Stanisław Jaworski
Wojciech Zieliński

Zbiór zadań z rachunku prawdopodobieństwa i statystyki matematycznej

Wersja 3/12/2014

Spis treści

Przedmowa

Niniejsza książeczka powstała na bazie wieloletnich doświadczeń dydaktycznych prowadzenia przedmiotów *Rachunek prawdopodobieństwa* oraz *Statystyka Matematyczna* na kierunku Ekonometria i Informatyka w Szkole Głównej Gospodarstwa Wiejskiego w Warszawie. Jej celem jest zebranie podstawowych informacji pojawiających się w trakcie wykładów oraz udostępnienie Studentom zadań.

Książeczka składa się z rozdziałów będących tematami kolejnych spotkań wykładowych i ćwiczeniowych. Na początku każdego rozdziału podane są najważniejsze definicje, twierdzenia i wzory, następnie rozwiązane są przykładowe zadania i podane są zadania do samodzielnego rozwiązania. Nie należy traktować książki jako wykładu z rachunku prawdopodobieństwa i statystyki matematycznej, lecz jako przewodnik do lepszego zrozumienia wykładów i ćwiczeń.

Do zadań nie podano odpowiedzi. Wynika to stąd, że celem tych zadań jest nie tyle uzyskanie konkretnego wyniku liczbowego, ile „zmuszenie" Czytelnika do maksymalnie samodzielnej analizy zagadnienia. Poza tym, każde z prezentowanych zadań można rozwiązać korzystając z różnych technik matematycznych, Na zakończenie podano garść informacji bibliograficznych. Literatura przedmiotu jest bardzo bogata i wybór tych, a nie innych książek po pierwsze nie wyczerpuje spektrum bibliograficznego, a po drugie podyktowany jest ich dostępnością w księgarniach i bibliotekach. Ze względu na podstawowy charakter wykładu w spisie literatury można znaleźć zarówno pozycje najnowsze, jak i sprzed kilkudziesięciu lat. Należy zwrócić uwagę, że ich „zabytkowość" w niczym nie umniejsza ich wartości merytorycznej i poznawczej.

Powtórka z analizy matematycznej i algebry liniowej

Pochodna funkcji. Znaleźć pochodne następujących funkcji:

1. $f(x) = x^p(1-x)^q$

2. $f(x) = (1+x)\sqrt{2+x^2}\sqrt[3]{3+x^3}$

3. $f(x) = \sqrt{x+\sqrt{x+\sqrt{x}}}$

4. $f(x) = e^{-x^2}$

5. $f(x) = e^x(x^2-2x+2)$

6. $f(x) = e^x + e^{e^x} + e^{e^{e^x}}$

7. $f(x) = \left[\frac{1-x^2}{2}\sin x - \frac{(1+x)^2}{2}\cos x\right]e^{-x}$

8. $f(x) = \ln(\ln^2(\ln^3 x))$

9. $f(x) = \ln\frac{1-\sqrt{x}}{\sqrt{1+\sqrt[3]{x}+\sqrt[3]{x^2}}}$

10. $f(x) = \ln\left(e^x + \sqrt{1+e^{2x}}\right)$

11. $f(x) = x + x^x + x^{x^x}$

12. $f(x) = \log_x e$

13. $f(x) = \frac{1}{4\sqrt{3}}\ln\frac{\sqrt{x^2+2}-x\sqrt{3}}{\sqrt{x^2+2}+x\sqrt{3}}$

Przebieg zmienności funkcji. Zbadać przebieg zmienności i narysować wykresy następujących funkcji:

1. $f(x) = \frac{x^n}{1-x^m}$ w zależności od parametrów $n, m > 0$

2. $f(x) = x^m\exp\{-x^n\}$ w zależności od parametrów n, m

3. $f(x) = x^4 + x^3 + x^2$

4. $f(x) = \frac{x^3}{x^2-1}$

5. $f(x) = x^3 - 3x^2 - 9x + 10$

6. $f(x) = \frac{2x^2}{(2-x)^2}$

7. $f(x) = \frac{2|x|^2}{(2-|x|)^2}$

8. $f(x) = \sqrt{2x-x^2}$

9. $f(x) = x^4 + 4x - 2$

10. $f(x) = x^3 - 3x^2$

11. $f(x) = \exp\left\{\frac{x^2}{x^2-1}\right\}$

12. $f(x) = (x-3) + \frac{2x-6}{x+2} + \frac{4x-12}{(x+2)^2} + \cdots$

13. $f(x) = \frac{\log x}{x}$

Różniczkowanie szeregu. Jeżeli wyrazy szeregu są w pewnym przedziale różniczkowalne i szeregi $\sum f_n(x)$ oraz $\sum f_n'(x)$ są na tym przedziale jednostajnie zbieżne, to $\left(\sum f_n(x)\right)' = \sum f_n'(x)$

Całka nieoznaczona. Wyznaczyć całki:

1. $\int \left(x^2 - x^{-5} + \frac{1}{x} + e^x\right)\, dx$,

2. $\int xe^x\, dx$,

3. $\int x^2 e^x\, dx$,

4. $\int (5^x + \cos x)\, dx$,

5. $\int \left(\frac{1}{\sin^2 x} + \frac{1}{\sqrt{1-x^2}}\right)\, dx$,

6. $\int (x\cos x + x\ln x)\, dx$

7. $\int \frac{f'(x)}{f(x)}\, dx$,

8. $\int \frac{1/x+1}{x+\ln x}\, dx$,

9. $\int \left(\frac{1}{x+3} + \frac{1}{(x-2)^2}\right)\, dx$,

10. $\int (x+3)^2 e^{x+3}\, dx$,

11. $\int \sin(x+\pi/2)\, dx$,

12. $\int \frac{dx}{\sqrt{1-(x-2)^2}}$

13. $\int \frac{1}{(x+3)^2+1}\, dx$,

Całka oznaczona. Wyznaczyć całki:

1. $\int\limits_0^{\pi/2} (x^2 + 3x - \sin x + \mathrm{arctg} x + \ln(x+1))\, dx$

2. $\int\limits_0^1 \frac{1}{x-1}\, dx$

3. $\int\limits_0^\infty \frac{1}{x}\, dx$

4. $\int\limits_2^3 \frac{1}{x-2}\, dx$

5. $\int\limits_2^3 (x-2)^2\, dx$

6. $\int\limits_{1}^{\infty} \ln(x-1)\, dx$

7. $\int\limits_{-\pi/2}^{\pi/2} \sin(x+\pi/2)\, dx$

8. $\int\limits_{2}^{3} x \ln x\, dx$

9. $\int\limits_{0}^{\infty} (x^2e^{-x} - xe^{-x} - e^{-x})\, dx$

10. $\int\limits_{-\infty}^{\infty} x^2 I_{(-1,1)}(x)\, dx$

11. $\int\limits_{-\infty}^{\infty} \frac{1}{x^2}\, dx$

12. $\int\limits_{-\infty}^{\infty} \frac{1}{(x+2)^2}\, dx$

13. $\int\limits_{0}^{\infty} \left(x^3 I_{(-1,1)}(x) + x^4 I_{(1,5)}(x) + x^2e^{-x} I_{(1,15)}\right)\, dx$

Wyznaczanie całek dwukrotnych.

1. **O zamianie całki podwójnej na iterowaną.** Niech funkcja f będzie ciągła na prostokącie $A =< a, b > \times < c, d >$. Wtedy

$$\int_A f(x,y)\,dx\,dy = \int_a^b \left(\int_c^d f(x,y)\,dy\right)\,dx = \int_c^d \left(\int_a^b f(x,y)\,dx\right)\,dy$$

2. **Całka podwójna z funkcji o rozdzielonych zmiennych.** Jeżeli funkcja g jest ciągła na $< a, b >$ oraz funkcja h jest ciągła na $< c, d >$ to

$$\int_A g(x)h(y)\,dx\,dy = \left(\int_a^b g(x)\,dx\right) \cdot \left(\int_c^d h(y)\,dy\right)$$

gdzie $A =< a, b > \times < c, d >$.

3. **Całki iterowane po obszarach normalnych.** Jeżeli funkcja f jest ciągła na obszarze normalnym $D = \{(x,y)|\ a \le x \le b, g(x) \le y \le h(x)\}$ to

$$\int_D f(x,y)\,dx\,dy = \int_a^b \left(\int_{g(x)}^{h(x)} f(x,y)\,dy\right)\,dx$$

Jeżeli funkcja f jest ciągła na obszarze normalnym $D = \{(x,y)|\ c \le y \le d, g(y) \le x \le h(y)\}$ to

$$\int_D f(x,y)\,dx\,dy = \int_c^d \left(\int_{g(y)}^{h(y)} f(x,y)\,dx\right)\,dy$$

4. Współrzędne biegunowe w całce podwójnej. Niech obszar Δ we współrzędnych biegunowych będzie obszarem normalnym, funkcja f będzie ciągła na obszarze D, który jest obrazem obszaru Δ przy przekształceniu biegunowym:

$$D = B(\Delta) = \{(\varrho\cos(\varphi), \varrho\sin(\varphi))|\ (\varrho, \varphi) \in \Delta\}$$

Wtedy

$$\int_D f(x,y)\,dxdy = \int_\Delta f(\varrho\cos(\varphi), \varrho\sin(\varphi))\varrho\,d\varrho\,d\varphi$$

Wyznaczanie całek wielokrotnych. Przy wyznaczaniu całek wielokrotnych posługujemy się ogólnym twierdzeniem, że jeżeli funkcja f jest funkcją mierzalną oraz $f \geq 0$ lub jedna z całek iterowanych funkcji $|f|$ jest skończona, to wolno zamieniać porządek całkowania.

Wyznaczyć całki:

1. $\int\limits_D (x+y)\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ 0 \leq y \leq x,\ x \leq 1\}$

2. $\int\limits_D \exp(x/y)\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ x^2 \leq y \leq 1,\ x > 0\}$

3. $\int\limits_D \frac{1}{\sqrt{x^2+y^2}}\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ 1 \leq x^2+y^2 \leq 4\}$

4. $\int\limits_D dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ \arcsin(x) < y < \pi/2,\ x > 0\}$

5. $\int\limits_D x^2\sqrt{x^2+y^2}\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ x, y \geq 0,\ x^2+y^2 \leq 4\}$

6. $\int\limits_D \ln(x^2+y^2)\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ e^2 \leq x^2+y^2 \leq e^4\}$

7. $\int\limits_D e^{-x^2}\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ 0 < y < x\}$

8. $\int\limits_D (x^2+xy^2)\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ 1 < x < 2,\ 0 < y < 3\}$

9. $\int\limits_D \sin(x+y)\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ -\frac{\pi}{4} < x < \frac{\pi}{4},\ 0 < y < \frac{\pi}{4}\}$

10. $\int\limits_D \frac{xy}{\sqrt{x^2+y^2+1}}\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ 0 < x < 1,\ 0 < y < 1\}$

11. $\int\limits_D (x^2-xy)\,dxdy$, gdzie $D = \{(x,y) \in \mathbb{R}^2 :\ y \geq x,\ y \leq -x^2+3x\}$

12. $\int\limits_D xy\,dxdy$, gdzie D jest obszarem ograniczonym krzywymi: $xy = 1,\ |x-y| = 1$

13. $\int\limits_D y\,dxdy$, gdzie D jest obszarem ograniczonym krzywymi: $y = 2-y^2,\ y = -1,\ y = 1,\ x = 1-\sqrt{1-y^2}$

Hesjan. **Hesjanem** funkcji $f : R^n \to R$ w punkcie $\boldsymbol{x} = (x_1, \ldots, x_n)$ nazywamy macierz

$$H(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\boldsymbol{x}) & \ldots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\boldsymbol{x}) & \ldots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\boldsymbol{x}) \end{bmatrix}$$

Ekstrema funkcji wielu zmiennych. Niech funkcja $f : \mathbb{R}^n \to \mathbb{R}$ ma ciągłe pochodne cząstkowe rzędu drugiego oraz $(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x})) = (0, \ldots, 0)$. Jeżeli macierz $H(\boldsymbol{x})$ jest dodatnio określona, to funkcja f osiąga minimum lokalne w punkcie $\boldsymbol{x}$. Jeżeli macierz $H(\boldsymbol{x})$ jest ujemnie określona, to funkcja f osiąga maksimum lokalne w punkcie $\boldsymbol{x}$.

Ekstrema warunkowe. Warunkiem koniecznym dla istnienia ekstremum funkcji $f : \mathbb{R}^n \to \mathbb{R}$ przy ograniczeniach

$$\varphi_i(x_1, x_2, \ldots, x_n) = 0, \quad i = 1, \ldots, k$$

jest, aby funkcja Lagrange'a $L : \mathbb{R}^{n+k} \to \mathbb{R}$ określona równaniem

$$L(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_k) = f(x_1, x_2, \ldots, x_n) - \lambda_1 \varphi(x_1) - \lambda_2 \varphi(x_k) - \ldots - \lambda_k \varphi(x_k)$$

miała ciągłe pochodne cząstkowe oraz spełniała układ równań:

$$\begin{cases} \frac{\partial L}{\partial x_i}(x_1, x_2, \ldots, x_n, \lambda_1, \lambda_2, \ldots, \lambda_k) = 0 & \text{dla } i = 1, \ldots, n; \\ \varphi_i(x_1, x_2, \ldots, x_n) = 0 & \text{dla } i = 1, \ldots, k. \end{cases}$$

Liczby $\lambda_1, \ldots, \lambda_k$ nazywane są **mnożnikami Lagrange'a**.

Wyznaczyć ekstrema lokalne funkcji f:

1. $f(x, y) = (x - y)^2 + (y + 2)^3$

2. $f(x, y) = xy \ln(x^2 + y^2)$

3. $f(x, y) = e^{-x^2 - y^2}$

4. $f(x, y) = x^2 + xy + y^2 - 4 \ln(x) - 10 \ln(y)$

5. $f(x, y) = (x + y^2) e^{\frac{x}{2}}$

6. $f(x, y) = x^4 - y^4$

7. $f(x, y) = x^3 + y^3 - 9xy + 12$

8. $f(x, y) = 4xy(1 - x - y) + 7$

9. $f(x, y) = xy + 1/x + 1/y$

10. $f(x, y) = x^3 - 2y^3 - 3x + 6y$

11. $f(x,y) = 8x^2 + 3y^2 - 4xy$ przy ograniczeniu $x^2 + y^2 = 1$

12. $f(x,y) = x^2 + y^2$ przy ograniczeniu $x^3 + y^3 - 8xy = 0$

13. $f(x,y) = \exp((x-5)^2 + (y-5)^2)$ przy ograniczeniu $\int\limits_x^y t^{10} \ln(t)\, dt = 10$

Funkcja Gamma. Funkcja Gamma określona jest dla $x > 0$ wzorem

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

Najważniejsze własności funkcji Gamma:

1. $\Gamma(x+1) = x\Gamma(x)$

2. $\Gamma(n) = (n-1)!$ dla naturalnych n

3. $\Gamma(1/2) = \sqrt{\pi}$

Funkcja Beta. Funkcja Beta określona jest dla $x, y > 0$ wzorem

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt$$

Najważniejsze własności funkcji Beta:

1. $B(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$

2. $n\binom{n+m-1}{n} B(n,m) = 1$ dla naturalnych n, m

Cześć całkowita. Częścią całkowitą $\lfloor x \rfloor$ liczby x nazywamy największą liczbę całkowitą n taką, że $n \leq x$.

Cześć ułamkowa. Częścią ułamkową $\lceil x \rceil$ liczby x nazywamy $x - \lfloor x \rfloor$.

O małe, o duże, $\sim$. Symbole $O(\cdot)$, $o(\cdot)$, $\sim$ oznaczają sposób wzajemnego zachownia się dwóch funkcji $f(x)$ oraz $g(x)$, gdy argument x zmierza do pewnej (niekoniecznie skończonej) granicy a. Zapis $f(x) = O(g(x))$, $x \to a$, oznacza, że $|f(x)/g(x)|$ pozostaje ograniczone, gdy $x \to a$. Zapis $f(x) = o(g(x))$, $x \to a$, oznacza, że $\lim\limits_{x \to a} \frac{f(x)}{g(x)} = 0$. Zapis $f(x) \sim g(x)$, $x \to a$, oznacza, że $\lim\limits_{x \to a} \frac{f(x)}{g(x)} = 1$

Nierówność Schwarza. Niech $x_1, \ldots, x_n \in \mathbb{R}$ oraz $y_1, \ldots, y_n \in \mathbb{R}$. Wówczas

$$\left(\sum_{i=1}^n x_i y_i \right)^2 \leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2$$

Niech f i g będą funkcjami całkowalnymi. Wówczas

$$\left(\int f(x) g(x) dx \right)^2 \leq \left(\int f^2(x) dx \right) \left(\int g^2(x) dx \right).$$

Najważniejsze pojęcia algebry liniowej.

Niech $\boldsymbol{A}$ będzie macierzą o m wierszach i n kolumnach:

$$\boldsymbol{A} = [a_{ij}]_{\substack{i=1,\dots,m\\ j=1,\dots,n}} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}.$$

Macierz o jednej kolumnie (tzn. $n = 1$) nazywamy **wektorem**.

