

1. **Twierdzenia o dodawaniu.** Niech X_1, \dots, X_k będą **niezależnymi** zmiennymi losowymi i niech $Y = \sum_{i=1}^k X_i$. Udowodnić następujące stwierdzenia:

- Jeżeli $X_i, i=1, \dots, k$, mają rozkłady dwumianowe $B(n_i, p)$, to Y ma rozkład dwumianowy $B(\sum_{i=1}^k n_i, p)$.

- Jeżeli $X_i, i=1, \dots, k$, mają rozkłady Poissona $P(\lambda_i)$, to Y ma rozkład dwumianowy $P(\sum_{i=1}^k \lambda_i)$.

- Jeżeli $X_i, i=1, \dots, k$, mają rozkłady gamma $G(\alpha_i, \lambda)$, to Y ma rozkład gamma $G(\sum_{i=1}^k \alpha_i, \lambda)$.

2. Wykonujemy n doświadczeń losowych z których każde kończy się sukcesem z prawdopodobieństwem θ . Wiadomo, że $\theta \in [\theta_1, \theta_2]$, gdzie $\theta_1, \theta_2 \in [0, 1]$ są ustalone. Sformułować model statystyczny tego eksperymentu.

3. Pewne urządzenie techniczne pracuje dopóty, dopóki nie uszkodzi się któryś z k elementów typu A lub któryś z l elementów typu B . Czas życia elementów typu A jest zmienną losową o rozkładzie wykładniczym z gęstością $f_\alpha(x) = \alpha^{-1} \exp(-x/\alpha)$, a czas życia elementów typu B jest zmienną losową o rozkładzie wykładniczym z gęstością $f_\beta(x) = \beta^{-1} \exp(-x/\beta)$. Obserwuje się czas życia T całego urządzenia. Sformułować model statystyczny tej obserwacji.

4. Przeprowadza się $n = \sum_{i=1}^k n_i$ eksperymentów w taki sposób, że n_i eksperymentów wykonuje się na poziomie $x_i, i=1, \dots, k$. Prawdopodobieństwo sukcesu w eksperymencie przeprowadzonym na poziomie x jest równe $p(x) = \frac{1}{1 + e^{-(\alpha + \beta x)}}$, $\alpha \in R, \beta > 0$, gdzie (α, β) jest nieznanym parametrem. Sformułować model statystyczny tego eksperymentu.

5. **Pewna optymalna własność mediany.** Medianą zmiennej losowej o rozkładzie P nazywamy liczbę m_e taką, że $P(X \leq m_e) \geq \frac{1}{2}$ i $P(X \geq m_e) \geq \frac{1}{2}$. Niech X będzie całkowalną zmienną losową (tzn. $E|X| < \infty$). Pokazać, że funkcja $\varphi(c) = E|X - c|$ osiąga najmniejszą wartość, gdy $c = m_e$.

Wskazówka Rozważyć 2 przypadki 1) $c < m_e$ 2) $c > m_e$ i w każdym z nich pokazać że

$\varphi(c) - \varphi(m_e) = E|X - c| - E|X - m_e| = \psi(c)$ przy czym $\psi(c) \geq 0$ dla każdego c i $\psi(m_e) = 0$.

6. **Rodziny wykładnicze.** Pokazać, że rodziny rozkładów

- dwumianowych
- ujemnych dwumianowych
- Poissona
- normalnych
- gamma
- beta
- wielomianowych

są rodzinami wykładniczymi. W każdym przypadku znaleźć naturalną parametryzację.

7. **Statystyki pozycyjne.** Niech X_1, \dots, X_n będzie próbą prostą z rozkładu o dystrybuancie F .

Wyznaczyć rozkłady następujących statystyk pozycyjnych:

- $X_{(n)}$
- $X_{(1)}$
- $(X_{(1)}, X_{(n)})$

Czy statystyki $X_{(1)}$ i $X_{(n)}$ są niezależne?

8. Niech X_1, \dots, X_n będzie próbą prostą z rozkładu o dystrybuancie F . Statystykę $R = X_{(n)} - X_{(1)}$ nazywamy rozstępem z próby.

a) wyznaczyć rozkład R

b) wykazać, że $ER = \int_{-\infty}^{\infty} \{1 - [F(x)]^n - [1 - F(x)]^n\} dx$

9. Niech X_1, \dots, X_n będzie próbą prostą z rozkładu o dystrybuancie F . Wykazać, że

$$F_{(k)}(x) = \frac{n!}{(n-1)!(n-k)!} \int_0^{F(x)} t^{k-1} (1-t)^{n-k} dt$$

10. Niech X_1, \dots, X_n będzie próbą prostą z rozkładu ciągłego o gęstości f .

- Wykazać, że $\mathbf{X}^* = (X_{(1)}, \dots, X_{(n)})$ ma gęstość $g(y_1, \dots, y_n) = n! \prod_{j=1}^n f(y_j) \mathbf{1}_{\{y_1 < y_2 < \dots < y_n\}}(y_1, \dots, y_n)$
- Wyznaczyć gęstość k -tej statystyki pozycyjnej $X_{(k)}$.
- Wykazać, że $F(X_{(k)}) \sim B(k, n-k+1)$

11. Niech X_1, \dots, X_n będzie próbą prostą z rozkładu wykładniczego $Exp(\lambda)$

- znaleźć rozkład $X_{(k)}$
- Wykazać, że $X_{(k)}, X_{(m)} - X_{(k)}$; $1 \leq k < m \leq n$ są niezależne.
- Znaleźć rozkład zmiennej: $X_{(k+1)} - X_{(k)}$; $1 \leq k \leq n-1$

12. Rozkład wielomianowy i jego asymptotyka

Statystyki dostateczne, swobodne, zupełne

13. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Poissona $P(\lambda)$. Korzystając z definicji pokazać,

że $T(\mathbf{X}) = \sum_{i=1}^n X_i$ jest statystyką dostateczną dla parametru λ .

14. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu wykładniczego $Exp(\lambda)$. Korzystając z definicji

pokazać, że $T(\mathbf{X}) = \sum_{i=1}^n X_i$ jest statystyką dostateczną dla parametru λ .

15. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu ujemnego wykładniczego (ozn. $NE(a, \lambda)$) o

funkcji gęstości $f(x) = \frac{1}{\lambda} e^{-\frac{x-a}{\lambda}} \mathbf{1}_{(a, \infty)}(x)$, gdzie $\lambda > 0$. Korzystając z kryterium faktoryzacji wyznaczyć statystykę dostateczną dla parametru (a, λ) .

16. Wykazać, że dla próby prostej z dwuwymiarowego rozkładu normalnego o znanym współczynniku korelacji (przy nieznanymi pozostałymi parametrami) współczynnik korelacji z

próby $W = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$ jest statystyką swobodną.

17. Niech X_1, \dots, X_n i Y_1, \dots, Y_n będą niezależnymi próbkami z rozkładów $N(m_x, \sigma_x)$ i $N(m_y, \sigma_y)$ odpowiednio. Można pokazać, że statystyką dostateczną i zupełną dla $(m_x, \sigma_x, m_y, \sigma_y)$ jest

statystyka $T = (\bar{X}, \sum_{i=1}^n X_i^2, \bar{Y}, \sum_{i=1}^n Y_i^2)$, a statystyka $W = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$ jest

niezależna od T

18. Czy rodzina rozkładów normalnych $N(1, \sigma^2)$ jest zupełna? (nie $\varphi(X) = X - 1 \neq 0$ i $E_\sigma(\varphi(X)) = 0 \quad \forall \sigma$.)
19. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Cauchy'ego $C(\theta, 1)$ o gęstości $p_\theta(x) = \frac{1}{\pi} \frac{1}{1+(x-\theta)^2}$, $\theta \in \mathbb{R}$. Znaleźć minimalną statystykę dostateczną dla θ .
20. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Pareto (ozn. $Pa(x_0, a)$) o funkcji gęstości $f(x) = \frac{\alpha}{x_0} (\frac{x_0}{x})^{\alpha+1} \mathbf{1}_{(x_0, \infty)}(x)$, gdzie $a > 1$. Korzystając z kryterium faktoryzacji wyznaczyć statystykę dostateczną dla parametru (x_0, a) .
21. Rozważmy rodzinę rozkładów o gęstości $p_\theta(x) = A(\theta)e^{-(x-\theta)^4}$. Pokazać, że jest to rodzina wykładnicza. Znaleźć rząd tej rodziny. Czy jest to regularna rodzina wykładnicza? Czy jest to rodzina zupełna? (Odp. Rodzina wykładnicza rzędu 3, naturalne parametry leżą na krzywej w \mathbb{R}^3 , nie jest to rodzina zupełna)
22. Udowodnić twierdzenie. Niech $\mathcal{P}_0 \subset \mathcal{P}$ będzie podrodziną rodziny \mathcal{P} równoważną z \mathcal{P} . Jeżeli S jest minimalną statystyką dostateczną dla \mathcal{P}_0 i dostateczną dla \mathcal{P} , to S jest minimalną statystyką dostateczną dla \mathcal{P} .
- Dowód.** Niech T będzie dowolną statystyką dostateczną dla \mathcal{P} . Oczywiście tym bardziej T jest statystyką dostateczną dla \mathcal{P}_0 . Ponieważ S jest minimalną statystyką dostateczną dla \mathcal{P}_0 istnieje więc funkcja h taka, że $S = h(T)$ prawie wszędzie \mathcal{P}_0 a wobec równoważności rodzin \mathcal{P}_0 i \mathcal{P} prawie wszędzie \mathcal{P} . wynika, że $(T_1(X), \dots, T_k(X))$ **jest minimalną statystyką dostateczną dla \mathcal{P} .**
23. Niech $\mathbf{X} = (X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Gamma(a, λ) o funkcji gęstości $f(x; \alpha, \lambda) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\lambda}} \mathbf{1}_{(0, \infty)}(x)$ Korzystając z kryterium faktoryzacji wyznaczyć statystykę dostateczną dla parametru (a, λ)
24. Niech $\mathbf{Z} = ((X_1, Y_1), \dots, (X_n, Y_n))$ będzie próbą z dwuwymiarowego rozkładu normalnego $N(\mathbf{0}, \Sigma)$, gdzie $\mathbf{0}$ jest wektorem zerowym a $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. Wyznaczyć minimalną statystykę dostateczną dla ρ . Czy ta statystyka jest zupełna? Udowodnić, że statystyki $T_1(\mathbf{Z}) = \sum_{i=1}^n X_i^2$ i $T_2(\mathbf{Z}) = \sum_{i=1}^n Y_i^2$ są swobodne a statystyka (T_1, T_2) nie jest swobodna.

