(\theta,x)}{\partial\theta}=0$, czyli

$$10\ln 2+\frac{10}{\theta}-\sum_{i=1}^{10}\ln(2+x_i)=0.$$

To znaczy

$$\theta=\frac{10}{\sum_{i=1}^{10}(\ln(2+x_i)-\ln 2)}.$$

Wyznaczmy rozkład $\ln(2+X_i)-\ln 2$. Dla $t>0$

$$\mathbf{P}(\ln(2+X_i)-\ln 2>t)=\mathbf{P}(X_i>2e^t-2)=2^\theta(2e^t)^{-\theta}=e^{-\theta t},$$

czyli rozkładem $\ln(2+X_i)-\ln 2$ jest rozkład wykładniczy z parametrem θ. Stąd $\sum_{i=1}^{10}(\ln(2+X_i)-\ln 2)$ ma rozkład $Gamma(10,\theta)$. Zatem $1/T$ ma rozkład $10^{-1}Gamma(10,\theta)=20^{-1}\theta^{-1}Gamma(\frac{20}{2},\frac{1}{2})$ czyli $20^{-1}\theta^{-1}\chi^2(20)$. Niech Z będzie z rozkładu $\chi^2(20)$, zachodzi

$$\mathbf{P}_\theta(\theta<cT)=\mathbf{P}(Z<20c)=0,025,\ \ \mathbf{P}_\theta(\theta>dT)=\mathbf{P}(Z>20d)=0,025.$$

Z tablic rozkładu χ^2 otrzymujemy $20c\simeq 9,591$, $20d\simeq 34,170$, czyli $c\simeq 0,48$, $d\simeq 1,71$.

■

13. (Eg 60/1) Niech $X_1,X_2,...,X_n$ będą niezależnymi zmiennymi losowymi z rozkładu o gęstości

$$f_\theta(x)=\frac{1}{\theta}x^{-\frac{1}{\theta}-1}1_{x>1},$$

gdzie $\theta\in(0,1)$ jest nieznanym parametrem. Rozważamy nieobciążony estymator parametru θ postaci $T_n=aY$, gdzie $Y=\min(\ln X_1,\ln X_2...,\ln X_n)$ i a jest odpowiednio dobraną stałą (być może zależną od liczebności próby n). Dla $\theta=\frac{1}{3}$ i $\varepsilon=\frac{1}{6}$ zachodzi
Odp: C-> $\mathbf{P}_\theta(|T_n-\theta|>\varepsilon)=1-e^{-\frac{1}{2}}+e^{\frac{3}{2}}$.

Rozwiązanie. Jak nietrudno stwierdzić $\ln X_i$ ma rozkład wykładniczy $Exp(\frac{1}{\theta})$. Dalej rozkład Y ma postać $Exp(\frac{n}{\theta})$ i w końcu aY ma rozkład $Exp(\frac{n}{a\theta})$. Stąd $a = n$, niech teraz Z będzie z rozkładu $Exp(1)$, zachodzi

$$\mathbf{P}_\theta(|T_n - \theta| > \varepsilon) = \mathbf{P}(|\theta Z - \theta| > \varepsilon) = \mathbf{P}(|Z - 1| > \frac{1}{2}).$$

Czyli

$$\lim_{n\to\infty} \mathbf{P}_\theta(|T_n - \theta| > \varepsilon) = 1 - \mathbf{P}(Z \in [\frac{1}{2}, \frac{3}{2}]) = 1 - e^{-\frac{1}{2}} + e^{-\frac{3}{2}}.$$

■

14. (Eg 61/7) Zmienna losowa X ma rozkład o gęstości

$$f_\theta(x) = \frac{\theta}{x^{\theta+1}} 1_{x>1},$$

gdzie $\theta > 0$ jest nieznanym parametrem. Nie obserwujemy zmiennej X ale zmienną Y równą X, gdy X jest większe od 2. Nie wiemy, ile było obserwacji zmiennej X nie większych niż 2 ani jakie były ich wartości. W wyniku tego eksperymentu otrzymujemy próbkę losową $Y_1, Y_2, ..., Y_8$ Na podstawie próbki budujemy przedział ufności dla parametru θ postaci $[c_1 T, c_2 T]$, gdzie T jest estymatorem największej wiarogodności parametru θ, a stałe c_1 i c_2 dobrane są tak, by

$$\mathbf{P}(\theta < c_1 T) = \mathbf{P}(\theta > c_2 T) = 0,05$$

Wtedy długość przedziału ufności jest równa
Odp: C-> $1,146T$.