Macierz $m \times n$ złożoną z samych zer nazywamy **zerową** i oznaczamy przez $\mathbf{0}_{m\times n}$.

Macierz $\boldsymbol{A}$ nazywamy **kwadratową stopnia n**, jeżeli $m = n$.

Macierz kwadratową $[a_{ij}]$ stopnia n taką, że $a_{ii} = 1$ dla wszystkich i oraz $a_{ij} = 0$ dla $i \neq j$ nazywamy **jednostkową** i oznaczamy przez $\boldsymbol{I}_n$.

Wektor złożony z n jedynek oznaczamy przez $\mathbf{1}_n$.

Macierz $m\times n$ złożoną z samych jedynek oznaczamy przez $\boldsymbol{J}_{m\times n}$. Jak łatwo sprawdzić: $\boldsymbol{J}_{m\times n} = \mathbf{1}_m \mathbf{1}'_n$.

Transpozycją macierzy $\boldsymbol{A} = [a_{ij}]_{\substack{i=1,\dots,m\\ j=1,\dots,n}}$ nazywamy macierz $\boldsymbol{A}^T = [a'_{ij}]_{\substack{i=1,\dots,n\\ j=1,\dots,m}}$ taką, że $a'_{ij} = a_{ji}$.

Długością wektora $\boldsymbol{a} = (a_1, a_2, \dots, a_n)^T$ nazywamy liczbę

$$\|\boldsymbol{a}\| = \sqrt{\boldsymbol{a}^T\boldsymbol{a}} = \sqrt{\sum_{i=1}^{n} a_i^2}.$$

Wektory $\boldsymbol{a}$ oraz $\boldsymbol{b}$ nazywamy **liniowo niezależnymi**, jeżeli jedynym rozwiązaniem równania $\alpha\boldsymbol{a} + \beta\boldsymbol{b} = \mathbf{0}$ jest $\alpha = \beta = 0$.

Rzędem macierzy $\boldsymbol{A}$ (rz$\boldsymbol{A}$) nazywamy liczbę liniowo niezależnych kolumn tej macierzy. Macierz $\boldsymbol{A}$ jest pełnego rzędu, jeżeli rz$\boldsymbol{A} = \min\{m, n\}$.

Obrazem macierzy $\boldsymbol{A}$ (Im$\boldsymbol{A}$) nazywamy podprzestrzeń liniową $\{\boldsymbol{A}\boldsymbol{a} : \boldsymbol{a} \in \mathbb{R}^n\}$ przestrzeni $\mathbb{R}^m$.

Wektory $\boldsymbol{a}$ i $\boldsymbol{b}$ nazywamy **ortogonalnymi** (prostopadłymi), jeżeli $\boldsymbol{a}^T\boldsymbol{b} = 0$.

Dopełnieniem algebraicznym elementu a_{ij} macierzy $\boldsymbol{A}$ nazywamy wyznacznik A_{ij} macierzy powstałej z macierzy $\boldsymbol{A}$ przez skreślenie i-tego wiersza oraz j-tej kolumny.

Niech $\boldsymbol{A}$ będzie macierzą kwadratową. **Wyznacznikiem macierzy $\boldsymbol{A}$** nazywamy liczbę $\det \boldsymbol{A}$ określoną w następujący rekurencyjny sposób:

$$\det[a_{11}] = a_{11}, \; \det \boldsymbol{A} = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} A_{i1},$$

gdzie A_{i1} jest dopełnieniem algebraicznym elementu a_{i1}.

Wyznacznik macierzy można obliczać także z następującego wzoru:

$$\det \boldsymbol{A} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} A_{ij}, \text{ dla dowolnego } 1 \le j \le n,$$

lub

$$\det \boldsymbol{A} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} A_{ij}, \text{ dla dowolnego } 1 \le i \le n.$$

W szczególności

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Niech $\boldsymbol{A}$ będzie macierzą kwadratową stopnia n pełnego rzędu ($\mathrm{rz}\boldsymbol{A} = n$). **Macierzą odwrotną** do macierzy $\boldsymbol{A}$ nazywamy taką macierz $\boldsymbol{A}^{-1}$, że $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$.

Jeżeli $\det \boldsymbol{A} \neq 0$, to $\boldsymbol{A}^{-1} = [(-1)^{i+j} A_{ij} / \det \boldsymbol{A}]_{i,j=1,\ldots,n}$. W szczególności

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \begin{bmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{bmatrix} / \det \boldsymbol{A} \;.$$

Jeżeli $\boldsymbol{A}$ i $\boldsymbol{D}$ są symetrycznymi macierzami takimi, że odpowiednie odwrotności istnieją, to

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^T & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{F}\boldsymbol{E}^{-1}\boldsymbol{F}^T & -\boldsymbol{F}\boldsymbol{E}^{-1} \\ -\boldsymbol{E}^{-1}\boldsymbol{F}^T & \boldsymbol{E}^{-1} \end{bmatrix},$$

gdzie $\boldsymbol{E} = \boldsymbol{D} - \boldsymbol{B}^T\boldsymbol{A}^{-1}\boldsymbol{B}$ oraz $\boldsymbol{F} = \boldsymbol{A}^{-1}\boldsymbol{B}$.

Niech $\boldsymbol{A}$ będzie macierzą nieosobliwą i niech $\boldsymbol{u}$ oraz $\boldsymbol{v}$ będą dwoma wektorami kolumnowymi. Wówczas

$$(\boldsymbol{A} + \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \boldsymbol{A}^{-1} - \frac{(\boldsymbol{A}^{-1}\boldsymbol{u})(\boldsymbol{v}^T\boldsymbol{A}^{-1})}{1 + \boldsymbol{v}^T\boldsymbol{A}^{-1}\boldsymbol{u}}.$$

Macierz symetryczną $\boldsymbol{A}$ ($\boldsymbol{A}^T = \boldsymbol{A}$) nazywamy **dodatnio określoną**, jeżeli dla dowolnego wektora $\boldsymbol{a} \neq 0$ zachodzi $\boldsymbol{a}^T\boldsymbol{A}\boldsymbol{a} > 0$. Na mocy **kryterium Sylvestera** macierz symetryczna $\boldsymbol{A} = [a_{ij}]_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$ jest dodatnio określona wtedy i tylko wtedy, gdy wszystkie jej wiodące minory główne są dodatnie, tzn.

$$det \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & \vdots & \ddots & \\ a_{k1} & a_{k2} & \ldots & a_{kk} \end{bmatrix} > 0, \text{ dla } k = 1, \ldots, n$$

Macierz symetryczną $\boldsymbol{A}$ ($\boldsymbol{A}^T = \boldsymbol{A}$) nazywamy **ujemnie określoną**, jeżeli macierz $-\boldsymbol{A}$ jest dodatnio określona.

1. Elementy rachunku prawdopodobieństwa

Przestrzeń zdarzeń elementarnych. Pojęciem pierwotnym w rachunku prawdopodobieństwa jest **przestrzeń zdarzeń elementarnych** Ω.

σ-ciało. Rodzinę $\mathcal{F}$ spełniającą warunki

1. $\mathcal{F} \neq \emptyset$

2. Jeśli $A \in \mathcal{F}$, to $\Omega \setminus A \in \mathcal{F}$

3. Jeśli $A_i \in \mathcal{F}$ dla $i = 1, 2, \ldots$, to $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$

nazywamy σ-ciałem podzbiorów zbioru Ω

Zdarzenie losowe. Elementy rodziny $\mathcal{F}$ nazywamy **zdarzeniami losowymi**.

Prawdopodobieństwo. Miarą prawdopodobieństwa nazywamy dowolną funkcję $P : \mathcal{F} \to \mathbb{R}$ taką, że

1. $(\forall A \in \mathcal{F})\ P(A) \geq 0$

2. $P(\Omega) = 1$

3. Jeśli $(A_i)_{i=1,2,\ldots} \in \mathcal{F}$, $A_i \cap A_j = \emptyset$ dla $i \neq j$, to

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

Przestrzeń probabilistyczna. **Przestrzenią probabilistyczną** nazywamy trójkę

$$(\Omega, \mathcal{F}, P)$$

Własności prawdopodobieństwa. Niech $A,\ B,\ A_1, \ldots, A_n \in \mathcal{F}$. Wówczas

1. $P(\emptyset) = 0$

2. jeżeli $A_i \cap A_j = \emptyset$ dla $i \neq j$, to

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

3. $P(A') = 1 - P(A)$, gdzie $A' = \Omega \setminus A$

4. Jeśli $A \subseteq B$, to $P(B \setminus A) = P(B) - P(A)$

5. Jeśli $A \subseteq B$, to $P(A) \leq P(B)$

6. $P(A) \leq 1$

7. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Prawdopodobieństwo warunkowe. Niech B będzie takim zdarzeniem, że $P(B) > 0$. **Prawdopodobieństwem warunkowym** zajścia zdarzenia A pod warunkiem zajścia zdarzenia B nazywamy liczbę

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Jeśli $P(A_1 \cap \ldots \cap A_{n-1}) > 0$, to

$$P(A_1 \cap \ldots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \ldots \cap A_{n-1}).$$

Prawdopodobieństwo całkowite. **Rozbiciem** przestrzeni Ω nazywamy rodzinę zdarzeń $\{H_i\}_{i=1,\ldots,n}$, które wzajemnie wykluczają się, zaś ich suma jest równa Ω.

Twierdzenie 1.1. *Niech $\{H_i\}_{i=1,\ldots,n}$ będzie rozbiciem przestrzeni Ω na zdarzenia o dodatnich prawdopodobieństwach. Wówczas dla dowolnego zdarzenia A zachodzi*

$$P(A) = \sum_{i=1}^{n} P(A|H_i)P(H_i)$$

Wzór Bayesa. Niech $\{H_i\}_{i=1,\ldots,n}$ będzie rozbiciem Ω na zdarzenia o dodatnich prawdopodobieństwach i $P(A) > 0$. Wówczas dla dowolnego $j \in \{1, \ldots, n\}$ mamy

$$P(H_j|A) = \frac{P(A|H_j)P(H_j)}{\sum_{i=1}^{n} P(A|H_i)P(H_i)}$$

Niezależność zdarzeń. Zdarzenia A oraz B nazywamy **niezależnymi**, gdy

$$P(B|A) = P(B), \quad P(A) > 0.$$

Zdarzenia A oraz B są niezależne, gdy

$$P(A \cap B) = P(A)P(B).$$

Zdarzenia $A_1, \ldots, A_n$ nazywamy niezależnymi, gdy

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k})$$

dla $1 \leq i_i < i_2 < \cdots < i_k \leq n, \; k = 2, 3, \ldots, n$.

Zadania do samodzielnego rozwiązania

1.1. Tarcza strzelecka składa się z trzech koncentrycznych kół o promieniach odpowiednio 1, 2 i 3. Za trafienie w środkowe koło zdobywa się trzy punkty, za trafienie w kolejne pierścienie (licząc od środka koła) odpowiednio dwa i jeden punkt. Jakie jest prawdopodobieństwo uzyskania co najmniej trzech punktów w dwóch strzałach? (Zakładamy, że każdy strzał trafia w tarczę.)

1.2. W grupie studenckiej jest 20 osób. Na ćwiczeniach Student do odpowiedzi losowany jest na podstawie wyniku rzutu kostką dwudziestościenną. Jakie jest prawdopodobieństwo, że ten sam Student zostanie wyrwany do odpowiedzi trzykrotnie z rzędu?

1.3. Autobus przyjeżdża na przystanek co piętnaście minut. Jakie jest prawdopodobieństwo tego, że przychodząc na przystanek w losowym momencie będziemy czekać na autobus nie dłużej niż pięć minut?

1.4. Pan Roztargniony zapomniał ostatniej cyfry telefonu do znajomego. W związku z tym wykręcając numer telefonu ostatnią cyfrę wybiera losowo. Jakie jest prawdopodobieństwo tego, że dodzwoni się, jeżeli ma do dyspozycji cztery żetony telefoniczne?

1.5. Na egzamin przygotowanych jest 100 pytań. Student zna odpowiedź na 80 z nich. Egzaminator przerywa egzamin w chwili, gdy Student nie umie odpowiedzieć na pytanie, lecz nie później niż po piątym pytaniu. Ocena końcowa jest równa liczbie pytań, na które odpowiedział Student. Jakie jest prawdopodobieństwo tego, że Student otrzyma ocenę co najmniej dobrą?

1.6. Rzucono trzema kostkami. Jakie jest prawdopodobieństwo, że przynajmniej na jednej kostce wypadnie jedynka, jeżeli na każdej kostce wypadnie inna liczba oczek?

1.7. Z talii 52 kart wyciągnięto losowo jedną kartę. Jakie jest prawdopodobieństwo, że jest to siódemka, jeżeli wiadomo, że wyciągnięta karta nie jest ani figurą ani asem?

1.8. Z talii 52 kart wyciągamy losowo jedną kartę. Rozpatrzmy zdarzenia:
A - wyciągnęliśmy asa,
B - wyciągnęliśmy kartę koloru czerwonego,
C - wyciągnęliśmy asa karo,
D - wyciągnęliśmy dziewiątkę.
Które z par zdarzeń są niezależne?

1.9. Rzucamy czterokrotnie symetryczną monetą. Obliczyć prawdopodobieństwo uzyskania $0, 1, 2, 3$ oraz 4 orłów. Dane są następujące zdarzenia:
A - wypadły cztery orły,
B - wypadła parzysta liczba orłów,
C - wypadło więcej orłów niż reszek.
Obliczyć prawdopodobieństwa następujących zdarzeń:

$$P(A), P(B), P(C), P(A|B), P(B|A), P(A|C), P(C|A), P(B|C), P(C|B).$$

1.10. Troje dzieci: Ania, Basia i Czesio zmywają szklanki. Najstarsza Ania zmywa dwa razy częściej niż młodsza Basia, zaś Basia trzy razy częściej niż najmłodszy Czesio. Wiadomo, że prawdopodobieństwo zbicia szklanki w czasie zmywania wynosi dla Ani 0.01, dla Basi wynosi 0.04 natomiast dla Czesia 0.5. Jakie jest prawdopodobieństwo, że w czasie zmywania zostanie zbita jedna szklanka?

Pewnego dnia po powrocie z pracy mama zauważyła, że jedna ze szklanek jest zbita, a żadne z dzieci nie chciało się przyznać do zniszczenia szklanki. Które z dzieci najprawdopodobniej zmywało tego dnia?

1.11. Przedsiębiorstwo zawarło umowy z zakładami Z_1, Z_2 oraz Z_3 na dostawę podzespołów. Zakład Z_1 dostarcza 50%, zakład Z_2 dostarcza 35% natomiast zakład Z_3 dostarcza 15% potrzebnych podzespołów. Wiadomo, że 95% dostaw zakładu Z_1, 80% dostaw zakładu Z_2 oraz 85% dostaw zakładu Z_3 odpowiada wymaganiom technicznym. Jakie jest prawdopodobieństwo, że jeden wylosowany podzespół odpowiada wymaganiom technicznym?

Do punktu serwisowego zgłasza się klient z urządzeniem, w którym uszkodzony jest podzespół. Jakie jest prawdopodobieństwo, że producentem zepsutego podzespołu był zakład Z_1?

1.12. Na wspólnej klasówce z matematyki spotkali się Studenci I roku z dwóch grup. W pierwszej grupie jest 15 pań oraz 10 panów, zaś w drugiej jest 12 panów i 13 pań. Prawdopodobieństwo, że pani z grupy pierwszej rozwiąże zadanie na klasówce wynosi 0.8, natomiast prawdopodobieństwo to dla pana wynosi 0.7. W drugiej grupie prawdopodobieństwa te kształtują się odpowiednio 0.9 oraz 0.85. Jak duży odsetek wszystkich Studentów rozwiąże zadanie na klasówce? Przy sprawdzaniu prac okazało się, że ktoś przygotował ściągawkę. Określić, kim najprawdopodobniej był autor ściągawki (tzn. określić płeć i grupę autora).

1.13. Wśród 300 zdających egzamin wstępny z matematyki jest 200 absolwentów klas matematyczno-fizycznych, 75 absolwentów klas ogólnokształcących oraz 25 absolwentów klas humanistycznych. Prawdopodobieństwo zdania egzaminu przez absolwenta klasy matematyczno-fizycznej wynosi 0.9, klasy ogólnokształcącej wynosi 0.25, zaś klasy humanistycznej 0.1. Jakie jest prawdopodobieństwo, że losowo wybrany przystępujący do egzaminu zda go pomyślnie? Jaki jest odsetek absolwentów klas matematyczno-fizycznych, klas humanistycznych oraz klas ogólnokształcących wśród osób, które zdały egzamin?

1.14. Sklep jest zaopatrywany w żarówki pochodzące z trzech fabryk, przy czym 20% żarówek pochodzi z pierwszej fabryki, 30% z drugiej, 50% z trzeciej. Produkcja pierwszej fabryki zawiera 1% żarówek wadliwych, produkcja drugiej – 5% żarówek wadliwych, a produkcja trzeciej fabryki 10% żarówek wadliwych. Obliczyć prawdopodobieństwo, że losowo wybrana żarówka będzie wadliwa.