Estymacja nieobciążona

25. Niech $\mathbf{X} = (X_1, \dots, X_n)$ i $\mathbf{Y} = (Y_1, \dots, Y_n)$ będą niezależnymi próbkami prostymi z rozkładów odpowiednio $N(m_x, \sigma^2)$ i $N(m_y, \sigma^2)$. Który z dwóch następujących estymatorów: $T_1(\mathbf{X}, \mathbf{Y}) = \bar{X} \bar{Y}$, $T_2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n} \sum_{i=1}^n X_i Y_i$ należy przyjąć za ocenę $m_x m_y$.
26. 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 .

27. 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. $c_1, \dots, c_{10} = \frac{4}{45}$, $c_{11}, \dots, c_{15} = \frac{1}{45}$)
28. 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. $c = \frac{1}{90}$)
29. 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 .

Estymacja nieobciążona o jednostajnie minimalnej wariancji

30. Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Poissona $P(\theta)$. Wyznaczyć estymator nieobciążony o jednostajnie minimalnej wariancji $g(\theta)=p_\theta(0)=e^{-\theta}$. Wyznaczyć wariancję tego estymatora. (Odp. $\hat{g}(X) = (1 - \frac{1}{n}) \sum_{i=1}^n X_i$, $V_\theta(\hat{g}) = e^{-2\theta} (e^{\frac{\theta}{n}} - 1)$).
Uwaga. Kształtujące wydaje się rozwiązanie powyższego zadania dwoma sposobami (1) równanie 2) tw. Lehmana Scheffee'go)
31. Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu $N(m, \sigma^2)$ σ -znane. Wyznaczyć estymator nieobciążony o minimalnej wariancji a) $g(m)=m$, b) $g(m)=m^2$.
(Odp: a) $g(X) = \bar{X}$ b) $g(X) = \bar{X}^2 - \frac{\sigma^2}{n}$.
32. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Bernoulliego $p(x) = \theta^x(1-\theta)^{1-x}$, $x=0,1$; $\theta \in (0,1)$. Wyznaczyć estymator nieobciążony o minimalnej wariancji funkcji $g(\theta)=\theta^2$.
33. Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu $N(m, \sigma^2)$ m -znane. Wyznaczyć ENJMW[σ] i b) ENJMW[σ^2]
(Odp: a) $\frac{\Gamma(\frac{n}{2})}{\sqrt{2} \Gamma(\frac{n+1}{2})} S$ b) $\frac{1}{n} S^2$ (gdzie $S^2 = \sum_{i=1}^n (X_i - m)^2$)
34. Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu $N(m, \sigma^2)$. Wyznaczyć ENJMW dla a) m b) σ^2 c) σ d) $\frac{m}{\sigma}$.
(Odp. a) \bar{X} , b) $\frac{S^2}{n-1}$ c) $\frac{\Gamma(\frac{n-1}{2})}{\sqrt{2} \Gamma(\frac{n}{2})} S$ d) $\frac{\sqrt{2} \Gamma(\frac{n-1}{2}) \bar{X}}{\Gamma(\frac{n}{2}-1) S}$ (gdzie $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$)

35. Niech $\mathbf{X}=(X_1,\dots,X_n)$ będzie próbą prostą z rozkładu o absolutnie ciągłej dystrybuancie $F \in \mathcal{F}$. Zakładając istnienie $g_1(F)=E_F(X)$ i $g_2(F)=V_F(X)$ wyznaczyć a) ENJMW $[E_F(X)]$ i b) ENJMW $[V_F(X)]$.

Odp: a) \bar{X} b) $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ (Statystyką dostateczną zupełną jest wektor statystyk pozycyjnych)

36. Niech $\mathbf{X}=(X_1,\dots,X_n)$ będzie próbą prostą z rozkładu o gęstości $f(x) = \begin{cases} \frac{\beta}{x^2}, & \text{gd } x \geq \alpha \\ 0, & \text{gd } x < \alpha \end{cases}$ gdzie $\alpha > 0$.

Obliczyć β a następnie w oparciu o n elementową próbę prostą wyznaczyć estymator nieobciążony o minimalnej wariancji parametru α . Czy jest on zgodny?

37. Niech (X_1,\dots,X_m) i (Y_1,\dots,Y_n) będą niezależnymi próbkami prostymi z rozkładów $U(0, \theta_x)$ i $U(0, \theta_y)$: $\theta_x, \theta_y > 0$. Wyznaczyć estymator nieobciążony o minimalnej wariancji dla $\frac{\theta_x}{\theta_y}$.

38. Niech $\mathbf{X}=(X_1,\dots,X_n)$ będzie wektorem losowym o niezależnych współrzędnych takich, że X_i mają rozkłady normalne $N(\alpha + \beta t_i, \sigma^2)$, gdzie α, β, σ^2 są nieznanymi a t_i są znanymi stałymi różnymi między sobą, $i=1,\dots,n$. Wyznaczyć estymatory nieobciążone o minimalnej wariancji parametrów α i β na podstawie obserwacji wektora \mathbf{X} .