Rozwiązanie. Interesuje nas rozkład warunkowy X pod warunkiem, że $X > 2$ który ma gęstość

$$g_\theta(x) = \frac{\theta 2^\theta}{x^{\theta+1}} 1_{x>2}.$$

Dalej zadanie jest standardowe, wyznaczamy estymator $ENW(\theta)$ z równania $\frac{\partial \ln L(x,\theta)}{\partial\theta} = 0$, czyli

$$8\ln 2 + \frac{8}{\theta} - \sum_{i=1}^{8} \ln x_i = 0.$$

Stąd

$$T = \frac{8}{\sum_{i=1}^{8}(\ln X_i - \ln 2)}.$$

Dalej zauważamy, że $\ln X_i - \ln 2$ ma rozkład $Exp(\theta)$. Stąd $\sum_{i=1}^{8}(\ln X_i - \ln 2)$ ma rozkład $Gamma(8, \theta)$ i wreszcie T^{-1} ma rozkład $16^{-1}\theta^{-1}Gamma(\frac{16}{2}, \frac{1}{2}) = 16^{-1}\theta^{-1}\chi^2(16)$. To pozwala obliczyć stałe c_1 i c_2 z tablic rozkładu χ^2. Istotnie niech Z będzie z rozkładu $\chi^2(16)$, zachodzi

$$\mathbf{P}(\theta < c_1 T) = \mathbf{P}(Z < 16c_1) = 0,05, \quad \mathbf{P}(\theta > c_2 T) = \mathbf{P}(Z > 16c_2).$$

Zatem $16c_1 \simeq 7,962$, $16c_2 \simeq 26,296$, czyli $c_1 \simeq 0,498$, $c_2 = 1,644$. Długość przedziału ufności wynosi w przybliżeniu $1,146T$.

■

15. (Eg 62/7) Niech $X_1, ..., X_{10}, ..., X_{30}$ będzie próbką losową z rozkładu normalnego $\mathcal{N}(\mu, \sigma^2)$, z nieznanymi parametrami μ i σ^2. Niech

$$\bar{X}_{10} = \frac{1}{10}\sum_{i=1}^{10} X_i, \quad \bar{X}_{30} = \frac{1}{30}\sum_{i=1}^{30} X_i,$$

$$S^2 = S_{10}^2 = \frac{1}{9}\sum_{i=1}^{10}(X_i - \bar{X}_{10})^2.$$

Skonstruowano przedział $[\bar{X}_{10} - aS, \bar{X}_{10} + aS]$ taki, że

$$\mathbf{P}(\bar{X}_{30} \in [\bar{X}_{10} - aS, \bar{X}_{10} + aS]) = 0,95.$$

Liczba a jest równa
Odp: E-> $0,584$.

Rozwiązanie. W tym zadaniu istotne jest aby umiejętnie przejść do rozkładu t-Studenta. Mamy niezależność $\bar{X}_{10}$, $\bar{X}_{20} = \frac{1}{20}\sum_{i=11}^{30} X_i$ oraz S. Nadto $\bar{X}_{30} = \frac{1}{3}\bar{X}_{10} + \frac{2}{3}\bar{X}_{20}$. Stąd

$$\mathbf{P}(\bar{X}_{30} \in [\bar{X}_{10} - aS, \bar{X}_{10} + aS]) = \mathbf{P}(|\bar{X}_{20} - \bar{X}_{10}| < \frac{3}{2}aS)$$

Teraz $\bar{X}_{20} - \bar{X}_{10}$ ma rozkład $\mathcal{N}(0, \frac{3\sigma^2}{20})$. Natomiast $\sum_{i=1}^{10}(X_i - \bar{X}_{10}$ ma rozkład $\sigma^2\chi^2(n-1)$. Stąd

$$Z = \frac{\frac{\sqrt{20}}{\sqrt{3}}(\bar{X}_{20} - \bar{X}_{10})}{S}, \quad \text{ma rozkład t-Studenta}$$

a 9 stopniami swobody. Zatem

$$\mathbf{P}(\bar{X}_{30} \in [\bar{X}_{10} - aS, \bar{X}_{10} + aS]) = \mathbf{P}(|Z| > \sqrt{15}a) = 0,95.$$