1.15. W szpitalu na oddziale wewnętrznym przebywa rocznie średnio 2000 chorych. Wśród leczonych było 800 cierpiących na chorobę K_1, 600 cierpiących na chorobę K_2, 400 cierpiących na chorobę K_3 oraz 200 cierpiących na chorobę K_4. Prawdopodobieństwo pełnego wyleczenia z chorób wynosiło odpowiednio 0.9, 0.8, 0.7 oraz 0.5. Obliczyć prawdopodobieństwo, że losowo wypisany pacjent jest całkowicie wyleczony.

1.16. W magazynie znajdują się opony do samochodów osobowych. Pochodzą one w 20% z fabryki F_1, w 30% z fabryki F_2 i w 50% z fabryki F_3. Wiadomo, że produkcja wadliwych opon w poszczególnych fabrykach wynosi odpowiednio 5%, 4%, 3%. Obliczyć prawdopodobieństwo, że losowo wybrana opona będzie wadliwa.

1.17. W fabryce trzech robotników wykonuje te same detale, przy czym 30% detali wykonuje robotnik I, 45% – robotnik II, a resztę robotnik III. Robotnicy wytwarzają odpowiednio 2%, 1% oraz 3% braków. Jakie jest prawdopodobieństwo, że losowo wybrany detal okaże się dobry?

1.18. Mamy dwie urny z kulami. W urnie nr 1 są trzy białe i dwie czarne kule, w urnie nr 2 są cztery białe i cztery czarne. Z urny nr 1 przekładamy losowo dwie kule do urny nr 2, a następnie wyciągamy jedną kulę z urny nr 2. Jakie prawdopodobieństwo, że będzie to kula biała?

1.19. Mamy dwie urny z kulami. W każdej z nich jest pięć czarnych, dziesięć czerwonych i sześć białych kul. Jakie jest prawdopodobieństwo, że ciągnąc po jednej kuli z każdej urny wylosujemy kule tego samego koloru?

1.20. Spośród 64 pól szachownicy wybieramy losowo dwa różne pola i ustawiamy na nich dwie jednakowe figury: białą i czarną. Obliczyć prawdopodobieństwo, że figury te nie będą zagrażać sobie wzajemnie, jeżeli ustawiono

a. dwa hetmany;

b. dwa gońce;

c. dwa skoczki;

d. dwie wieże.

1.21. Z urny zawierającej trzy kule białe i dwie kule czarne przełożono dwie wyciągnięte losowo kule do urny zawierającej cztery białe i cztery czarne kule. Obliczyć prawdopodobieństwo wyciągnięcia białej kuli z drugiej urny.

1.22. W trzech urnach znajdują białe i czarne kule. W pierwszej z nich sa dwie kule białe i trzy czarne, w drugiej dwie białe i dwie czarne, a w trzeciej trzy białe i jedna czarna. Przekładamy wylosowaną kulę z pierwszej urny do drugiej, a następnie losowo wybraną kulę z drugiej urny do trzeciej. Wreszcie wybraną losowo kulę przekładamy z urny trzeciej do pierwszej.

a. Jaki skład pierwszej urny jest najbardziej prawdopodobny?

b. Obliczyć prawdopodobieństwo, że skład wszystkich trzech urn pozostanie niezmieniony.

1.23. Pewien Student nie zna odpowiedzi na niektóre z pytań na kartkach egzaminacyjnych. W jakim przypadku szansa wyciągnięcia przez niego kartki z pytaniem, na które nie zna odpowiedzi, będzie najmniejsza: jeżeli losuje pierwszy, czy jeżeli losuje ostatni?

1.24. Prawdopodobieństwo, że wyroby pewnej fabryki spełniają wymagane normy wynosi 0.96. Zakładamy uproszczony system sprawdzania, który daje rezultat dodatni z prawdopodobieństwem 0.98 dla sztuk dobrych i z prawdopodobieństwem 0.05 dla sztuk wadliwych. Jakie jest prawdopodobieństwo, że sztuka uznana za dobrą przez kontrolę rzeczywiście spełnia wymagania normy?

1.25. Załóżmy, że prawdopodobieństwo trafienia w cel przy pojedynczym strzale wynosi p, a prawdopodobieństwo zniszczenia celu przy $k \geq 1$ trafieniach wynosi $1 - q^k$. Jakie jest prawdopodobieństwo zniszczenia celu, jeżeli oddano n strzałów?

1.26. W partii towaru złożonej z N sztuk znajduje się $M < N$ wadliwych. Wybrano losowo $n < N$ sztuk, które poddano pobieżnej kontroli. Kontrola ta może popełnić błędy: z prawdopodobieństwem p może się zdarzyć, że wadliwą sztukę uzna się za „dobrą", a z prawdopodobieństwem q dobrą sztukę uzna się za „wadliwą". Obliczyć prawdopodobieństwo, że m sztuk zostanie uznanych za wadliwe.

1.27. Niech X będzie zmienną losową przyjmującą wartości całkowite, przy czym $X = k$ z prawdopodobieństwem $\frac{\lambda^k}{k!}e^{-\lambda}$, gdzie $k = 0, 1, 2, \ldots$ Doświadczenie polega na tym, że na odcinek $[0, 1]$ rzucamy losowo X punktów. Oznaczmy przez X_i liczbę punktów, które znajdą się w odcinku $[(i-1)/n, i/n]$, gdzie $i = 1, 2, \ldots, n$. Udowodnić, że dla $\lambda = n$ zmienne losowe X_i są niezależne.

1.28. Przypuśćmy, że pewien owad składa k jajeczek z prawdopodobieństwem $\frac{\lambda^k}{k!}e^{-\lambda}$, a każde z jajeczek wylęga się z prawdopodobieństwem p. Zakładając wzajemną niezależność wylęgania się jaj, znaleźć prawdopodobieństwo, że ilość potomków danego owada wyniesie dokładnie l.

1.29. Z urny zawierającej $m \geq 3$ kul białych i n kul czarnych zgubiono jedną kulę nieznanego koloru. Aby określić skład urny wybrano z niej losowo trzy kule. Znaleźć prawdopodobieństwo, że zgubiona kula była biała, jeżeli wiadomo, że wszystkie wybrane kule są białe.

1.30. Prawdopodobieństwo wyłączających się i wyczerpujących wszystkie możliwości hipotez $A_1, A_2, \ldots, A_k$ przed nastąpieniem zdarzenia B wynoszą odpowiednio $\alpha_1, \alpha_2, \ldots, \alpha_k$, prawdopodobieństwa zdarzenia B odpowiadające danym hipotezom wynoszą $p_1, p_2, \ldots, p_k$. Wiadomo, że w n_1 niezależnych doświadczeniach zdarzenie B nastąpiło m_1 razy. Wiadomo też, że w następnej serii złożonej z n_2 doświadczeń zdarzenie B nastąpiło m_2 razy. Udowodnić, następującą własność wzoru Bayesa:

Prawdopodobieństwo *a posteriori* hipotez obliczone po drugiej serii doświadczeń przy przyjęciu jako prawdopodobieństwa *a priori* odpowiednich prawdopodobieństw *a posteriori* po pierwszej serii doświadczeń równe są zawsze prawdopodobieństwom *a posteriori* obliczonym po prostu dla serii $n_1 + n_2$ doświadczeń, w których zdarzenie B zaszło $m_1 + m_2$ razy.

1.31. Przez kanał łączności można przekazać jedną z następujących serii liter: $AAAA$, $BBBB$, $CCCC$, przy czym odpowiednie prawdopodobieństwa *a priori* wynoszą 0.3, 0.4, 0.3. Wiadomo, że działanie szumów zmniejsza prawdopodobieństwo poprawnego odebrania nadanej litery do 0.6, a prawdopodobieństwo odebrania każdej z dwóch innych liter wynosi po 0.2. Zakładamy, że litery przekazywane są niezależnie od siebie. Obliczyć prawdopodobieństwo, że przekazano ciąg $AAAA$, jeżeli odebrano ciąg $ABCA$.

1.32. Prawdopodobieństwo, że cząsteczka, która w chwili $t = 0$ zderzyła się z inną cząsteczką i do chwili t nie uległa żadnym innym zderzeniom, ulegnie zderzeniu w ciągu odcinka czasu od t do $t+\Delta t$, wynosi $\lambda\Delta t+o(\Delta t)$. Znaleźć prawdopodobieństwo, że czas wolnego przebiegu (tzn. okres między dwoma kolejnymi zderzeniami) będzie większy niż t.

1.33. Zakładamy, że przy rozmnażaniu się bakterii przez podział (na dwie bakterie) prawdopodobieństwo podziału w ciągu odcinka czasu o długości Δt wynosi $a\Delta t + o(\Delta t)$ i nie zależy od ilości bakterii w koloni ani od ilości poprzedzających podziałów. Obliczyć prawdopodobieństwo, że jeżeli w chwili $t = 0$ kolonia składa się z jednej bakterii, to w chwili t będzie ona zawierać i bakterii.

1.34. Załóżmy dodatkowo (do warunków poprzedniego zadania), że w ciągu odcinka czasu o długości Δt każda z bakterii niezależnie od ogólnej ilości bakterii może umrzeć z prawdopodobieństwem $b\Delta t+o(\Delta t)$. Wyprowadzić równanie różniczkowe dla funkcji $p_r(t)$, gdzie $p_r(t)$ oznacza prawdopodobieństwo, że w chwili t kolonia ma r bakterii.

1.35. Dwaj gracze A i B posiadający początkowo kapitały odpowiednio a i b grają w grę hazardową składającą się z oddzielnych partii. W każdej z partii pierwszy z graczy wygrywa z prawdopodobieństwem 0.5 i przegrywa z prawdopodobieństwem 0.5. Wypłata odbywa się po każdej partii i wynosi 1, przy czym gra toczy się tak długo, aż jeden z graczy nie zostanie zruinowany. Obliczyć prawdopodobieństwo ruiny drugiego gracza.

Rozwiązać to zadanie w przypadku, gdy pierwszy gracz wygrywa z prawdopodobieństwem $p > 0.5$ i przegrywa z prawdopodobieństwem $q = 1 - p$.

1.36. Znaleźć liczbę β o tej własności, żeby przy rzucaniu kostką do gry prawdopodobieństwo zdarzenia A polegającego na wyrzuceniu serii trzech kolejnych jedynek przed serią β kolejnych nie-jedynek było w przybliżeniu równe 0.5. (Wskazówka. Wprowadzić prawdopodobieństwa warunkowe u i v zdarzenia A przy warunku, że wynikiem pierwszego rzutu były odpowiednio jedynak i nie-jedynka. Używając wzoru na prawdopodobieństwo całkowite ułożyć równanie wiążące u i v.)

1.37. Rozpatrzmy ciąg niezależnych doświadczeń, z których każde może dać trzy możliwe wyniki A, B, C z odpowiednimi prawdopodobieństwami p, q r $(p+q+r = 1)$. Obliczyć prawdopodobieństwo, że

a. seria wyników A o długości α pojawi się wcześniej niż seria wyników B o długości β;

b. seria wyników A o długości α pojawi się wcześniej niż seria wyników B o długości β lub seria wyników C o długości γ.

1.38. Mamy dwie partie wyprodukowanych przedmiotów, przy czym wiadomo, że wszystkie przedmioty jednej partii odpowiadają wymaganiom technicznym, a 15% przedmiotów drugiej partii jest złej jakości. Przedmiot wzięty z losowo wybranej partii był dobrej jakości. Obliczyć prawdopodobieństwo, że drugi przedmiot wzięty z tej samej partii będzie dobrej jakości, jeżeli pierwszy przedmiot po sprawdzeniu zwrócono z powrotem do partii, z której pochodził?

1.39. Wiadomo, że 5% wszystkich mężczyzn i 0.25% wszystkich kobiet jest daltonistami. Wybrana losowo osoba okazała się daltonistą. Obliczyć prawdopodobieństwo, że następna osoba tej samej płci co pierwsza, też okaże się daltonistą.

1.40. Z badań genealogicznych wynika, że kobieta jest nośnikiem hemofilii z prawdopodobieństwem p. Jeżeli kobieta jest nośnikiem hemofilii, to każdy jej syn dziedziczy tę chorobę z prawdopodobieństwem 0.5. Kobieta, która nie jest nośnikiem hemofilii rodzi zdrowych synów. Obliczyć prawdopodobieństwo, że drugi syn będzie zdrowy, jeśli pierwszy syn jest zdrowy.

1.41. Pewien zakład produkuje 25% produktów pierwszej jakości, 12% drugiej jakości oraz 63% trzeciej jakości. Wylosowano dwa produkty i okazało się, że jeden z nich ma lepszą jakość niż drugi. Jakie jest prawdopodobieństwo, że produkt lepszej jakości jest produktem pierwszej jakości?

1.42. Pewien matematyk nosi przy sobie dwa pudełka zapałek. Ilekroć chce on zapalić papierosa, wydobywa jedno pudełko losowo z kieszeni. Znaleźć prawdopodobieństwo, że w chwili gdy po raz pierwszy wydobędzie on puste pudełko, drugie będzie zawierało r zapałek ($r = 0, 1, \ldots, n$; n jest tu liczbą zapałek, które na początku znajdowały się w każdym z pudełek).

1.43. Jest n osób: $A_1, A_2, \ldots, A_n$. Osoba A_1 dostaje kartkę ze znakiem $+$. Z prawdopodobieństwem p zmienia znak na przeciwny i podaje kartkę osobie A_2, która z prawdopodobieństwem p zmienia go na przeciwny i podaje kartkę osobie A_3, itd. Na zakończenie, po oddaniu kartki przez osobę A_n, zaobserwowano znak $+$. Jakie jest prawdopodobieństwo, że osoba A_1 nie zmieniła znaku?

1.44. W fabryce produkuje się dwa rodzaje śrubek. Wylosowano dwie śrubki. Wiadomo, że prawdopodobieństwo wylosowania dwóch identycznych śrubek wynosi p, a wylosowania śrubki pierwszego rodzaju wynosi q. Wyznaczyć prawdopodobieństwo, że jedna ze śrubek jest pierwszego rodzaju, jeżeli wiadomo, że druga też jest pierwszego rodzaju.

1.45. Grześ i Jaś rzucają na przemian monetą. Jaś wygrywa, gdy pojawią się kolejno OOR, Grześ – gdy ROR. Jakie są prawdopodobieństwa wygranych dla obu chłopców?

1.46. Na rynku są odkurzacze trzech firm, przy czym firma A ma 50% udziału w rynku, firma B – 30%, zaś firma C – 20%. Wadliwość silników w odkurzaczach poszczególnych firm wynosi odpowiednio 1%, 2% oraz 5%. Jaka powinna być struktura procentowa silników w magazynie punktu serwisowego, by można było bezzwłocznie przystąpić do wymiany zepsutego silnika w odkurzaczu przyniesionym do tego punktu?

1.47. Wśród kandydatów na studia jest 50% absolwentów klas matematyczno-fizycznych, 30% absolwentów klas biologiczno-chemicznych oraz 20% absolwentów klas humanistycznych. Szanse na zdanie egzaminu z matematyki dla poszczególnych grup absolwentów wynoszą odpowiednio 90%, 80% oraz 50%. Jakiej struktury procentowej absolwentów można się spodziewać wśród studentów pierwszego roku?

1.48. Na podstawie tabel życia wiadomo, że 89.835% kobiet dożywa wieku 60 lat, a 57.062% do wieku 80 lat. Jakie jest prawdopodobieństwo, że Pani Ala, która ma 60 lat przeżyje jeszcze lat dwadzieścia?

1.49. Pewien uczeń miał 50% szans, że zda na Uniwersytet oraz 30% szans, że zda do SGGW. Uczeń ten zdawał na obie uczelnie, przy czym wiadomo było, że ma 20% szans, że zda na obie uczelnie. Jakie są szanse, że dostał się na studia?

1.50. Gra polega na wykonaniu dwóch kolejnych rzutów sześcienną kostką do gry, przy czym za udział w grze należy wnieść opłatę w wysokości siedmiu paciorków. Wygrana jest równa sumie wyrzuconych oczek. Jaka jest oczekiwana wygrana w tej grze? Czy wiadomość o tym, że w pierwszym rzucie uzyskano trzy oczka zmienia szansę na wygraną?

1.51. Gra polega na wykonaniu dwóch kolejnych rzutów sześcienną kostką do gry. Wygrana jest równa różnicy oczek wyrzuconych w pierwszym oraz drugim rzucie. Jaka jest oczekiwana wygrana w tej grze? Czy wiadomość o tym, że w pierwszym rzucie uzyskano trzy oczka zmienia szansę na wygraną?