39. Niech $\mathbf{X}=(X_1,\dots,X_n)$ będzie próbą prostą z rozkładu o rozkładzie jednostajnym $U(0, \theta)$, $\theta \in (0, \infty)$. Wyznaczyć ENJMW funkcji $g(\theta)$ różniczkowalnej na $(0, \theta)$. Następnie podać postać estymatora dla $g_1(\theta) = \frac{\theta}{2} = E_\theta(X_1)$ i $g_2(\theta) = \frac{\theta^2}{12} = V_\theta(X_1)$.

Odp: $\hat{g}(X_{(n)}) = g(X_{(n)}) + \frac{1}{n} X_{(n)} g'(X_{(n)})$, $\hat{g}_1(X_{(n)}) = \frac{n+1}{2n} X_{(n)}$, $\hat{g}_2(X_{(n)}) = g(X_{(n)}) + \frac{n+2}{12n} X_{(n)}^2$

40. Niech $\mathbf{X}=(X_1,\dots,X_m)$ i $\mathbf{Y}=(Y_1,\dots,Y_n)$ będą niezależnymi próbkami z rozkładów $N(m_x, \sigma_x^2)$ i $N(m_y, \sigma_y^2)$.

- a) Wyznaczyć ENJMW $g_1 = m_x - m_y$ i $g_2 = \left(\frac{\sigma_x}{\sigma_y}\right)^r$, $r > 0$

- b) Niech $\sigma_x^2 = \sigma_y^2$. Wyznaczyć ENJMW $g_1 = m_x - m_y$ i $g_2 = \frac{m_x - m_y}{\sigma_x}$

- c) Niech $m_x = m_y$ i $\frac{\sigma_x^2}{\sigma_y^2} = \gamma$, gdzie γ znane. Wyznaczyć ENJMW $[m_x]$. Odp ($\hat{m}_x = \frac{m}{m+n\gamma} \bar{X} + \frac{n\gamma}{m+n\gamma} \bar{Y}$)

- d) Niech $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Wyznaczyć ENJMW $[\sigma^2]$.

- e) Niech $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Wyznaczyć estymator nieobciążony o minimalnej wariancji parametru funkcji ENJMW $[m_x - m_y + \sigma]$.

Estymacja bayesowska i minimaksowa

41. Niech $\mathbf{X}=(X_1,\dots,X_n)$ będzie próbą prostą z rozkładu wykładniczego $p(x|\theta) = \theta e^{-\theta x} \mathbf{1}_{(0,\infty)}$, $\theta > 0$.

Wyznaczyć estymator bayesowski parametru θ (przy kwadratowej funkcji straty) jeżeli rozkładem „a priori” tego parametru jest rozkład Gamma(p, a) o funkcji gęstości

$$f(\theta) = \frac{a^p}{\Gamma(p)} \theta^{p-1} e^{-a\theta} \mathbf{1}_{(0,\infty)}(\theta).$$