Zatem $\sqrt{15}a = 2,262$, czyli $a \simeq 0,584$

■

16. (Eg 63/9) Niech $X_1, X_2, ..., X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie Laplace'a o gęstości

$$f_\theta(x) = \exp(-2|x - \theta|) \ \text{dla} \ x \in \mathbb{R},$$

gdzie θ jest nieznanym parametrem rzeczywistym. Rozważamy estymator $\bar{\theta}$ parametru θ równy medianie z próby $X_1, X_2, ..., X_n$

$$\bar{\theta} = X_{[0,5n]:n}.$$

W oparciu o ten estymator budujemy przedział ufności dla parametru θ postaci

$$(\bar{\theta} - a, \bar{\theta} + a),$$

gdzie a dobrane jest tak, aby dla każdego $\theta \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbf{P}_\theta(\theta \in (\bar{\theta} - a, \bar{\theta} + a)) = 0,95.$$

Wtedy a jest równe
Odp: C-> $\frac{0,98}{\sqrt{n}}$.

Rozwiązanie. Niech F_θ będzie dystrybuantą rozkładu f_θ, czyli

$$F_\theta(t) = \begin{cases} 1 - \frac{1}{2}\exp(-2|t - \theta|) & t \geqslant \theta \\ \frac{1}{2}\exp(-2|t - \theta|) & t \leqslant \theta. \end{cases}$$

Przypomnijmy, że rozkład $X_{[0,5n]:n}$ można wyznaczyć korzystając ze zmiennej Borelowskiej $S_n(t)$ z rozkładu $\mathcal{B}(n, F_\theta(t))$ to znaczy

$$\mathbf{P}_\theta(X_{[0,5n]:n} \leqslant t) = \mathbf{P}(S_n(t) \geqslant [0,5n]).$$

Z warunków zadania

$$\begin{aligned} \mathbf{P}_\theta(\theta \in (\bar{\theta} - a, \bar{\theta} + a)) &= \mathbf{P}_\theta(\theta - a < \bar{\theta} < \theta + a) = \\ &= \mathbf{P}(S_n(\theta - a) \geqslant [0,5n]) - \mathbf{P}(S_n(\theta + a) \geqslant [0,5n]), \end{aligned}$$

gdzie korzystamy z tego, że $X_{[0,5n]:n}$ ma rozkład ciągły. Będziemy tak dobierać ciąg a aby zachodziło CTG. Niech Z będzie z rozkładu $\mathcal{N}(0,1)$, wówczas jeśli $a = \frac{c}{\sqrt{n}}$ wtedy na mocy CTG

$$\lim_{n\to\infty} \mathbf{P}(S_n(\theta - a) \geqslant [0,5n]) = \mathbf{P}(Z \geqslant -2c),$$

gdzie korzystamy z równości

$$\lim_{n\to\infty} \frac{[0,5n] - F_\theta(\theta - a)}{\sqrt{nF_\theta(\theta - a)(1 - F_\theta(a - \theta))}}) = -2c.$$

Istotnie zauważmy, że

$$\lim_{n\to\infty} \sqrt{F_\theta(\theta - a)(1 - F_\theta(a - \theta))} = \frac{1}{2}$$

nadto na mocy przybliżenia $e^{-x} \simeq 1 - x$ dla małych x otrzymujemy

$$\lim_{n\to\infty} \frac{2([0,5n] - F_\theta(\theta - a))}{\sqrt{n}} = -2c$$

Analogicznie

$$\lim_{n\to\infty} \mathbf{P}(S_n(\theta + a) \geqslant [0,5n]) = \mathbf{P}(Z \geqslant 2c).$$

Dobieramy zatem $2c$ tak aby $\mathbf{P}(Z \in (-2c, 2c)) = 0,95$. Zatem $2c = 1,96$, czyli $c = 0,98$.

$\blacksquare$