2. Zmienne losowe

Zmienna losowa. **Zmienna losowa** jest to funkcja rzeczywista $\xi : \Omega \to \mathbb{R}$ o własności:

$$(\forall x \in \mathbb{R})\, \{\omega \in \Omega :\ \xi(\omega) \le x\} \in \mathcal{F}$$

Rozkład zmiennej losowej. **Rozkładem zmiennej losowej** ξ nazywamy rozkład prawdopodobieństwa P_ξ na $\mathbb{R}$ określony wzorem

$$P_\xi(A) = P(\{\omega \in \Omega :\ \xi(\omega) \in A\}) = P(\xi^{-1}(A)) \quad \text{dla } A \subset \mathbb{R}$$

Dystrybuanta. **Dystrybuantą** zmiennej losowej ξ nazywamy funkcję $F : \mathbb{R} \to [0, 1]$ określona wzorem

$$F_\xi(x) = P(\xi \le x)$$

Najważniejsze własności dystrybuanty:

1. $F(-\infty) = 0,\ F(\infty) = 1$

2. dystrybuanta jest funkcją niemalejącą

3. dystrybuanta jest funkcją prawostronnie ciągłą

Ponadto

1. $P\{a < \xi \le b\} = F(b) - F(a)$

2. $P\{a \le \xi \le b\} = F(b) - \lim\limits_{x \to a-} F(x)$

3. $P(a < \xi < b) = \lim\limits_{x \to b-} F(x) - F(a)$

4. $P(\xi = a) = F(a) - \lim\limits_{x \to a-} F(x)$

Zmienna losowa skokowa (dyskretna). Zmienna losowa ξ jest typu skokowego (dyskretna), jeżeli istnieje zbiór skończony lub przeliczalny $\mathbb{X} \subset \mathbb{R}$ taki, że

$$P_\xi(\mathbb{X}) = 1$$

Funkcja rozkładu prawdopodobieństwa. Funkcją rozkładu prawdopodobieństwa nazywamy

$$f(x) = \begin{cases} P\{\xi = x\}, & \text{dla } x \in \mathbb{X}, \\ 0, & \text{dla } x \notin \mathbb{X}, \end{cases}$$

Zmienna losowa ciągła. Zmienna losowa ξ o dystrybuancie F jest typu ciągłego, jeżeli istnieje taka funkcja $f \ge 0$, że dla każdego x zachodzi równość

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

Funkcję f nazywamy gęstością prawdopodobieństwa zmiennej losowej ξ lub w skrócie **gęstością**.

Najważniejsze własności funkcji gęstości:

1. $\int_{-\infty}^{\infty} f(x)dx = 1$

2. $f(x) = \begin{cases} F'(x), & \text{jeżeli } F'(x) \text{ istnieje,} \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$

3. $P\{a < \xi \le b\} = \int_a^b f(x)dx$

Momenty zmiennej losowej. Momentem rzędu k zmiennej losowej ξ nazywamy wielkość

$$E\xi^k = \int_{\mathbb{R}} x^k F(dx)$$

$$E\xi^k = \begin{cases} \sum_{x \in \mathbb{X}} x^k P\{\xi = x\}, & \text{dla dyskretnej zmiennej losowej,} \\ \int_{\mathbb{R}} x^k f(x)dx, & \text{dla ciągłej zmiennej losowej.} \end{cases}$$

Wartość oczekiwana. Wartością oczekiwaną $E\xi$ nazywamy moment rzędu 1.

Momenty centralne zmiennej losowej. Momentem centralnym rzędu k zmiennej losowej ξ nazywamy wielkość

$$E(\xi - E\xi)^k = \int_{\mathbb{R}} (x - E\xi)^k F(dx)$$

$$E(\xi - E\xi)^k = \begin{cases} \sum_{x \in \mathbb{X}} (x - E\xi)^k P\{\xi = x\}, & \text{dla dyskretnej zmiennej losowej,} \\ \int_{\mathbb{R}} (x - E\xi)^k f(x)dx, & \text{dla ciągłej zmiennej losowej.} \end{cases}$$

Wariancja. Wariancją $D^2\xi$ nazywamy moment centralny rzędu 2.

Jak łatwo sprawdzić:

$$D^2\xi = E\xi^2 - (E\xi)^2.$$

Odchylenie standardowe. Odchyleniem standardowym nazywamy $D\xi = \sqrt{D^2\xi}$

Kwantyl. Kwantylem rzędu p zmiennej losowej ξ nazywamy taką liczbę x_p, że

$$F(x_p) = p.$$

Kwantyl rzędu 0.5 nazywamy **medianą**.

Frakcja. Jeżeli A jest danym podzbiorem zbioru wartości zmiennej losowej ξ, to **frakcją** nazywamy liczbę

$$p = P\{\xi \in A\}.$$

Moda. **Modą** zmiennej losowej ξ nazywamy „najbardziej” prawdopodobną wartość tej zmiennej losowej.

W przypadku zmiennej losowej typu skokowego jest to wartość zmiennej losowej, której odpowiada największe prawdopodobieństwo, natomiast w przypadku zmiennej losowej typu ciągłego - wartość, dla której funkcja gęstości rozkładu prawdopodobieństwa przyjmuje maksimum lokalne.

Funkcje zmiennych losowych. Niech $h : \mathbb{R} \to \mathbb{R}$ i niech $\eta = h(\xi)$. Wówczas

$$F_\eta(x) = P\{h(\xi) \le x\} = P\{\xi \in h^{-1}(-\infty, x\rangle\}$$

Twierdzenie 2.1. *Niech ξ będzie zmienną losową typu ciągłego. Niech h będzie funkcją określoną na zbiorze*

$$\bigcup_{k=1}^{n} [a_k, b_k],$$

która na każdym przedziale otwartym (a_k, b_k) jest funkcją ściśle monotoniczną oraz ma ciągłą pochodną $h'(x) \neq 0$. Niech $g_k(y)$ będzie funkcją odwrotną do funkcji $h(x)$ na przedziale

$$I_k = h((a_k, b_k)) = \{y : \ x \in (a_k, b_k), h(x) = y\}.$$

Wówczas funkcja gęstości zmiennej losowej $\eta = h(\xi)$ ma postać

$$f_\eta(y) = \sum_{k=1}^{n} f_\xi(g_k(y)) \cdot |h'(y)| \cdot I_{I_k}(y)$$

Zadania do samodzielnego rozwiązania

2.1. Dla jakiego a funkcja

$$f(x) = \begin{cases} \dfrac{a}{x^4}, & \text{dla } |x| \ge 1, \\ 0, & \text{dla } |x| < 1, \end{cases}$$

jest funkcją gęstości rozkładu prawdopodobieństwa zmiennej losowej ξ? Znaleźć prawdopodobieństwo, że zmienna losowa przyjmie wartość większą od dwóch.

2.2. Dla jakiego a funkcja

$$f(x) = \begin{cases} 0, & \text{dla } x < 0, \\ axe^{-x}, & \text{dla } x \ge 0, \end{cases}$$

jest funkcją gęstości rozkładu prawdopodobieństwa zmiennej losowej ξ? Znaleźć dystrybuantę F_ξ. Obliczyć $P(\xi = 1)$.

2.3. Dla jakich a i b funkcja

$$F(x) = \begin{cases} 0, & \text{dla } x \le -3, \\ \frac{1}{2} + \frac{1}{a} \arcsin\left(\frac{x}{3}\right), & \text{dla } -3 < x \le 3, \\ b, & \text{dla } x > 3, \end{cases}$$

jest dystrybuantą ciągłej zmiennej losowej? Wyznaczyć funkcję gęstości rozkładu prawdopodobieństwa tej zmiennej losowej.

2.4. Dobrać stałe A i B tak, by funkcja określona wzorem

$$F(x) = A + B \arctan(x), \qquad \text{dla } -\infty < x < \infty$$

była dystrybuantą zmiennej losowej ξ. Wyznaczyć funkcję gęstości rozkładu prawdopodobieństwa tej zmiennej losowej.

2.5. Dla pewnego a funkcja

$$f(x) = \begin{cases} \dfrac{1}{4|ax|^3}, & \text{dla } |x| \ge 1, \\ 0, & \text{dla } |x| < 1, \end{cases}$$

jest funkcją gęstości rozkładu prawdopodobieństwa zmiennej losowej ξ. Wyznaczyć stałą a oraz znaleźć dystrybuantę zmiennej losowej ξ. Obliczyć $P(\xi \in (0.5; 1.5))$.

2.6. Dobrać stałe a oraz $b > 0$ tak, aby funkcja

$$f(x) = \begin{cases} a \cos x, & \text{dla } x \in [0, b], \\ 0, & \text{dla } \notin [0, b], \end{cases}$$

była funkcją gęstości rozkładu prawdopodobieństwa zmiennej losowej ξ. Wyznaczyć dystrybuantę zmiennej losowej ξ. Obliczyć $P(\xi \in (-1, b/2))$.

2.7. Dla jakich a, b oraz c funkcja określona wzorem

$$F(x) = \begin{cases} 0, & \text{dla } x \le 0, \\ a \sin(x), & \text{dla } x \in (0, b], \\ c, & \text{dla } x > b, \end{cases}$$

jest dystrybuantą pewnej zmiennej losowej ξ typu ciągłego. Wyznaczyć gęstość rozkładu prawdopodobieństwa tej zmiennej losowej. Obliczyć $P(\xi \in (-1, b/2))$.

2.8. Dla jakich a, b oraz c funkcja określona wzorem

$$F(x) = \begin{cases} 0, & \text{dla } x \le b, \\ a \ln(x), & \text{dla } x \in (b, 4], \\ c, & \text{dla } x > 4, \end{cases}$$

jest dystrybuantą pewnej zmiennej losowej ξ typu ciągłego. Wyznaczyć gęstość rozkładu prawdopodobieństwa tej zmiennej losowej. Obliczyć $P(\xi \in (-1, 3b/2))$.

2.9. Dystrybuanta F_ξ zmiennej losowej ξ wyraża się wzorem:

$$F_\xi(t) = 1 - e^{-t/8} \quad \text{dla} \quad t > 0$$

Obliczyć $P(\xi \in (-8, 16))$.

2.10. Z urny, w której jest pięć kul białych, trzy czerwone i dwie zielone losujemy ze zwracaniem cztery kule. Podać rozkład liczby wylosowanych kul białych, dystrybuantę odpowiedniej zmiennej losowej i jej wartość oczekiwaną.

2.11. W urnie jest pięć kul białych i dziesięć kul czarnych. Losujemy po jednej kuli bez zwracania do momentu, aż wśród wylosowanych kul znajdą się kule obydwu kolorów. Jaka jest wartość oczekiwana liczby wylosowanych kul czarnych?

2.12. W urnie znajduje się dziesięć kul białych i dziesięć czarnych. Wybieramy z urny kolejno bez zwracania po jednej kuli aż do momentu wyciągnięcia po raz pierwszy kuli czarnej. Wyznaczyć wartość oczekiwaną liczby wyciągniętych kul białych.

2.13. W urnie znajduje się 20 kul, w tym 10 kul białych i 10 czarnych. Ciągniemy losowo bez zwracania 18 kul. Niech N oznacza liczbę wyciągniętych kul białych. Obliczyć wariancję zmiennej losowej N.

2.14. W umie znajduje się dziesięć kul, ponumerowanych liczbami $1, 2, \ldots, 10$. Losujemy ze zwracaniem czterokrotnie po jednej kuli. Niech S oznacza sumę numerów wylosowanych kul. Umawiamy się przy tym, że każdy wylosowany numer występuje w sumie tylko raz, (np. jeśli wylosowaliśmy kule o numerach $3, 1, 5, 3$, to $S = 3+1+5 = 9$). Obliczyć wartość oczekiwaną zmiennej losowej S.

2.15. W urnie znajduje się 25 kul, z których 15 jest białych i 10 czarnych. Losujemy bez zwracania kolejno po jednej kuli. Kończymy losowanie w momencie, kiedy wyciągnięte zostaną wszystkie czarne kule. Obliczyć wartość oczekiwaną liczby pozostałych w urnie białych kul.

2.16. Rozważmy uproszczoną wersję gry w „wojnę". Talia składa się z 52 kart. Dobrze potasowane karty rozdajemy dwóm graczom, każdemu po 26 i układamy w dwie kupki. Gracze wykładają kolejno po jednej karcie z wierzchu swojej kupki i sprawdzają wysokość obu kart. Jeśli obie wyłożone karty są równej wysokości (dwa asy lub dwa króle itd.) to mówimy, że następuje wojna. Po sprawdzeniu, obie karty odkładamy na bok i nie biorą już one udziału w dalszej grze. Powtarzamy tę procedurę 26 razy; gra kończy się, gdy obaj gracze wyłożą wszystkie karty. Obliczyć wartość oczekiwaną liczby wojen.

2.17. Rzucamy dwiema kostkami. Niech zmienną losową będzie różnica oczek. Podać rozkład tej zmiennej losowej, jej dystrybuantę oraz wartość oczekiwaną.

2.18. Rzucamy kością do gry dotąd, aż uzyskamy przynajmniej po jednym z sześciu możliwych wyników. Jaka jest wartość oczekiwana liczby rzutów?

2.19. W ciągu 20 rzutów monetą liczymy serie pięciu orłów. Każdy ciąg sąsiadujących ze sobą pięciu orłów uznajemy za serię. Przyjmujemy zatem, że serie mogą „zachodzić na siebie”, na przykład w ciągu

Nr rzutu	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Wynik	R	O	O	O	O	O	O	O	R	O	O	O	R	O	O	O	O	O	R	R

mamy cztery serie, zaczynające się od miejsc 2, 3, 4 i 14. Obliczyć wartość oczekiwaną liczby serii pięciu orłów w 20 rzutach.

2.20. Wykonujemy rzuty monetą aż do otrzymania po raz pierwszy sekwencji jednakowych wyników w dwóch kolejnych rzutach. Obliczyć wartość oczekiwaną liczby wykonanych rzutów.

2.21. Zysk w pewnej populacji gospodarstw ma rozkład określony funkcją gęstości

$$f(x) = \begin{cases} \frac{1}{18}x, & 0 \le x \le 6, \\ 0, & x \notin \langle 0; 6 \rangle. \end{cases}$$

Jaki jest średni zysk gospodarstw oraz zróżnicowanie zysków? Jakie jest prawdopodobieństwo natrafienia na takie trzy gospodarstwa, z których każde będzie miało zysk większy od 3?

2.22. Z pracy do domu możemy dojechać autobusem jednej z trzech linii: 333, 666 lub 999. Czas dojazdu autobusem do domu wynosi odpowiednio 12, 15 oraz 18 minut. Autobusy z przystanku odjeżdżają co piętnaście minut, przy czym autobusy linii 666 odjeżdżają o cztery minuty później niż linii 333, zaś autobusy linii 999 o sześć minut później niż linii 666. Przyjmując, że przychodzimy na przystanek autobusowy w losowym momencie wyznaczyć rozkład czasu dojazdu do domu, średni czas dojazdu i jego odchylenie standardowe. Podać rozwiązanie zarówno bez, jak i z uwzględnieniem czasu oczekiwania na przystanku.

2.23. Ciągła zmienna losowa ξ ma dystrybuantę F i gęstość rozkładu prawdopodobieństwa f takie, że $f(x)$ jest ciągła dla $x > 1$, $F(1) = 0$, $\frac{f(x)}{1-F(x)} = \frac{1}{x}$ dla $x > 1$. Obliczyć $P(\xi > 2)$.

2.24. Wiadomo, że dla każdej zmiennej losowej ξ mającej skończone momenty do czwartego rzędu włącznie zachodzi $E(\xi - E\xi)^4 \ge \{E(\xi - E\xi)^2\}^4$. Pokazać, że równość zachodzi wtedy i tylko wtedy, gdy ξ ma rozkład $D\,(0.5)$.

2.25. Pobieramy osiem niezależnych realizacji jednowymiarowej zmiennej losowej o nieznanym (ale ciągłym) rozkładzie. Po uporządkowaniu zaobserwowanych wartości w ciąg rosnący $\{z_1, \ldots, z_8\}$ tworzymy przedział (z_2, z_7). Z jakim prawdopodobieństwem tak określony przedział pokrywa wartość mediany rozkładu badanej zmiennej losowej?

2.26. Niech $\xi_1, \ldots, \xi_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie, oraz niech

$$\xi_{\max}^{(n)} = \max\{\xi_1, \ldots, \xi_n\},\ \xi_{\min}^{(n)} = \min\{\xi_1, \ldots, \xi_n\}.$$

Pokazać, że zależność

$$E(\xi_{\max}^{(3)} - \xi_{\min}^{(3)}) = \frac{3}{2} E(\xi_{\max}^{(2)} - \xi_{\min}^{(2)})$$

zachodzi wtedy i tylko wtedy, gdy $E\xi_1 < \infty$.

2.27. Niech ξ ma funkcję gęstości rozkładu prawdopodobieństwa

$$f(x) = \begin{cases} 0.5x + 0.5, & \text{dla } -1 < x < 1, \\ 0, & \text{poza tym.} \end{cases}$$

Wyznaczyć gęstość rozkładu prawdopodobieństwa rozkładu zmiennej losowej $\eta = \xi^2$.

2.28. Na okręgu o promieniu 1 wybieramy losowo i niezależnie dwa punkty. Obliczyć wartość oczekiwaną odległości między nimi (odległość mierzymy wzdłuż cięciwy).

2.29. Na okręgu o obwodzie 1 wybieramy punkt ξ_0, a następnie losowo i niezależnie wybieramy punkty $\xi_1, \ldots, \xi_n$. Niech η oznacza odległość od ξ_0 do najbliższego spośród punktów $\xi_1, \ldots, \xi_n$ liczoną wzdłuż okręgu. Obliczyć $E\eta$.