42. Niech X_1, \dots, X_n będzie próbą prostą z rozkładu geometrycznego $p(x|\theta) = \theta(1-\theta)^{x-1}$, $\theta \in (0,1)$, $x=1,2,\dots$. Wyznaczyć estymator bayesowski parametru θ przy kwadratowej funkcji strat, jeżeli nie dysponujemy żadną wiedzą „a priori” odnośnie parametru θ .
43. Niech X będzie zmienną losową o rozkładzie geometrycznym $p(x,\theta) = \theta^x(1-\theta)$, $\theta \in (0,1)$, $x=1,2,\dots$. Załóżmy, że parametr θ jest realizacją zmiennej losowej Θ o rozkładzie $\pi(\theta) = 3\theta^2 \mathbf{1}_{(0,1)}(\theta)$. Wyznaczyć estymator bayesowski parametru θ . (Odp $\hat{\theta}(X) = \frac{X+3}{X+5}$)
44. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Poissona $p(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, $\lambda > 0$, $x=0,1,\dots$. Wyznaczyć estymator bayesowski parametru λ^2 (przy kwadratowej funkcji straty) jeżeli rozkładem „a priori” tego parametru jest rozkład wykładniczy o funkcji gęstości $f(\lambda) = e^{-\lambda} \mathbf{1}_{(0,\infty)}(\lambda)$.
45. Towarzystwo ubezpieczeniowe szacuje, że około 20% pasażerów linii lotniczych wykupuje polisę ubezpieczeniową na przelot. Proporcja ta jest różna dla różnych terminali, przy czym zmienność charakteryzowana odchyleniem standardowym wynosi 10%. Losowa próbka 50 pasażerów wybranego terminalu pokazała, że tylko 5 nabyło polisę. Przyjmując jako rozkład a priori odpowiedni rozkład Beta oszacować prawdziwą proporcję pasażerów w tym terminalu, którzy nabyli polisę.
46. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Poissona $p(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$, $\lambda > 0$, $x=0,1,\dots$. Wyznaczyć estymator bayesowski parametru λ (przy kwadratowej funkcji straty) jeżeli rozkład „a priori” tego parametru ma gęstość $f(\lambda) = \alpha e^{-\alpha\lambda} \mathbf{1}_{(0,\infty)}(\lambda)$.
47. Przypuśćmy, że X ma rozkład $U(0,\theta)$. Załóżmy, że rozkładem „a priori” parametru θ jest rozkład $\pi(\theta) = \theta e^{-\theta}$, $\theta > 0$. Wyznaczyć estymator bayesowski parametru θ przy funkcji strat a) $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$. b) $L(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$
48. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu geometrycznego $p(x|\theta) = \theta(1-\theta)^{x-1}$, $\theta \in (0,1)$, $x=1,2,\dots$. Niech rozkładem „a priori” dla θ będzie rozkład Beta(α, β). Wyznaczyć estymator bayesowski parametru θ przy kwadratowej funkcji strat.
49. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu geometrycznego $p(x|\theta) = \theta(1-\theta)^x$, $0 < \theta < 1$, $x=0,1,\dots$. Wyznaczyć estymator bayesowski parametru θ (przy kwadratowej funkcji straty) jeżeli rozkładem „a priori” tego parametru jest rozkład jednostajny $U(0,1)$ o funkcji gęstości $f(\theta) = \mathbf{1}_{(0,1)}(\theta)$.
50. Przypuśćmy, że prawdopodobieństwo θ sukcesu w ciągu n prób Bernoulliego ma rozkład (a priori) jednostajny $U(0,1)$. Wyznaczyć estymator bayesowski parametru θ^2 (przy kwadratowej funkcji strat).
51. Niech (X_1, \dots, X_n) będzie próbą prostą z rozkładu Poissona $P(\lambda)$. Wykazać, że statystyka \bar{X} jest minimaxowym estymatorem parametru λ przy funkcji straty $L(\lambda, d) = \frac{(\lambda - d)^2}{\lambda}$. Czy \bar{X} jest estymatorem dopuszczalnym?
52. Czas świecenia żarówek jest zmienną losową o rozkładzie wykładniczym z funkcją gęstości $p(x|\theta) = \theta e^{-\theta x} \mathbf{1}_{(0,\infty)}(x)$. Przypuśćmy, że θ ma rozkład „a priori” wykładniczy z funkcją gęstości $f(\theta) = e^{-\theta} \mathbf{1}_{(0,\infty)}(\theta)$. Wyznaczyć estymator bayesowski parametru θ (przy kwadratowej funkcji straty) oparty na n elementowej próbie prostej.

53. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu normalnego $N(\theta, 1)$. Wyznaczyć estymator bayesowski parametru θ (przy kwadratowej funkcji straty) jeżeli rozkładem „a priori” tego parametru jest rozkład wykładniczy o funkcji gęstości $f(\theta) = e^{-\theta} \mathbf{1}_{(0, \infty)}(\theta)$.

(Uwaga. Rozkładem a posteriori jest ucięty rozkład normalny - zadanie na ćwiczenia)

54. Niech X będzie zmienną losową o rozkładzie jednostajnym $U(0, \theta)$. Wyznaczyć estymator bayesowski parametru θ (przy kwadratowej funkcji straty) jeżeli rozkładem „a priori” tego parametru jest rozkład $f(\theta) = \theta e^{-\theta} \mathbf{1}_{(0, \infty)}(\theta)$.

Metoda największej wiarygodności

55. 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. $\hat{k} = X_{(n)}$

56. 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. $(\hat{\theta}_1, \hat{\theta}_2) = (X_{(1)}, X_{(n)})$).

57. 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 $1-\alpha$.

58. 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. $\hat{\lambda} = -\ln(1 - \bar{M})$)

59. Skonstruować estymator największej wiarygodności parametru θ i asymptotyczny przedział ufności na poziomie $1-\alpha=0.95$ oparty na n elementowej próbie prostej niezależnych obserwacjach X_1, X_2, \dots, X_n z rozkładu

a) geometrycznego $p(x, \theta) = \theta(1-\theta)^{x-1}$, $\theta \in (0, 1)$, $x=1, 2, \dots$,

b) geometrycznego $p(x, \theta) = (1-\theta)\theta^{x-1}$, $\theta \in (0, 1)$, $x=1, 2, \dots$.

c) wykładniczego $p(x, \theta) = \theta \exp(-\theta x)$, $\theta > 0$, $x > 0$,

d) wykładniczego $p(x, \theta) = \theta^{-1} \exp(-x/\theta)$, $\theta > 0$, $x > 0$.

e) o gęstości $p(x, \theta) = \theta^2 x \exp(-\theta x)$, $\theta > 0$, $x > 0$,

f) normalnego $p(x, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-1)^2}{2\theta}}$, $\theta > 0$.

g) normalnego $p(x, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\theta)^2}{2\theta}}$, $\theta > 0$.

h) Poissona $p(x, \theta) = \frac{\theta^x}{x!} e^{-\theta}$, $\theta > 0$, $x=0, 1, 2, \dots$.

i) z rozkładu Bernoulliego $p(x, \theta) = \theta^x (1-\theta)^{1-x}$, $\theta \in (0, 1)$, $x=0, 1$.

j) z rozkładu Pareto (ozn. $Pa(1, a)$) o funkcji gęstości $f(x) = ax^{-a-1} \mathbf{1}_{(1, \infty)}(x)$, gdzie $a > 1$

60. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu normalnego z funkcją gęstości $f(x) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\theta)^2}{2\theta}}$. Metodą największej wiarygodności wyznaczyć estymator parametru θ . Wyznaczyć asymptotyczną wariancję tego estymatora.