2.30. Niech $\xi_1, \ldots, \xi_{10}$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie prawdopodobieństwa $P(\xi_i = 1) = 2/3 = 1 - P(\xi_i = -1)$. Niech $\eta_k = \sum_{i=1}^{k} \xi_i$ dla $k = 1, \ldots, 10$. Obliczyć $P(\eta_{10} = 2 \text{ i } \eta_1 \leq 5, \eta_2 \leq 5, \ldots, \eta_9 \leq 5)$.

2.31. Z odcinka $[0, 1]$ wybieramy losowo punkt ξ_1. Następnie z odcinka $[0, \xi_1]$ wybieramy losowo punkt ξ_2, z odcinka $[0, \xi_2]$ - punkt ξ_3 i tak dalej. Obliczyć $\frac{D\xi_n}{E\xi_n}$.

2.32. Niech $\xi_1, \xi_2, \xi_3, \xi_4$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie z gęstością rozkładu prawdopodobieństwa

$$f(x) = \begin{cases} \frac{2}{x^3}, & \text{dla } x > 0, \\ 0, & \text{poza tym.} \end{cases}$$

Obliczyć $E\left(\frac{\min\{\xi_1,\xi_2,\xi_3,\xi_4\}}{\max\{\xi_1,\xi_2,\xi_3,\xi_4\}}\right)$.

2.33. Załóżmy, że $\xi_1, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o jednakowym, ciągłym rozkładzie prawdopodobieństwa, mającymi momenty rzędu 1, 2 i 3. Znamy $\mu = E\xi_i$ i $\sigma^2 = D^2\xi_i$. Niech $f(x)$ oznacza gęstość rozkładu prawdopodobieństwa pojedynczej zmiennej ξ_i. Wiemy, że rozkład jest symetryczny w tym sensie, że $f(\mu + x) = f(\mu - x)$ dla każdego x. Obliczyć $E\eta_n^3$, gdzie $\eta_n = \xi_1 + \cdots + \xi_n$.

2.34. Jeden z boków prostokąta ma rozkład $U(0,3)$. Jaki powinien mieć rozkład drugi z boków, jeżeli wiadomo, że zmiana długości boków tego prostokąta nie powoduje zmiany jego pola, które jest równe 1. Podać funkcję gęstości rozkładu prawdopodobieństwa drugiego boku.

2.35. Promień koła jest zmienną losową o rozkładzie $E(1)$. Znaleźć gęstość rozkładu prawdopodobieństwa pola tego koła.

2.36. Na jednostkowym okręgu ustalono punkt A i wybrano losowo punkt M. Niech ξ oznacza długość cięciwy AM. Znaleźć gęstość rozkładu prawdopodobieństwa zmiennej losowej ξ.

2.37. Pole kwadratu jest zmienną losową o gęstości rozkładu prawdopodobieństwa

$$f(x) = \begin{cases} \frac{1}{6\sqrt{x}}, & \text{dla } 0 < x < 9, \\ 0, & \text{dla pozostałych } x. \end{cases}$$

Znaleźć funkcję gęstości rozkładu prawdopodobieństwa długości boku tego kwadratu.

2.38. Znaleźć gęstość rozkładu prawdopodobieństwa zmiennej losowej będącej polem kwadratu, którego długość boku jest zmienną losową o rozkładzie $U(0,4)$.

3. Wektory losowe

Wektor losowy. **Wektorem losowym** $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ nazywamy odwzorowanie

$$\boldsymbol{\xi} : (\Omega, \mathcal{F}, P) \to \mathbb{X} \subseteq \mathbb{R}^n$$

o własności:

$$\{\omega \in \Omega : \ \xi_1(\omega) \le x_1, \ldots, \xi_n(\omega) \le x_n\} \in \mathcal{F}$$

dla dowolnego $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

Rozkład łączny. **Rozkładem wektora losowego** $\boldsymbol{\xi}$ nazywamy rozkład prawdopodobieństwa $P_{\boldsymbol{\xi}}$ określony wzorem

$$P_{\boldsymbol{\xi}}(A) = P(\{\omega \in \Omega : \ \boldsymbol{\xi}(\omega) \in A\}) \quad \text{dla } A \in \mathcal{B}(\mathbb{X})$$

Dystrybuanta. Funkcję $F_{\boldsymbol{\xi}} : \mathbb{R}^n \to [0,1]$ postaci

$$F_{\boldsymbol{\xi}}(x_1, \ldots, x_n) = P(\xi_1 \le x_1, \ldots, \xi_n \le x_n)$$

nazywamy **dystrybuantą** wektora losowego $\boldsymbol{\xi}$

Rozkład skokowy. Wektor losowy jest typu skokowego, jeżeli istnieje zbiór przeliczalny $\mathbb{X} \subset \mathbb{R}^n$, taki że $P_{\boldsymbol{\xi}}(\mathbb{X}) = 1$

Rozkład ciągły. Wektor losowy jest typu ciągłego, jeżeli istnieje nieujemna funkcja $f_{\boldsymbol{\xi}}(\boldsymbol{x})$ taka, że dla każdego $\boldsymbol{x} \in \mathbb{R}^n$

$$F_{\boldsymbol{\xi}}(\boldsymbol{x}) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_2} f_{\boldsymbol{\xi}}(u_1, \ldots, u_n)\, du_1 \ldots du_n$$

Funkcja $f_{\boldsymbol{\xi}}(\boldsymbol{x})$ nazywa się gęstością.

Rozkład brzegowy. **Rozkładem brzegowym zmiennej losowej** ξ_1 nazywamy

$$f_{\xi_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{\xi}}(x_1, \ldots, x_n)\, dx_2 \ldots dx_n$$

Rozkładem brzegowym wektora losowego $(\xi_1 \xi_2)$ nazywamy

$$f_{(\xi_1,\xi_2)}(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\boldsymbol{\xi}}(x_1, \ldots, x_n)\, dx_3 \ldots dx_n$$

Rozkład warunkowy. Niech $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$. **Rozkładem warunkowym** wektora $\boldsymbol{\xi}_1$ pod warunkiem wektora $\boldsymbol{\xi}_2$ nazywamy

$$f_{\boldsymbol{\xi}_1|\boldsymbol{\xi}_2=\boldsymbol{x}_2}(\boldsymbol{x}_1) = \frac{f_{\boldsymbol{\xi}}(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f_{\boldsymbol{\xi}_2}(\boldsymbol{x}_2)}$$

Niezależność. Wektory losowe $\boldsymbol{\xi}_1$ i $\boldsymbol{\xi}_2$ nazywamy **niezależnymi**, jeżeli dla wszystkich $\boldsymbol{x}_1$ oraz $\boldsymbol{x}_2$

$$f_{\boldsymbol{\xi}}(\boldsymbol{x}_1, \boldsymbol{x}_2) = f_{\boldsymbol{\xi}_1}(\boldsymbol{x}_1) f_{\boldsymbol{\xi}_2}(\boldsymbol{x}_2)$$

Momenty rzędu k. **Momentem rzędu k** wektora losowego nazywamy

$$E\xi_1^{k_1}\cdots\xi_n^{k_n} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1^{k_1}\cdots x_n^{k_n} f_{\boldsymbol{\xi}}(\boldsymbol{x})d\boldsymbol{x}$$

dla wszystkich naturalnych $k_1, \ldots, k_n$ takich, że $k_1 + \cdots + k_n = k$.

Wartość oczekiwana. Wektorem wartości oczekiwanych nazywamy

$$E\boldsymbol{\xi} = \begin{bmatrix} E\xi_1 \\ \vdots \\ E\xi_n \end{bmatrix}$$

Momenty centralne. **Momentem centralnym rzędu k** nazywamy

$$E(\xi_1 - E\xi_1)^{k_1}\cdots(\xi_n - E\xi_n)^{k_n}$$

dla wszystkich naturalnych $k_1, \ldots, k_n$ takich, że $k_1 + \cdots + k_n = k$.

Kowariancja. **Kowariancją** zmiennych losowych ξ_i oraz ξ_j nazywamy

$$Cov(\xi_i, \xi_j) = E(\xi_i - E\xi_i)(\xi_j - E\xi_j)$$

Macierz kowariancji. Macierzą kowariancji wektora losowego $\boldsymbol{\xi}$ nazywamy

$$D\boldsymbol{\xi} = \begin{bmatrix} D^2\xi_1 & Cov(\xi_1,\xi_2) & Cov(\xi_1,\xi_3) & \cdots & Cov(\xi_1,\xi_n) \\ Cov(\xi_1,\xi_2) & D^2\xi_2 & Cov(\xi_2,\xi_3) & \cdots & Cov(\xi_2,\xi_n) \\ Cov(\xi_1,\xi_3) & Cov(\xi_2,\xi_3) & D^2\xi_3 & \cdots & Cov(\xi_3,\xi_n) \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ Cov(\xi_1,\xi_n) & Cov(\xi_2,\xi_n) & Cov(\xi_3,\xi_n) & \cdots & D^2\xi_n \end{bmatrix}$$

Współczynnik korelacji liniowej. **Współczynnikiem korelacji liniowej** zmiennych ξ_i oraz ξ_j nazywamy

$$Corr(\xi_i, \xi_j) = \frac{Cov(\xi_i, \xi_j)}{\sqrt{D^2\xi_i D^2\xi_j}}$$

Momenty: ważna własność. Jeżeli $\boldsymbol{A}$ jest macierzą macierz odpowiednich wymiarów, to

$$E\boldsymbol{A\xi} = \boldsymbol{A}E\boldsymbol{\xi}$$
$$D\boldsymbol{A\xi} = \boldsymbol{A}D\boldsymbol{\xi}\boldsymbol{A}^T$$

Warunkowa wartość oczekiwana. **Warunkowa wartość oczekiwana** $E(\xi_1|\xi_2)$ zmiennej losowej ξ_1 pod warunkiem zmiennej losowej ξ_2 to nowa zmienna losowa, która jest postaci $E(\xi_1|\xi_2) = m(\xi_2)$ z prawdopodobieństwem 1 dla borelowskiej funkcji $m(x)$ takiej, że

$$\int_B m(y)dF_{\xi_2}(y) = \int\int_{B\times\mathbb{R}} xdF_{\xi_1,\xi_2}(x,y)$$

dla każdego borelowskiego zbioru B.

Własności warunkowej wartości oczekiwanej.

$$E(E(\xi_1|\xi_2)) = E\xi_1$$
$$D^2\xi_1 = E\left(D^2(\xi_1|\xi_2)\right) + D^2\left(E(\xi_1|\xi_2)\right)$$

Przykład. *Rozważmy rzut dwiema sześciennymi kostkami. Niech ξ_1 będzie wynikiem rzutu pierwszą kostką, a ξ_2 wynikiem rzutu drugą kostką. Wówczas $\boldsymbol{\xi} = (\xi_1, \xi_2)$ jest wynikiem rzutu parą kostek.*

Rozwiązanie. *Funkcja rozkładu prawdopodobieństwa ma postać*

$$P\{(\xi_1,\xi_2) = (i,j)\} = p_{ij}, \quad i,j = 1,\ldots,6$$

Prawdopodobieństwa p_{ij} są takie, że $\sum_{i,j} p_{ij} = 1$.

Dystrybuanta rozkładu wektora $\boldsymbol{\xi}$ ma postać

$$F_{(\xi_1,\xi_2)}(x,y) = P\{\xi_1 \le x, \xi_2 \le y\} = \sum_{(i,j): i\le x, j\le y} p_{ij}$$

dla $(x,y) \in \mathbb{R}^2$

Funkcja rozkładu brzegowego zmiennej losowej ξ_1:

$$P\{\xi_1 = i\} = \sum_{j=1}^{6} P\{(\xi_1,\xi_2) = (i,j)\} = \sum_{j=1}^{6} p_{ij}, \quad i = 1,\ldots,6$$

Dystrybuanta rozkładu zmiennej losowej ξ_1

$$F_{\xi_1}(x) = P\{\xi_1 \le x\} = \sum_{i \le x} P\{\xi_1 = i\}$$

Podobnie można znaleźć rozkład brzegowy zmiennej losowej ξ_2.

Funkcja rozkładu prawdopodobieństwa zmiennej losowej ξ_2 pod warunkiem $\xi_1 = i$ ma postać

$$P\{\xi_2 = j | \xi_1 = i\} = \frac{P\{(\xi_1, \xi_2) = (i, j)\}}{P\{\xi_1 = i\}}, \quad j = 1, \ldots, 6$$

Wyniki rzutów kostkami sa niezależne, tzn. zmienne losowe ξ_1 oraz ξ_2 wtedy i tylko wtedy, gdy

$$P\{\xi_2 = j | \xi_1 = i\} = P\{\xi_2 = j\}, \quad j = 1, \ldots, 6, \; i = 1, \ldots, 6$$

lub równoważnie

$$P\{(\xi_1, \xi_2) = (i, j)\} = P\{\xi_1 = i\} P\{\xi_2 = j\}, \quad i, j = 1, \ldots, 6$$

Moment rzędu k. Jeżeli $k_1 + k_2 = k$, to z definicji

$$E\xi_1^{k_1} \xi_2^{k_2} = \sum_{i,j} i^{k_1} j^{k_2} p_{ij}.$$

Podobnie dla momentów centralnych.

W szczególności

Wektor wartości oczekiwanych

$$E\boldsymbol{\xi} = \begin{bmatrix} E\xi_1 \\ E\xi_2 \end{bmatrix}$$

Z definicji

$$E\xi_1 = \sum_{i,j} i p_{ij} = \sum_i i P\{\xi_1 = i\}$$

$$E\xi_2 = \sum_{i,j} j p_{ij} = \sum_j j P\{\xi_2 = j\}$$

Macierz kowariancji

$$D\boldsymbol{\xi} = \begin{bmatrix} D^2\xi_1 & Cov(\xi_1, \xi_2) \\ Cov(\xi_1, \xi_2) & D^2\xi_2 \end{bmatrix},$$

gdzie

$$D^2\xi_1 = \sum_{i,j} (i - E\xi_1)^2 p_{ij} = \sum_i (i - E\xi_1)^2 P\{\xi_1 = i\}$$

$$Cov(\xi_1, \xi_2) = \sum_{i,j} (i - E\xi_1)(j - E\xi_2) p_{ij}$$

$$D^2\xi_2 = \sum_{i,j} (j - E\xi_2)^2 p_{ij} = \sum_j (j - E\xi_2)^2 P\{\xi_2 = j\}$$

Warunkowa wartość oczekiwana $E(\xi_1|\xi_2)$ jest to zmienna losowa o rozkładzie

$$\begin{array}{ll} E(\xi_1|\xi_2=1) & P\{\xi_2=1\} \\ E(\xi_1|\xi_2=2) & P\{\xi_2=2\} \\ E(\xi_1|\xi_2=3) & P\{\xi_2=3\} \\ E(\xi_1|\xi_2=4) & P\{\xi_2=4\} \\ E(\xi_1|\xi_2=5) & P\{\xi_2=5\} \\ E(\xi_1|\xi_2=6) & P\{\xi_2=6\} \end{array}$$

gdzie

$$E(\xi_1|\xi_2=j) = \sum_{i=1}^{6} iP\{\xi_1=i|\xi_2=j\} = \sum_{i=1}^{6} i\frac{P\{\xi_1=i,\xi_2=j\}}{P\{\xi_2=j\}} = \sum_{i=1}^{6} i\frac{p_{ij}}{\sum_{k=1}^{6} p_{kj}}$$

□

Przykład. *Strzelamy do okrągłej tarczy. Umówmy się, że środek układu współrzędnych znajduje się w środku tarczy. Niech ξ_1 będzie odciętą punktu trafienia, a ξ_2 jego rzędną. Wówczas wektor losowy $\boldsymbol{\xi} = (\xi_1, \xi_2)$ opisuje współrzędne punktu trafienia. Zakładamy, że każdy punkt tarczy ma jednakowe szanse być trafionym (rozkład wektora $\boldsymbol{\xi}$ jest jednostajny na kole).*

Rozwiązanie. Funkcja gęstości rozkładu prawdopodobieństwa ma postać

$$f_{(\xi_1,\xi_2)}(x,y) = a\mathbf{1}_{\{x^2+y^2\le r^2\}}(x,y)$$

gdzie stała $a > 0$ jest taka, że $\int_{\mathbb{R}^2} f(x,y)dxdy = 1$. Jak łatwo obliczyć

$$\int_{\mathbb{R}^2} f(x,y)dxdy = \int_{\{x^2+y^2\le r^2\}} dxdy = \int_{-r}^{r}\left[\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dy\right]dx = \pi r^2$$

Zatem funkcja gęstości rozkładu prawdopodobieństwa ma postać

$$f_{(\xi_1,\xi_2)}(x,y) = \frac{1}{\pi r^2}\mathbf{1}_{\{x^2+y^2\le r^2\}}(x,y)$$

Dystrybuanta wyrażona jest wzorem $F_{(\xi_1,\xi_2)}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(t,u)dudt$

Funkcja gęstości rozkładu prawdopodobieństwa zmiennej losowej ξ_1

$$\begin{aligned} f_{\xi_1}(x) &= \int_{-\infty}^{\infty} f_{(\xi_1,\xi_2)}(x,u)du \\ &= \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} \frac{1}{\pi r^2}du \\ &= \frac{2\sqrt{r^2-x^2}}{\pi r^2}, \qquad \text{dla } x\in(-r,r) \end{aligned}$$

Podobnie funkcja gęstości rozkładu prawdopodobieństwa zmiennej losowej ξ_2

$$f_{\xi_2}(y) = \frac{2\sqrt{r^2 - y^2}}{\pi r^2}, \quad \text{dla } y \in (-r, r)$$

Dystrybuanta

$$F_{\xi_1}(x) = \int_{-\infty}^{x} f_{\xi_1}(t)dt$$

Funkcja gęstości rozkładu prawdopodobieństwa zmiennej losowej ξ_2 *pod warunkiem* $\xi_1 = x$ *ma postać*

$$\begin{aligned} f_{\xi_2|\xi_1=x}(y) &= \frac{f_{(\xi_1,\xi_2)}(x,y)}{f_{\xi_1}(x)} \\ &= \frac{1/\pi r^2}{2\sqrt{r^2 - x^2}/\pi r^2} = \frac{1}{2\sqrt{r^2 - x^2}}, \quad \text{dla } y \in (-\sqrt{r^2 - x^2}, \sqrt{r^2 - x^2}) \end{aligned}$$

Podobnie

$$f_{\xi_1|\xi_2=y}(x) = \frac{1}{2\sqrt{r^2 - y^2}}, \quad \text{dla } x \in (-\sqrt{r^2 - y^2}, \sqrt{r^2 - y^2})$$

Moment rzędu k. *Jeżeli* $k_1 + k_2 = k$, *to z definicji*

$$E\xi_1^{k_1}\xi_2^{k_2} = \int_{\{x^2+y^2 \le r^2\}} x^{k_1} y^{k_2} dxdy.$$

Podobnie dla momentów centralnych.

W szczególności

Wektor wartości oczekiwanych

$$E\boldsymbol{\xi} = \begin{bmatrix} E\xi_1 \\ E\xi_2 \end{bmatrix}$$

Z definicji

$$\begin{aligned} E\xi_1 &= \int_{\{x^2+y^2 \le r^2\}} xdxdy = \int_0^1 \left[\int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dy \right] xdx = \int_{-r}^{r} xf_{\xi_1}(x)dx = 0 \\ E\xi_2 &= \int_{\{x^2+y^2 \le r^2\}} ydxdy = 0 \end{aligned}$$

Macierz kowariancji

$$D\boldsymbol{\xi} = \begin{bmatrix} D^2\xi_1 & Cov(\xi_1, \xi_2) \\ Cov(\xi_1, \xi_2) & D^2\xi_2 \end{bmatrix},$$

gdzie

$$D^2\xi_1 = \frac{1}{\pi r^2}\int_{\{x^2+y^2\le r^2\}}(x-E\xi_1)^2dxdy = \int_{-r}^{r}(x-E\xi_1)^2 f_{\xi_1}(x)dx = \frac{r^2}{4}$$
$$Cov(\xi_1,\xi_2) = \frac{1}{\pi r^2}\int_{\{x^2+y^2\le r^2\}}(x-E\xi_1)(y-E\xi_2)dxdy = 0$$
$$D^2\xi_2 = \frac{1}{\pi r^2}\int_{\{x^2+y^2\le r^2\}}(y-E\xi_2)^2dxdy = \int_{-r}^{r}(y-E\xi_2)^2 f_{\xi_2}(y)dy = \frac{r^2}{4}$$

Wyznaczamy rozkład warunkowej wartości oczekiwanej $E(\xi_1|\xi_2)$. Najpierw znajdujemy $E(\xi_1|\xi_2 = y)$ dla ustalonego y (oczywiście $y \in \langle -r, r\rangle$).

$$E(\xi_1|\xi_2 = y) = \int_{-\infty}^{\infty} x f_{\xi_1|\xi_2=y}(x)dx = \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} x\frac{1}{2\sqrt{r^2-y^2}}dx = 0$$

A zatem $E(\xi_1|\xi_2) = 0$.
□

Zadania do samodzielnego rozwiązania

3.1. Pokazać, że $Cov(\xi,\eta) = E\xi\eta - E\xi E\eta$.

3.2. Pokazać, że dla dowolnych $a, b \in R$ zachodzi $Cov(aX + b, Y) = Cov(aX, Y) = aCov(X, Y)$.

3.3. Udowodnić, że jeżeli zmienne losowe ξ oraz η są niezależne, to $Cov(\xi,\eta) = 0$. Podać przykład, że implikacji przeciwna nie jest prawdziwa.

3.4. Pokazać, że $|Corr(\xi,\eta)| \le 1$ (skorzystać z nierówności Schwarza).

3.5. Pokazać, że $|Corr(\xi,\eta)| = 1$ wtedy i tylko wtedy, gdy istnieją takie stałe a, b, że $\xi = a\eta + b$.

3.6. Udowodnić, że $E(E(\xi|\eta)) = E\xi$.

3.7. Udowodnić, że $D^2\xi = E\left(D^2(\xi|\eta)\right) + D^2\left(E(\xi|\eta)\right)$.

3.8. Niech K będzie zmienną losową taką, że $P(K = k) = 0.1$ dla $k = 1, \ldots, 10$. Niech

$$\xi_k = \begin{cases} 1, & \text{gdy } K = k, \\ 0, & \text{gdy } K \ne k, \end{cases} \qquad \eta_5 = \xi_1 + \xi_2 + \xi_3 + \xi_4 + \xi_5.$$

Obliczyć $Cov(\xi_1, \eta_5)$.

3.9. W urnie znajdują się kule, z których każda oznaczona jest jedną z liter alfabetu: 10 kul oznaczonych literą A, 20 kul oznaczonych literą B, 30 kul oznaczonych literą C i x kul oznaczonych innymi literami alfabetu. Losujemy ze zwracaniem siedem razy po jednej kuli z urny. Zmienne losowe N_A, N_B, N_C oznaczają odpowiednio liczbę tych ciągnięć, w których pojawiła się litera A, B, C. Jakie musi być x, aby zmienne losowe $N_A + N_B$ oraz $N_B + N_C$ były nieskorelowane?

3.10. W urnie znajduje się 25 kul, z których $m = 15$ jest białych, $r - m = 10$ czarnych. Losujemy bez zwracania najpierw $n_1 = 6$ kul, a następnie spośród pozostałych w urnie, losujemy bez zwracania $n_2 = 8$ kul. Niech S_1 oznacza liczbę białych kul wybranych w pierwszym losowaniu, a S_2 oznacza liczbę białych kul wybranych w drugim losowaniu. Obliczyć $Cov(S_1, S_2)$.

3.11. W urnie znajduje się 40 kul, z których 25 jest białych i 15 czarnych. Losujemy bez zwracania najpierw 13 kul, a następnie z pozostałych kul w urnie losujemy bez zwracania 8 kul. Niech S_1 oznacza liczbę kul białych w pierwszym losowaniu, a S_2 liczbę kul białych w drugim losowaniu. Obliczyć $Cov(S_1, S_2)$.

3.12. W urnie znajduje się trzydzieści kul, na każdej narysowana jest litera i cyfra. Mamy dziesięć kul oznaczonych $X1$, osiem kul oznaczonych $Y1$, osiem kul oznaczonych $X2$ oraz cztery kule oznaczone $Y2$. Losujemy bez zwracania piętnaście kul. Niech N_X określa liczbę kul oznaczonych literą X wśród wylosowanych, a N_2 liczbę kul z cyfrą 2 wśród kul wylosowanych. Obliczyć $E(N_X|N_2)$.

3.13. Zmienne losowe ξ i η są niezależne i mają następujące rozkłady prawdopodobieństwa: $P(\xi = n) = 2^{-n}$ dla $n = 1, 2, \ldots$, $P(\eta > x) = 2^{-x}$ dla $x > 0$. Obliczyć $P(\eta > \xi)$.

3.14. Niech dwuwymiarowy wektor losowy ma gęstość:

$$f_{\xi,\eta}(x, y) = \begin{cases} 2 - x - y, & \text{dla } (x, y) \in (0, 1) \times (0, 1), \\ 0, & \text{poza tym.} \end{cases}$$

Obliczyć prawdopodobieństwo

$$P\left((\xi, \eta) \in \left(\frac{1}{2}, 1\right) \times \left(\frac{1}{2}, 1\right)\right).$$

3.15. Zmienne losowe ξ i η mają łączny rozkład prawdopodobieństwa o gęstości

$$f(x, y) = \begin{cases} e^{-y+x}, & \text{dla } 0 < x < 1 \text{ i } y > x, \\ 0, & \text{poza tym.} \end{cases}$$

Obliczyć wartość oczekiwaną $E(\xi + \eta)$.

3.16. Łączny rozkład zmiennych losowych ξ i η ma gęstość:

$$f_{\xi,\eta}(x, y) = \begin{cases} e^{-y+x}, & \text{dla } 0 < x < 1 \text{ i } y > x, \\ 0, & \text{poza tym.} \end{cases}$$

Obliczyć $D^2\eta$

3.17. O zmiennych losowych $\xi_1, \ldots, \xi_n$ o tej samej wartości oczekiwanej równej μ oraz tej samej wariancji σ^2 zakładamy, że $Cov(X_i, X_j) = \varrho\sigma^2$ dla $i \neq j$. Zmienne losowe $\varepsilon_1, \ldots, \varepsilon_n$ są nawzajem niezależne oraz niezależne od zmiennych $\xi_1, \ldots, \xi_n$ i mają rozkłady prawdopodobieństwa $P(\varepsilon_i = -1) = P(\varepsilon_i = 1) = 0.5$. Obliczyć wariancję zmiennej losowej $\sum_{i=1}^{n} \varepsilon_i \xi_i$.

3.18. Niech $\xi_1, \ldots, \xi_n, \ldots$ będą dodatnimi, niezależnymi zmiennymi losowymi o jednakowym ciągłym rozkładzie prawdopodobieństwa. Niech $\eta_n = \max\{\xi_1, \ldots, \xi_n\}$ dla $n > 0$ ($\eta_0 = 0$). Zmienne losowe N i M są od siebie niezależne i niezależne od $\xi_1, \ldots, \xi_n, \ldots$ Wiadomo, że obie te zmienne mają rozkłady $Po\,(\lambda)$ i $Po\,(\mu)$, odpowiednio. Obliczyć $P(\eta_{N+M} > \eta_N)$.

3.19. W urnie znajduje się 10 kul Amarantowych, 10 kul Białych i 10 kul Czarnych. Losujemy bez zwracania 12 kul. Niech A oznacza liczbę wylosowanych kul Amarantowych, B oznacza liczbę wylosowanych kul Białych, C oznacza liczbę wylosowanych kul Czarnych. Obliczyć współczynnik korelacji zmiennych losowych A i B. (Wskazówka: $Var(A + B + C) = 0$.)

3.20. O zmiennych losowych ξ_0 i ξ_1 zakładamy, że $E\xi_0 = E\xi_1 = 0$, $D^2\xi_0 = D^2\xi_1 = 1$ i $Cov(\xi_0, \xi_1) = \varrho$, gdzie $0 < \varrho < 1$. Niech $\xi_1 = \varrho\xi_0 + \eta$. Rozważmy zmienne losowe postaci $\hat{\eta} = z\xi_1 + (1 - z)\xi_0$ interpretowane jako predyktory nieobserwowanej zmiennej η. Znaleźć współczynnik z_*, dla którego błąd średniokwadratowy $E(\hat{\eta} - \eta)^2$ jest minimalny.

3.21. O zmiennych losowych ξ i η wiemy, że $0 \le \eta < \xi$, $P(\xi = 0) = 0$, $E(\eta|\xi) = \frac{\xi}{2}$ i $D^2\eta = \frac{1}{2}D^2\xi + \frac{1}{4}(E\xi)^2$. Pokazać, że $P(\eta = \xi) = 0.5$.

3.22. Funkcja gęstości wektora losowego (ξ, η) dana jest wzorem

$$f(x, y) = \begin{cases} \frac{3}{4}x + 2xy + \frac{1}{4}y, & \text{dla } (x, y) \in (0, 1) \times (0, 1), \\ 0, & \text{poza tym.} \end{cases}$$

Obliczyć $P\left(\xi > \frac{1}{2} | \eta > \frac{1}{2}\right)$.

3.23. Funkcja gęstości wektora losowego (ξ, η) dana jest wzorem

$$f(x, y) = \begin{cases} x + y, & \text{dla } (x, y) \in (0, 1) \times (0, 1), \\ 0, & \text{poza tym.} \end{cases}$$

Obliczyć $E(\xi|\eta = 0.5)$.

3.24. Zmienne losowe ξ i η są niezależne Zmienna losowa ξ ma rozkład $Pow\,(2)$. Zmienna losowa η ma rozkład $E\,(1)$. Obliczyć $E(\xi + \eta | \xi \le 0.5)$.

3.25. Losujemy ze zwracaniem po jednej karcie do gry z talii 52 kart tak długo aż wylosujemy pika. Niech η oznacza zmienną losową równą liczbie wyciągniętych kart, a ξ zmienną losową równą liczbie kart, w których uzyskaliśmy karo. Obliczyć $E(\eta|\xi = 4)$.

3.26. W urnie znajduje się 20 kul, na każdej z nich narysowana jest litera i cyfra. Mamy osiem kul oznaczonych A_1, cztery kule oznaczone A_2, sześć kul oznaczonych B_1 i dwie kule oznaczone B_2. Losujemy bez zwracania dziesięć kul. Niech N_A oznacza liczbę wylosowanych kul oznaczonych literą A, zaś N_1 - liczbę wylosowanych kul oznaczonych cyfrą 1. Obliczyć $E(N_1|N_A)$.

3.27. Zmienne losowe ξ i η mają łączny rozkład prawdopodobieństwa o gęstości

$$f(x,y) = \begin{cases} xe^{-x(y-x)}, & \text{dla } y > x,\ 0 < x < 1, \\ 0, & \text{poza tym.} \end{cases}$$

Obliczyć $P(\eta > \mu(\xi))$ wiedząc, że $\mu(\xi) = E(\eta|\xi)$.

3.28. Zmienne losowe $\xi_1, \ldots, \xi_5$ są niezależne i mają jednakowy rozkład $E(\theta)$, gdzie $\theta > 0$ jest ustaloną liczbą. Niech η oznacza zmienną losowa równą 1, gdy $\xi_1 \geq 3$ i równą 0 w pozostałych przypadkach. Niech $\zeta = \sum_{i=1}^{5} \xi_i$. Wyznaczyć $E(\eta|\zeta = 5)$.

3.29. Niech ξ będzie zmienną losową o rozkładzie o gęstości

$$f(x) = \begin{cases} \frac{64}{(2+x)^5}, & \text{dla } x > 0, \\ 0, & \text{poza tym.} \end{cases}$$

Niech η będzie zmienną losową równą

$$\eta = \begin{cases} 0, & \text{gdy } \xi \leq 3, \\ \xi - 3, & \text{gdy } X\xi > 3. \end{cases}$$

Wyznaczyć $D^2(\eta|\xi > 3)$.

3.30. Załóżmy, że zmienne losowe $\xi_1, \ldots, \xi_n, \ldots$ są niezależne, mają jednakowy rozkład prawdopodobieństwa, $E\xi_i = \mu$, $D^2\xi_i = \sigma^2$. Niech N będzie zmienną losową niezależną od ciągu $\xi_1, \ldots, \xi_n, \ldots$ o rozkładzie prawdopodobieństwa danym wzorem

$$P(N = n) = n(1-\theta)^{n-1}\theta^2 \quad \text{dla} \quad n = 1, 2, \ldots$$

Niech $\eta_n = \sum_{i=1}^{n} \xi_i$. Obliczyć $D^2\left(\frac{\eta_N}{N}\right)$.

3.31. Rzucamy dziesięć razy monetą. Niech ξ_n oznacza liczbę orłów w n pierwszych rzutach. Obliczyć $ED^2(\xi_5|\xi_{10})$.

3.32. Wiemy, że zmienne losowe $\xi_1, \ldots, \xi_m, \ldots, \xi_n$ są niezależne i mają jednakowy rozkład prawdopodobieństwa. Zakładamy, że $1 < m < n$ i znamy $D^2(\xi_i) = \sigma^2$. Niech $\eta_m = \xi_1 + \cdots + \xi_m$ i $\eta_n = \xi_1 + \cdots + \xi_m + \cdots + \xi_n$. Obliczyć $ED^2(\eta_m|\eta_n)$.

3.33. Macierz kowariancji wektora losowego $(\xi_1, \ldots, \xi_n)$ jest postaci $\sigma^2((1-\varrho)I + \varrho E)$, gdzie macierze I oraz E to, odpowiednio, macierz jednostkowa i macierz złożona z samych jedynek, a obie są wymiarów $n \times n$. Zakładamy, że macierz jest rzędu n. Jaki jest dopuszczalny zakres wartości parametru ϱ?

3.34. Dwuwymiarowy rozkład pary zmiennych losowych ξ oraz η dany jest za pomocą tablicy

ξ \ η	0	1	2
0	0.10	0.06	0.08
1	0.27	0.05	0.17
2	0.02	0.05	0.07
3	0.03	0.04	0.06

Obliczyć $E(\xi + \eta)$ oraz $P(\xi + \eta \geq 2|\eta < 1)$.