61. 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 $Y = -\ln X$)
62. 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 $p(x, \theta) = \frac{\theta}{2} e^{-\theta|x|}$, $\theta > 0$.

Efektywność estymatora

63. Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu $N(m, \sigma^2)$, gdzie m jest znane. Pokazać, że estymator $\tilde{S}^2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ jest efektywnym estymatorem parametru σ^2 .
64. Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu $N(m, \sigma^2)$, gdzie m jest znane. Pokazać, że estymator $U(\mathbf{X}) = \frac{\sqrt{n} \Gamma(\frac{n}{2})}{\sqrt{2} \Gamma(\frac{n+1}{2})} \tilde{S}(\mathbf{X})$ gdzie $\tilde{S}^2(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ jest
- nieobciążonym estymatorem parametru σ
 - nie jest efektywnym estymatorem parametru σ .
65. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Laplace'a $\Lambda(0, \beta)$. Wyznaczyć estymator nieobciążony o minimalnej wariancji funkcji parametru
- a) $g(\beta) = \beta$ b) $g(\beta) = \beta^2$ i sprawdzić, czy jego wariancja osiąga dolne ograniczenie Cramera-Rao.
66. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu wykładniczego $f(x, \theta) = \theta^{-1} e^{-\frac{x}{\theta}} \mathbf{1}_{(1, \infty)}(x)$. Wyznaczyć estymator nieobciążony o minimalnej wariancji funkcji parametru $g(\theta) = \theta^2$ i sprawdzić, czy jego wariancja osiąga dolne ograniczenie Cramera-Rao.
- (Odp. $\hat{g}(X) = \frac{(\sum_{i=1}^n X_i)^2}{n(n+1)}$, $V(\hat{g}) = \frac{4\theta^4}{n} (1 + \frac{1}{2(n+1)}) > \frac{4\theta^4}{n}$ (dolne ograniczenie CR)).
67. Wyznaczyć macierz informacji Fishera dla próby prostej z rozkładu $N(m, \sigma^2)$
68. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu Laplace'a $\Lambda(0, \beta)$ ($f(x) = \frac{1}{2\beta} e^{-\frac{|x|}{\beta}}$). Wyznaczyć estymator nieobciążony o minimalnej wariancji funkcji parametru $g(\beta) = \beta^2$ i sprawdzić, czy jego wariancja osiąga dolne ograniczenie Cramera-Rao.

Przedziały ufności

69. Niech X_1, \dots, X_m i Y_1, \dots, Y_n będą dwoma niezależnymi próbkami prostymi z rozkładów $N(\mu_1, \sigma^2)$ i $N(\mu_2, \sigma^2)$. Znaleźć najkrótszy przedział ufności dla $\mu_1 - \mu_2$ oparty na odpowiedniej statystyce dostatecznej gdy σ^2 jest znane.
70. Rozważmy próbę losową X_1, \dots, X_n z rozkładu jednostajnego na odcinku $[0, \theta]$ z nieznanym prawym końcem θ . Niech $M = \max(X_1, \dots, X_n)$. Należy zbudować przedział ufności dla θ na poziomie 90%. Chcemy aby ten przedział był postaci $[aM, bM]$, gdzie liczby a i b są tak dobrane aby $P(\theta < aM) = P(\theta > bM) = 0.05$. podaj długość tego przedziału.
71. Załóżmy, że X_1, \dots, X_4 jest próbą prostą z rozkładu normalnego $N(m, \sigma^2)$ o nieznannej wartości oczekiwanej i nieznannej wariancji, zaś X_5 jest zmienną losową z tego samego rozkładu niezależną od próbki. Interpretujemy zmienną X_5 jako kolejną obserwację, która pojawi się w przyszłości, ale obecnie jest nieznaną. Zbuduj „przedział ufności” $[L, U] = [L(X_1, \dots, X_4), U(X_1, \dots, X_4)]$ oparty na próbce X_1, \dots, X_4 taki, że $P(L(X_1, \dots, X_4) \leq X_5 \leq U(X_1, \dots, X_4)) = 0.95$, przy czym żądamy, aby przedział był symetryczny tzn. $\frac{1}{2}(L + U) = \bar{X}$. Używamy tu oznaczeń $\bar{X} = \frac{1}{4} \sum_{i=1}^4 X_i$, $S^2 = \frac{1}{3} \sum_{i=1}^4 (X_i - \bar{X})^2$.
72. Wyznaczyć najkrótszy przedział ufności dla parametru θ w rozkładzie jednostajnym $U(0, \theta)$ oparty na n elementowej próbce prostej. Funkcję centralną oprzeć na minimalnej statystyce dostatecznej. **Odp.** $[X_{(n)}, \frac{1}{\alpha} X_{(n)}]$, Funkcją centralną jest funkcja $T(X_1, \dots, X_n, \theta) = \frac{X_{(n)}}{\theta}$.