3.35. Dwuwymiarowy rozkład pary zmiennych losowych ξ oraz η dany jest za pomocą tablicy

ξ \ η	0	1	2
0	0.09	0.01	0.08
1	0.17	0.15	0.17
2	0.03	0.05	0.07
3	0.02	0.10	0.06

Obliczyć $E(\xi + \eta)$ oraz $P(\xi + \eta \leq 1|\eta < 2)$.

3.36. Dwuwymiarowy rozkład pary zmiennych losowych ξ oraz η dany jest za pomocą tablicy

ξ \ η	0	1	2
0	0.09	0.01	0.08
1	0.17	0.11	0.12
2	0.03	0.05	0.07
3	0.02	0.10	0.15

Obliczyć $E(\xi - 2\eta)$ oraz $P(\xi + \eta \leq 1|\eta > 0)$.

3.37. Dwuwymiarowy rozkład pary zmiennych losowych ξ oraz η dany jest za pomocą tablicy

ξ \ η	0	1	2
0	0.18	0.01	0.08
1	0.17	0.01	0.12
2	0.12	0.03	0.01
3	0.02	0.10	0.15

Obliczyć $E(2\xi + \eta)$ oraz $P(\xi + \eta \leq 1|\eta \geq 1)$.

3.38. Zmienne losowe ξ oraz η są niezależne oraz $E\xi = 2$, $D^2\xi = 4$, $E\eta = -3$, $D^2\eta = 4$. Obliczyć
a) $E((\xi - 3)(\eta^2 + 3))$
b) $Corr(\xi^2, \eta - 3)$
c) $\min\limits_t E(\xi - \eta\xi - t)^2$

3.39. Zmienne losowe ξ oraz η są niezależne oraz $E\xi = 2$, $D^2\xi = 3$, $E\eta = -1$, $D^2\eta = 14$. Obliczyć
a) $E((\xi+3)(\eta^2+6))$
b) $Corr(\xi-3, \eta^2-1)$
c) $\min_t E(\xi(1-2\eta)-t)^2$

3.40. Zmienne losowe ξ oraz η są niezależne oraz $E\xi = -2$, $D^2\xi = 4$, $E\eta = -3$, $D^2\eta = 7$. Obliczyć
a) $E((\xi-3)(\eta^2+3))$
b) $Corr(\xi^2+2, 2(\eta-2))$
c) $\min_t E(\xi+12\xi\eta-t)^2$

3.41. Zmienne losowe ξ oraz η są niezależne oraz $E\xi = 1$, $D^2\xi = 2$, $E\eta = -3$, $D^2\eta = 3$. Obliczyć
a) $E((\xi+103)(\eta^2+3))$
b) $Corr(\xi, (\eta+3)^2)$
c) $\min_t E(\eta\xi-3\xi-t)^2$

4. Funkcje wektorów losowych

Niech $\boldsymbol{\xi}$ będzie p wymiarowym wektorem losowym oraz niech $h : \mathbb{R}^p \to \mathbb{R}^m$. interesuje nas rozkład wektora losowego $\boldsymbol{\eta} = h(\boldsymbol{\xi})$. Dystrybuanta wektora $\boldsymbol{\eta}$ ma postać:

$$\begin{aligned} F_{\boldsymbol{\eta}}(\boldsymbol{y}) &= P\left\{(\eta_1, \ldots, \eta_m) \in (-\infty, y_1\rangle \times \cdots \times (-\infty, y_m\rangle\right\} \\ &= P\left\{(\xi_1, \ldots, \xi_p) \in h^{-1}\left((-\infty, y_1\rangle \times \cdots \times (-\infty, y_m\rangle\right)\right\} \\ &= \int \cdots \int_{\{(x_1,\ldots,x_p)\in h^{-1}((-\infty,y_1\rangle\times\cdots\times(-\infty,y_m\rangle)\}} f_{\boldsymbol{\xi}}(\boldsymbol{x}) d\boldsymbol{x} \end{aligned}$$

Rozkład sumy zmiennych losowych.

Niech $\boldsymbol{\xi} = (\xi_1, \xi_2)$ o gęstości $f_{\boldsymbol{\xi}}(x_1, x_2)$. Niech $\eta = \xi_1 + \xi_2$. Wyznaczamy rozkład zmiennej losowej η. Dystrybuanta

$$\begin{aligned} F_\eta(x) &= P\{\xi_1 + \xi_2 \leq x\} \\ &= \int_{-\infty}^{\infty} P\{\xi_1 \leq x - x_2 | \xi_2 = x_2\} f_{\xi_2}(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x-x_2} f_{\xi_1|\xi_2=x_2}(x_1) dx_1\right] f_{\xi_2}(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x} f_{\xi_1|\xi_2=x_2}(t - x_2) dt\right] f_{\xi_2}(x_2) dx_2 \\ &= \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f_{\xi_1|\xi_2=x_2}(t - x_2) f_{\xi_2}(x_2) dx_2\right] dt \end{aligned}$$

Funkcja gęstości

$$f_\eta(x) = \int_{-\infty}^{\infty} f_{\xi_1|\xi_2=x_2}(x - x_2) f_{\xi_2}(x_2) dx_2$$

Jeżeli ξ_1 oraz ξ_2 są niezależne, to

$$f_\eta(x) = \int_{-\infty}^{\infty} f_{\xi_1}(x - x_2) f_{\xi_2}(x_2) dx_2$$

Wartość oczekiwana sumy zmiennych losowych:

$$\begin{aligned} E\eta =& E(\xi_1 + \xi_2) \\ &= \int\int_{\mathbb{R}^2} (x + y) f_{(\xi_1,\xi_2)}(x, y) dx dy \\ &= \int_{\mathbb{R}} x \left[\int_{\mathbb{R}} f_{(\xi_1,\xi_2)}(x, y) dy\right] dx + \int_{\mathbb{R}} y \left[\int_{\mathbb{R}} f_{(\xi_1,\xi_2)}(x, y) dx\right] dy \\ &= \int_{\mathbb{R}} x f_{\xi_1}(x) dx + \int_{\mathbb{R}} y f_{\xi_2}(y) dy \\ &= E\xi_1 + E\xi_2 \end{aligned}$$

Wariancja sumy zmiennych losowych

$$D^2(\xi_1 + \xi_2) = D^2\xi_1 + D^2\xi_2 + 2Cov(\xi_1, \xi_2)$$

Jeżeli ξ_1 oraz ξ_2 są niezależne, to

$$D^2(\xi_1 + \xi_2) = D^2\xi_1 + D^2\xi_2$$

Zadania do samodzielnego rozwiązania

4.1. Niech (ξ, η) będzie dwuwymiarową zmienną losową o funkcji gęstości

$$f(x, y) = \begin{cases} \frac{2}{\pi}, & \text{dla } y > 0 \text{ i } x^2 + y^2 < 1, \\ 0, & \text{poza tym.} \end{cases}$$

Niech $\zeta = \frac{\xi}{\xi^2+\eta^2}$ i $\vartheta = \sqrt{\xi^2 + \eta^2}$. Wykazać, że rozkład brzegowy zmiennej ϑ jest rozkładem $U(0, 1)$.

4.2. Zmienne losowe ξ i η mają łączną gęstość prawdopodobieństwa

$$f(u, v) = \begin{cases} 4/\pi, & \text{dla } u > 0,\ v > 0 \text{ i } u^2 + v^2 \leq 1, \\ 0, & \text{poza tym.} \end{cases}$$

Znaleźć rozkład zmiennej losowej $\frac{\xi^2}{\xi^2+\eta^2}$.

4.3. Wektor losowy (ξ, η) ma łączną gęstość prawdopodobieństwa

$$f(x, y) = \begin{cases} 2, & \text{dla } x > 0,\ y > 0,\ 4x + y < 1, \\ 0, & \text{poza tym.} \end{cases}$$

Podać gęstość rozkładu zmiennej losowej $\frac{\xi}{\xi+\eta}$.

4.4. Dwuwymiarowy wektor losowy (ξ, η) ma gęstość

$$f(x, y) = \begin{cases} \frac{1}{8}, & \text{dla } |x - 2| + |y - 2| < 2, \\ 0, & \text{dla pozostałych } (x, y). \end{cases}$$

Znaleźć gęstość prawdopodobieństwa zmiennej losowej $\xi - \eta$.

4.5. Dwuwymiarowy wektor losowy (ξ, η) ma gęstość

$$f(x, y) = \begin{cases} \frac{1}{2}, & \text{dla } |x| + |y| \leq 1, \\ 0, & \text{dla pozostałych } (x, y). \end{cases}$$

Znaleźć gęstość rozkładu prawdopodobieństwa zmiennej losowej $\xi + \eta$.

4.6. Zmienne losowe ξ_1 oraz ξ_2 są niezależne i mają identyczny rozkład $U(0, 1)$. Niech $\eta_1 = \cos(2\pi\xi_1)f(\xi_2)$ oraz $\eta_2 = \sin(2\pi\xi_1)f(\xi_2)$. Dla jakiej funkcji f zmienne η_1 i η_2 są niezależnymi zmiennymi losowymi o rozkładach $N(0, 1)$?

5. Rozkład dwupunktowy

Zmienna losowa ξ ma **rozkład dwupunktowy** $D(p)$ z parametrem p, jeżeli

$$P\{\xi = 1\} = p = 1 - P\{\xi = 0\}.$$

Doświadczenie Bernoulliego. Wykonujemy dwuwynikowe doświadczenie. Wyniki nazywane są umownie *sukcesem* oraz *porażką*. Prawdopodobieństwo sukcesu wynosi p. Zmienną losową związaną z wynikiem pojedynczego doświadczenia nazywamy **dwupunktową** i oznaczamy $D(p)$.

Zadania do samodzielnego rozwiązania

5.1. Udowodnić, że $E\xi^k = p$ dla $k > 0$.

5.2. Udowodnić, że $D^2\xi = p(1-p)$.

5.3. Zmienne losowe $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ są niezależne i mają rozkład $D\,(0.5)$. Niech $\eta_n = \sum_{i=1}^n \xi_i$. Obliczyć $P(\eta_{10} = 4 \text{ i } \eta_n \leq 6 \text{ dla } n = 1, 2, \ldots, 9)$.

5.4. Zmienne losowe $\xi_1, \ldots, \xi_n, \ldots$ są warunkowo niezależne przy danej wartości $\theta \in (0,1)$ i mają rozkład $D\,(\theta)$. Zmienna losowa θ ma rozkład $Bet\,(3,2)$. Niech $\eta_n = \sum_{i=1}^n \xi_i$. Obliczyć $P(\eta_8 > 0 | \eta_6 = 0)$.

6. Rozkład dwumianowy

Zmienna losowa ξ ma **rozkład dwumianowy** $Bin(n,p)$ z parametrami n oraz p, jeżeli

$$P\{\xi = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Schemat Bernoulliego. Wykonujemy doświadczenie Bernoulliego. Wyniki nazywane są umownie *sukcesem* oraz *porażką*. Prawdopodobieństwo sukcesu wynosi p. Doświadczenie wykonujemy w sposób niezależny n krotnie. Niech zmienną losową ξ będzie ilość sukcesów. Zmienna ξ ma rozkład $Bin(n,p)$.

Zadania do samodzielnego rozwiązania

6.1. Udowodnić, że $P_{n,p}\{\xi = k\} = P_{n,1-p}\{\xi = n-k\}$.

6.2. Udowodnić, że $P_{n,p}\{\xi \geq k\} = \frac{1}{B(k+1,n-k+1)} \int_0^{1-p} x^k (1-x)^{n-k} dx$, tzn. ogon dystrybuanty rozkładu $Bin\,(n,p)$ jest dystrybuantą rozkładu $Bet\,(k+1, n-k+1)$.

6.3. Udowodnić, że $E\xi = np$.

6.4. Udowodnić, że $D^2\xi = np(1-p)$.

6.5. Udowodnić, że jeżeli $\xi_1, \xi_2, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o rozkładach $Bin\,(n_i, p)$, $i = 1, \ldots, n$, to zmienna losowa $\eta = \sum_{i=1}^n \xi_i$ ma rozkład

$$Bin\,(\textstyle\sum_{i=1}^n n_i, p).$$

W szczególności: jeżeli $\xi_1, \xi_2, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o rozkładach $D\,(p)$, $i = 1, \ldots, n$, to $\eta = \sum_{i=1}^n \xi_i \sim Bin\,(n,p)$.

6.6. Co jest bardziej prawdopodobne: wygrać z równorzędnym przeciwnikiem trzy partie z pięciu czy dwie z trzech?

6.7. Załóżmy, że prawdziwa jest hipoteza Mendla, iż dla krzyżówki grochu w drugim pokoleniu stosunek nasion żółtych do zielonych jest jak 3 : 1. Wylosowano dziesięć nasion. Obliczyć prawdopodobieństwo, że będą co najwyżej cztery nasiona żółte.

6.8. Środek owadobójczy zabija przeciętnie 90% owadów. Środek ten zastosowano na dziesięciu owadach. Obliczyć prawdopodobieństwo, że co najwyżej dwa osobniki przeżyją.

6.9. Wadliwość procesu produkcyjnego wynosi 10%. Obliczyć prawdopodobieństwo, że na osiem wylosowanych produktów będą co najwyżej dwa złe.

6.10. W pewnym gatunku zwierząt prawdopodobieństwo urodzenia osobnika płci męskiej wynosi 0.6. Obliczyć prawdopodobieństwo, że w miocie, w którym urodziło się pięcioro młodych będą co najmniej cztery osobniki męskie.

6.11. W stawie hodowlanym są dwa gatunki ryb w proporcji 8 : 2. Obliczyć prawdopodobieństwo, że wśród dziesięciu złowionych ryb będzie co najmniej siedem ryb liczniejszego gatunku.

6.12. W jeziorze jest tysiąc ryb, w tym sto ryb zaobrączkowanych. Obliczyć prawdopodobieństwo, że wśród dziesięciu złowionych ryb będzie co najmniej siedem ryb zaobrączkowanych.

6.13. Właściciel kurzej fermy stwierdził, że kogutków wykluwa się trzy razy więcej niż kurek. Obliczyć prawdopodobieństwo, że z pięciu losowo wybranych jajek wykluje się co najmniej jeden kogutek, ale nie mniej niż dwie kurki.

6.14. Producent podaje, że w co czwartym jajku niespodziance znajduje się zajączek Ribbon. Jakie jest prawdopodobieństwo, że wśród dwudziestu kupionych jajek jest a) przynajmniej pięć jajek z zajączkiem Ribbon; b) nie więcej niż piętnaście jajek bez zajączka. Jaka jest najbardziej prawdopodobna ilość jajek z zajączkami?

7. Rozkład ujemny dwumianowy

Zmienna losowa ξ ma **rozkład ujemny dwumianowy** $NB(r,p)$ z parametrami r oraz p, jeżeli

$$P\{\xi = k\} = \binom{r+k-1}{k} p^r (1-p)^k, \quad k = 0,1,2,\ldots$$

Wykonujemy doświadczenie Bernoulliego. Wyniki nazywane są umownie *sukcesem* oraz *porażką*. Prawdopodobieństwo sukcesu wynosi p. Doświadczenie wykonujemy w sposób niezależny aż do uzyskania r-tego sukcesu. Niech zmienną losową ξ będzie liczba porażek. Zmienna ξ ma rozkład $NB\,(r,p)$.

Rozkład geometryczny. Rozkład $NB(1,p)$ nazywamy **rozkładem geometrycznym**. Rozkład ten oznaczać będziemy $Ge(p)$.

Zadania do samodzielnego rozwiązania

7.1. Udowodnić, że $P_{n,p}\{\xi \le k\} = \frac{1}{B(r,k+1)} \int_0^p x^{r-1}(1-x)^k dx$, tzn. dystrybuanta rozkładu $NB\,(r,p)$ jest dystrybuantą rozkładu $Bet\,(r,k+1)$.

7.2. Udowodnić, że $E\xi = \frac{r(1-p)}{p}$.

7.3. Udowodnić, że $D^2\xi = \frac{r(1-p)}{p^2}$.

7.4. Udowodnić, że jeżeli $\xi_1, \xi_2, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o rozkładach $NB\,(r_i,p)$, $i = 1,\ldots,n$, to zmienna losowa $\eta = \sum_{i=1}^n \xi_i$ ma rozkład

$$NB\,(\textstyle\sum_{i=1}^n r_i, p).$$

W szczególności: jeżeli $\xi_1, \xi_2, \ldots, \xi_n$ są i.i.d. o rozkładach $Ge\,(p)$, $i = 1,\ldots,n$, to $\eta = \sum_{i=1}^n \xi_i \sim NB\,(n,p)$.

7.5. Własność braku pamięci rozkładu geometrycznego: jeżeli η ma rozkład $Ge\,(p)$, to $P\{\eta \ge m+n | \eta \ge m\} = P\{\eta \ge n\}$ dla wszystkich $n \ge 0$.