Model liniowy i MNK

73. Pewien deterministyczny proces x_0, x_1, \dots, x_n przebiega tak że $x_{i+1} = ax_i$, $i=0, \dots, n-1$, przy czym a jest znaną stałą. Wielkości x_i nie mogą być obserwowane bezpośrednio, lecz ich obserwacje mają postać $y_i = x_i + \varepsilon_i$, $i=0, \dots, n$, gdzie ε_i są nieskorelowanymi błędami o jednakowej wariancji σ^2 . Metodą najmniejszych kwadratów wyznaczyć estymatory x_0, x_1, \dots, x_n . Wyznaczyć również estymator wariancji błędów σ^2 .
74. W modelu regresji liniowej $y_i = \theta_1 + \theta_2 x_i + \varepsilon_i$, $i=1, \dots, n$, założmy, że $\text{Var}(\varepsilon_i) = w_i \sigma^2$, gdzie σ^2 jest nieznaną a w_1, \dots, w_n są znane. Mówimy, że $\tilde{\theta}_1$ i $\tilde{\theta}_2$ są estymatorami parametrów θ_1 i θ_2 uzyskanymi metodą ważonych najmniejszych kwadratów, gdy

$$\sum_{i=1}^n \frac{1}{w_i} (y_i - \tilde{\theta}_1 - \tilde{\theta}_2 x_i)^2 = \min_{\theta_1, \theta_2} \sum_{i=1}^n \frac{1}{w_i} (y_i - \theta_1 - \theta_2 x_i)^2.$$

Napisać odpowiednik układu równań normalnych dla tej sytuacji i wyznaczyć estymatory parametrów θ_1 , θ_2 oraz σ^2 metodą ważonych najmniejszych kwadratów.

75. Należy wyznaczyć ciężary netto $\beta_1, \beta_2, \beta_3$ trzech przedmiotów oraz ciężar opakowania β_0 za pomocą czterech wazek przy założeniu, że przy każdym ważeniu popełniamy losowy błąd o rozkładzie normalnym ze średnią 0 i wariancją σ^2 . Rozpatrzmy dwa plany eksperymentu. Pierwszy z nich polega na zważeniu każdego przedmiotu (z opakowaniem) z osobna, a na końcu zważeniu samego opakowania. W drugim schemacie ważymy najpierw opakowanie a następnie trzykrotnie po dwa przedmioty (każdy z opakowaniem), nie powtarzając doboru ważonych par. Niech Y_i , $i=1, 2, 3, 4$ oznaczają kolejne wskazania wagi. Wyznaczyć metodą najmniejszych kwadratów oszacowania ciężarów β_i . Sprawdzić, który z planów jest lepszy ze względu na wielkość wariancji uzyskanych estymatorów.

76. Dla wybranych n wartości x_i z przedziału $[-1,1]$ możemy wykonać pomiary wielkości $y_i = \beta_0 + \beta_1 x_i$ obarczone niezależnymi zmiennymi losowymi o zerowych wartościach oczekiwanych i wariancjach σ^2 i estymować wartości β_0 i β_1 metodą najmniejszych kwadratów. Załóżmy, że n jest parzyste. Jak należy wybrać wielkości x_i , aby wariancja estymatora β_1 była najmniejsza?

(Odp. $V(\hat{\beta}_1) = \frac{n}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$ -- połowa z x_i jest równa -1 a druga połowa 1.)

77. Na podstawie pomiarów wielkości $y_i = \beta_0 + \beta_1 x_i$ ($i=1, \dots, n$), które są obarczone niezależnymi błędami losowymi o zerowych wartościach oczekiwanych i wariancjach σ^2 , konstruujemy estymatory $\hat{\beta}_0$ i $\hat{\beta}_1$ metodą najmniejszych kwadratów. Prognozę dla wartości $y_0 = \beta_0 + \beta_1 x_0$ konstruujemy jako $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$. Wyznaczyć wariancję tej prognozy. Dla jakiego x_0 wariancja \hat{y}_0 jest najmniejsza.

(Odp. $V(\hat{y}_0) = \frac{\sigma^2 (\sum_{i=1}^n x_i^2 - 2x_0 \sum_{i=1}^n x_i + n x_0^2)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$ jest najmniejsza dla $x_0 = \frac{1}{n} \sum_{i=1}^n x_i$)

78.

Testowanie hipotez

79. X_1, X_2, \dots, X_{20} jest próbą losową z rozkładu normalnego o parametrach $N(0, \sigma^2)$. Znaleźć najmocniejszy test hipotezy: $H_0: \sigma^2 = 1$ przeciwko alternatywie: $H_1: \sigma^2 = 3$ na poziomie istotności $\alpha=0.01$. Oszacować moc tego testu (przy szacowaniu mocy można skorzystać z aproksymacji $\sqrt{2\chi_n^2} - \sqrt{2n-1} \sim N(0,1)$ (Odp. ok. 90%)

80. Zmienna losowa X ma funkcję gęstości $f(x) = \begin{cases} (1+b)x^b, & \text{dla } 0 \leq x \leq 1 \\ 0, & \text{poza tym} \end{cases}$. Skonstruować najmocniejszy test hipotezy $H_0: b=1$ przeciwko $H_1: b=2$ na poziomie $\alpha=0.1$ gdy dysponujemy pojedynczą obserwacją zmiennej X . (Odp. $C = \{ X : X > \sqrt{0.9} \}$)

81. Zmienna losowa X ma funkcję gęstości $f(x)$. Na podstawie pojedynczej obserwacji skonstruować najmocniejszy test hipotezy $H_0: f(x) = \mathbf{1}_{[0,1]}(x)$ przeciwko $H_1: f(x) = 5x^4 \mathbf{1}_{[0,1]}(x)$ na poziomie α . Wyznaczyć moc tego testu. (Odp. $C = \{ X : X > 1 - \alpha \}$ moc = $1 - (1 - \alpha)^5$)

82. Sformułować lemat Neymana-Pearsona. Zmienna losowa X ma funkcję gęstości $f(x) = \begin{cases} (1+b)x^b, & \text{dla } 0 \leq x \leq 1 \\ 0, & \text{poza tym} \end{cases}$. Skonstruować najmocniejszy test hipotezy $H_0: b=1$ przeciwko $H_1: b=2$ na poziomie $\alpha=0.1$ gdy dysponujemy pojedynczą obserwacją zmiennej X . (Odp. $C = \{ X : X > \sqrt{0.9} \}$)

83. Niech $\mathbf{X}=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu normalnego $N(0, \theta^2)$. Skonstruować najmocniejszy test hipotezy $H_0: \theta = \theta_0$ przeciwko alternatywie $H_1: \theta = \theta_1 > \theta_0$ na poziomie α . Wyznaczyć moc tego testu (wykorzystując dystrybuantę odpowiedniego rozkładu)

84. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu o gęstości $f(x) = \frac{x}{\theta^2} e^{-\frac{x}{\theta}} \mathbf{1}_{(0, \infty)}(x)$. Skonstruować najmocniejszy test hipotezy $H_0: \theta = \theta_0 > 0$ przeciwko alternatywie $H_1: \theta = \theta_1 > \theta_0$ na poziomie α . Wyznaczyć moc tego testu (Wskazówka. Tw. o dodawaniu dla rozkładów Gamma: Jeżeli $X_1 \sim \text{Gamma}(\alpha_1, \lambda)$, $X_2 \sim \text{Gamma}(\alpha_2, \lambda)$, X_1, X_2 niezależne, to $X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \lambda)$).
85. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu o funkcji gęstości $f(x, \theta) = \theta x^{\theta-1} \mathbf{1}_{(0,1)}(x)$. Skonstruować najmocniejszy test hipotezy $H_0: \theta = 1$ przeciwko alternatywie $H_1: \theta = \theta_1 > 1$ na poziomie α . Następnie przyjmując $n=1$ (1 obserwacja) wyznaczyć dokładnie (tzn. wyznaczyć odpowiednią stałą) postać testu i znaleźć jego moc.
86. Niech $X=(X_1, \dots, X_n)$ będzie próbą prostą z rozkładu o funkcji gęstości $f(x, \theta) = e^{-(x-\theta)} \mathbf{1}_{(\theta, \infty)}(x)$. Skonstruować najmocniejszy test hipotezy $H_0: \theta = \theta_0$ przeciwko alternatywie $H_1: \theta = \theta_1 > \theta_0$ na poziomie α . (nieco trudniejsze)
87. Obserwujemy parę zmiennych losowych (X, Y) . Zakładamy, że są to zmienne niezależne. X ma rozkład $N(m_x, 1)$ a Y ma rozkład $N(m_y, 1/3)$ (w nawiasie podane są wariancje a nie odchylenia standardowe). Rozważmy najmocniejszy test hipotezy $H_0: (m_x, m_y) = (0, 0)$ przeciwko alternatywie $H_1: (m_x, m_y) = (1, 1)$ na poziomie istotności $\alpha = 0.1$. Wyznaczyć moc tego testu.
88. Skonstruować test najmocniejszy na poziomie α dla $H_0: X \sim U[0, 1]$ przy alternatywie $H_1: X \sim U[0, 2]$ gdy dysponujemy pojedynczą obserwacją zmiennej X .
89. Niech X_1, \dots, X_n będzie próbą prostą z rozkładu normalnego $N(m, 1)$ o nieznannej wartości oczekiwanej znanej wariancji $\sigma^2 = 1$. Rozpatrzmy zadanie testowania hipotezy $H_0: m = 0$ przeciwko alternatywie $H_1: m = 0.5$. Należy zbudować taki test dla którego **suma** prawdopodobieństw błędów I i II rodzaju, oznaczanych odpowiednio α i β **jest najmniejsza**. Oblicz tę najmniejszą wartość $\alpha + \beta$.
90. Zmienna losowa X ma funkcję gęstości $f(x)$. Na podstawie pojedynczej obserwacji skonstruować najmocniejszy test hipotezy $H_0: f(x) = \mathbf{1}_{[0,1]}(x)$ przeciwko $H_1: f(x) = 5x^4 \mathbf{1}_{[0,1]}(x)$ na poziomie α . Wyznaczyć moc tego testu.