7.6. Prawdopodobieństwo sukcesu w pojedynczym doświadczeniu wynosi p, gdzie $p \in (0,1)$. Powtarzamy doświadczenie aż do momentu, kiedy po raz trzeci nastąpi sukces. Niech ξ oznacza liczbę porażek, które poprzedziły trzeci sukces. Liczba powtórzeń doświadczenia wynosi więc $\xi + 3$. Przy jakiej wartości parametru p zachodzi: $P(\xi = 1) = P(\xi = 2)$?

7.7. Zmienna losowa ξ ma rozkład z geometrycznym ogonem, tzn. rozkład dany wzorem:

$$P(\xi = k) = \begin{cases} p_0, & \text{dla } k = 0, \\ (1-p_0)p(1-p)^{k-1}, & \text{dla } k = 1,2,\ldots, \end{cases}$$

gdzie $p_0 = 0.5$, $p = 0.25$. Obliczyć wartość oczekiwaną tej zmiennej losowej.

7.8. Niech ξ_1, ξ_2 będą niezależnymi zmiennymi losowymi o tym samym rozkładzie $NB\,(2, 0.75)$. Wyznaczyć $P(\xi_1 = 3 | \xi_1 + \xi_2 = 6)$.

8. Rozkład Poissona

Zmienna losowa ξ ma **rozkład Poissona** $Po(\lambda)$ z parametrem $\lambda > 0$, jeżeli

$$P\{\xi = k\} = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, 2, \ldots.$$

Zadania do samodzielnego rozwiązania

8.1. Udowodnić, że $E\xi = \lambda$.

8.2. Udowodnić, że $D^2\xi = \lambda$.

8.3. Udowodnić, że jeżeli $\xi_1, \xi_2, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o rozkładach $Po(\lambda_i)$, $i = 1, \ldots, n$, to zmienna losowa $\eta = \sum_{i=1}^n \xi_i$ ma rozkład

$$Po\left(\sum_{i=1}^n \lambda_i\right)$$

8.4. Niech ξ będzie zmienną losową o rozkładzie Poissona takim, że $P(\xi \leq 1) = \frac{8}{9}P(\xi = 2)$. Obliczyć $E\xi$.

8.5. Niech ξ_1 i ξ_2 będą niezależnymi zmiennymi losowymi o rozkładach Poissona z wartościami oczekiwanymi $E\xi_1 = 20$ i $E\xi_2 = 30$. Obliczyć $D^2(\xi_1|\xi_1 + \xi_2 = 50)$.

8.6. Załóżmy, że ξ_1, ξ_2, ξ_3 są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $Po(5)$. Obliczyć $D^2(\xi_2 + \xi_3|\xi_1 + \xi_2 = 5)$.

8.7. Zmienna losowa ξ ma rozkład dany wzorem

$$P(\xi = k) = \begin{cases} p_0, & \text{dla } k = 0, \\ \frac{1-p_0}{e^\lambda - 1} \cdot \frac{\lambda^k}{k!}, & \text{dla } k = 1, 2, \ldots, \end{cases}$$

gdzie $p_0 \in (0, 1)$ oraz $\lambda > 0$. Obliczyć wartość oczekiwaną tej zmiennej losowej.

9. Rozkład hipergeometryczny

Zmienna losowa ξ ma **rozkład hipergeometryczny** $H(n, N, M)$ z parametrami n, N oraz M, jeżeli

$$P\{\xi = k\} = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k \in \langle \max\{0, n-(N-M)\}, \min\{n, M\}\rangle\,.$$

Schemat urnowy. Z urny zawierającej N kul, w tym M białych, losujemy bez zwracania n kul. Zmienną losową jest liczba wylosowanych kul białych. Ta zmienna losowa ma rozkład hipergeometryczny $H(n, N, M)$

Zadania do samodzielnego rozwiązania

9.1. Udowodnić, że $E\xi = n\frac{M}{N}$.

9.2. Udowodnić, że $D^2\xi = \frac{N-n}{N-1}n\frac{M}{N}\left(1-\frac{M}{N}\right)$.

10. Rozkład jednostajny

Zmienna losowa ξ ma **rozkład jednostajny** $U(a,b)$ na przedziale (a,b), jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{1}{b-a}\mathbf{1}_{(a,b)}(x).$$

Zadania do samodzielnego rozwiązania

10.1. Narysować wykres funkcji gęstości rozkładu $U(0,1)$.

10.2. Wyznaczyć dystrybuantę rozkładu $U(a,b)$.

10.3. Pokazać, że $E\xi^k = \frac{1}{b-a}\frac{b^{k+1}-a^{k+1}}{k+1}$.

10.4. Pokazać, że $D^2\xi = \frac{(b-a)^2}{12}$.

10.5. Na odcinku $(0,1)$ losujemy punkt zgodnie z rozkładem jednostajnym. W ten sposób odcinek zostaje podzielony na dwa odcinki. Obliczyć wartość oczekiwaną stosunku długości odcinka krótszego do dłuższego.

10.6. Zmienna losowa ξ ma rozkład jednostajny $U(0,1)$, natomiast zależna od niej zmienna η ma rozkład warunkowy (przy danej wartości $\xi = x$) jednostajny $U(0,x)$. Obliczyć prawdopodobieństwo (bezwarunkowe) $P(\eta < 0.5)$.

10.7. Niech ξ_1 i ξ_2 będą niezależnymi zmiennymi losowymi o rozkładzie $U(0,1)$. Rozważmy zmienną losową równą bezwzględnej wartości różnicy zmiennych ξ_1 i ξ_2. Obliczyć wartość oczekiwaną i wariancję tej zmiennej losowej.

10.8. Zmienna losowa ξ_1 ma rozkład $U(0,2)$, a zmienna losowa ξ_2 ma rozkład $U(0,1)$. Zmienne są niezależne. Obliczyć $P(|2\xi_2 - \xi_1| < 0.5)$.

10.9. Dane są dwie niezależne zmienne losowe ξ oraz η o rozkładach $U(-1,1)$ oraz $U(-2,2)$, odpowiednio. Znaleźć gęstość rozkładu prawdopodobieństwa zmiennej losowej $\xi - \eta$.

10.10. Niech $\xi_0, \xi_1, \ldots, \xi_n, \ldots$ będą niezależnymi zmiennymi losowymi o rozkładzie $U(0,1)$. Obliczyć $E(\xi_N - \xi_0)$, gdzie $N = \min\{k : \xi_k > \xi_0\}$.

10.11. Zmienne losowe ξ_1 i ξ_2 są niezależne i mają rozkłady $U(0,2)$. Udowodnić, że $Cov(\max\{\xi_1,\xi_2\}, \min\{\xi_1,\xi_2\}) = \frac{1}{9}$.

10.12. Niech $\xi_1, \ldots, \xi_n$ będą niezależnymi zmiennymi losowymi o rozkładzie $U(a,b)$. Obliczyć współczynnik korelacji liniowej $Corr(\min\{\xi_1,\ldots,\xi_n\}, \max\{\xi_1,\ldots,\xi_n\})$.

10.13. Taśma blachy szerokości jednego metra przepuszczana jest przez gilotynę, która z pewnym błędem tnie ją na kawałki jednometrowej długości. Powstałe kawałki zwija się w cylindry o wysokości jednego metra i nominalnej objętości $1 \cdot \pi \left(\frac{1}{2\pi}\right)^2$ m^3. Cylinder jest klasyfikowany do dalszej produkcji, jeżeli jego faktyczna objętość odchyla się od nominalnej o nie więcej niż 10%. Jaki procent cylindrów klasyfikowanych jest do dalszej produkcji, jeżeli odchylenie (w metrach) rzeczywistego miejsca cięcia od teoretycznego ma rozkład $U(-0.2, 0.2)$?

10.14. Wiadomo, że zmienna losowa ξ ma rozkład $U(0, 1)$, zaś η ma rozkład dyskretny $P(\eta = 1) = P(\eta = -1) = 0.5$. Niech ϱ będzie współczynnikiem korelacji między zmiennymi ξ i η. Jakie są dopuszczalne wartości współczynnika ϱ?

11. Rozkład beta

Zmienna losowa ξ ma **rozkład beta** $Bet(\alpha,\beta)$ z parametrami $\alpha > 0$ oraz $\beta > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\mathbf{1}_{(0,1)}(x).$$

Rozkład $Bet(1,1)$ jest rozkładem $U(0,1)$.

Rozkład $Bet(0.5,0.5)$ nazywany jest rozkładem **arcsinusa**.

Rozkład $Bet(\alpha,1)$ nazywany jest rozkładem **potęgowym** $Pow(\alpha)$.

Zmienna losowa ξ ma **rozkład beta** $Bet(\alpha,\beta;a,b)$ **na przedziale** (a,b) z parametrami $\alpha > 0$ oraz $\beta > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{1}{B(\alpha,\beta)}(x-a)^{\alpha-1}(b-x)^{\beta-1}\mathbf{1}_{(a,b)}(x).$$

Zadania do samodzielnego rozwiązania

11.1. Narysować wykres funkcji gęstości rozkładu $Bet(\alpha,\beta)$.

11.2. Pokazać, że

$$E\xi^k = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+k-1)}.$$

11.3. Pokazać, że $D^2\xi = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

11.4. Udowodnić, że dla $\alpha > 1$ i $\beta > 1$ rozkład $Bet(\alpha,\beta)$ jest jednomodalny. Wyznaczyć modę tego rozkładu.

12. Rozkład gamma

Zmienna losowa ξ ma **rozkład Gamma** $G(\alpha, \lambda)$ z parametrami $\alpha > 0$ oraz $\lambda > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left\{-\frac{x}{\lambda}\right\} \mathbf{1}_{(0,\infty)}(x).$$

Rozkład wykładniczy. Rozkład $G(1, \lambda)$ nazywamy **rozkładem wykładniczym**. Rozkład ten oznaczać będziemy $E(\lambda)$.

Rozkład chi-kwadrat. Rozkład $G(\alpha/2, 2)$ nazywamy **rozkładem chi-kwadrat** z α stopniami swobody. Rozkład ten oznaczać będziemy $\chi^2(\alpha)$.

Rozkład chi-kwadrat ma związek z jednowymiarowym rozkładem normalnym, a także z wielowymiarowym rozkładem normalnym.

Wartością krytyczną rzędu α rozkładu $\chi^2(r)$ nazywamy kwantyl rzędu $1 - \alpha$. Oznaczenie: $\chi^2(\alpha; r)$.

Zmienna losowa ξ ma **przesunięty rozkład gamma** $G(\alpha, \lambda; a)$ **na przedziale** (a, ∞) z parametrami $\alpha > 0$ oraz $\lambda > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} (x-a)^{\alpha-1} \exp\left\{-\frac{x-a}{\lambda}\right\} \mathbf{1}_{(a,\infty)}(x).$$

Rozkład $G(1, \lambda; a)$ nazywamy **przesuniętym rozkładem wykładniczym** i oznaczamy przez $E(\lambda; a)$.

Zadania do samodzielnego rozwiązania

12.1. Narysować wykres funkcji gęstości rozkładu $G(\alpha, \lambda)$.

12.2. Pokazać, że $E\xi^k = \alpha(\alpha+1)\cdots(\alpha+k-1)\lambda^k$

12.3. Pokazać, że $D^2\xi = \alpha\lambda^2$.

12.4. Udowodnić, że dla $\alpha > 1$ rozkład gamma jest jednomodalny. Wyznaczyć modę tego rozkładu.

12.5. Wyznaczyć dystrybuantę rozkładu wykładniczego $E(\lambda)$. Narysować jej wykres.

12.6. Udowodnić, że jeżeli $\xi_1, \xi_2, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o rozkładach $G(\alpha_i, \lambda)$, $i = 1, \ldots, n$, to zmienna losowa $\eta = \sum_{i=1}^n \xi_i$ ma rozkład

$$G\left(\sum_{i=1}^n \alpha_i, \lambda\right)$$

Jeżeli $\xi_1, \xi_2, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o takich samych rozkładach $E(\lambda)$, $i = 1, \ldots, n$, to $\eta = \sum_{i=1}^n \xi_i \sim G(n, \lambda)$.

12.7. **Własność braku pamięci rozkładu wykładniczego**: jeżeli η ma rozkład $E(\lambda)$, to $P\{\eta \geq x + t | \eta \geq x\} = P\{\eta \geq t\}$ dla wszystkich $x \geq 0$.

12.8. Zmienna losowa ξ ma rozkład wykładniczy o wartości oczekiwanej 0.5. Niezależna zmienna losowa η ma rozkład wykładniczy o wartości oczekiwanej 2. Obliczyć $E(\xi|\xi + \eta = 5)$.

12.9. $\xi_1, \ldots, \xi_{10}$ są niezależnymi zmiennymi losowymi o rozkładzie $E(0.2)$. Wiadomo, że $P(\max\{\xi_1, \ldots, \xi_{10}\} \leq x) = 0.95$. Obliczyć x.

12.10. Mamy dwie niezależne zmienne losowe ξ oraz η. Jedna nich (nie wiadomo która) ma rozkład wykładniczy z wartością oczekiwaną równą 1, druga zaś ma rozkład wykładniczy z wartością oczekiwaną równą 2. Obliczyć wartość ilorazu

$$\frac{E\max\{\xi, \eta\}}{E\min\{\xi, \eta\}.}$$

12.11. Mamy dwie niezależne zmienne losowe ξ oraz η. Zmienna ξ ma rozkład wykładniczy $E(1)$, zmienna η zaś ma rozkład wykładniczy z wartością oczekiwaną równą 2. Zdefiniujmy nową zmienną ζ jako udział zmiennej ξ w sumie obu zmiennych: $\zeta = \frac{\xi}{\xi+\eta}$. Obliczyć medianę zmiennej losowej ζ.

12.12. Załóżmy, że zmienna losowa ξ ma rozkład wykładniczy $E(\lambda)$. Wyznaczyć wartość oczekiwaną zmiennej losowej $\eta = \lfloor \xi + 0.5 \rfloor$.

12.13. Niech ξ będzie zmienną losową o rozkładzie $E(1)$. Obliczyć współczynnik korelacji liniowej $Corr(\lceil \xi \rceil, \lfloor \xi \rfloor)$.

12.14. Zmienna losowa ξ ma rozkład $E(\lambda)$. Obliczyć $E\lceil \xi \rceil$ w zależności od $E(\lfloor \xi \rfloor)$.

12.15. Załóżmy, że ξ_1, ξ_2, ξ_3 są niezależnymi zmiennymi losowymi o jednakowym rozkładzie wykładniczym. Niech $\eta = \xi_1 + \xi_2 + \xi_3$. Obliczyć $P(\xi_1 > \eta/2$ lub $\xi_2 > \eta/2$ lub $\xi_3 > \eta/2)$.

12.16. Niezależne zmienne losowe ξ_1, ξ_2 mają identyczny rozkład $E(\mu)$. Obliczyć warunkową wartość oczekiwaną $E(\min\{\xi, \xi_2\}|\xi_1 + \xi_2 = M)$, gdzie M jest pewną dodatnią liczbą.

12.17. Załóżmy, że ξ_1, ξ_2 są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $E(\mu)$. Obliczyć $E(\xi_1 | \min\{\xi_1, \xi_2\})$.

12.18. Niech $\xi_0, \xi_1, \ldots, \xi_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie $E(1)$. Obliczyć warunkową wartość oczekiwaną $E(\min\{\xi_0, \xi_1, \ldots, \xi_n\}|\xi_0)$.

12.19. Zmienne losowe ξ i η są niezależne o rozkładach $E(\lambda)$. Obliczyć warunkową wariancję $D^2(\min\{\xi, \eta\}|\xi + \eta = 2)$.

12.20. Załóżmy, że niezależne zmienne losowe $\xi_1, \xi_2, \xi_3, \xi_4$ mają rozkłady wykładnicze o wartościach oczekiwanych odpowiednio 1, 2, 3 i 4. Obliczyć prawdopodobieństwo $P(\xi_1 = \min\{\xi_1, \xi_2, \xi_3, \xi_4\})$.

12.21. Niech ξ i η będą niezależnymi zmiennymi losowymi o rozkładach wykładniczych, przy czym $E\xi = 4$ i $E\eta = 6$. Wyznaczyć medianę zmiennej losowej $\frac{\xi}{\xi+\eta}$.

12.22. Dane są dwie niezależne zmienne losowe ξ oraz η o rozkładach $E\,(2)$ oraz $E\,(4)$, odpowiednio. Znaleźć gęstość rozkładu prawdopodobieństwa zmiennej losowej $\xi + 2\eta$.

12.23. Niech ξ_1, ξ_2, ξ_3 będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $E\,(\lambda)$. Obliczyć medianę zmiennej losowej $\frac{\xi_1}{\xi_2+\xi_3}$.

12.24. Załóżmy, że $\xi_1, \ldots, \xi_n, \ldots$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie wykładniczym, $E(\xi_n) = 1/\lambda$. Niech $\eta_0 = 0$ i $\eta_n = \sum_{i=1}^n \xi_i$ dla $n = 1, 2, \ldots$. Załóżmy, że ζ jest zmienną losową o rozkładzie wykładniczym, $E\zeta = 1/\alpha$ i niezależną od zmiennych ξ_i. Niech $N = \max\{n \geq 0 : \eta_n \leq \zeta\}$. Podać rozkład prawdopodobieństwa zmiennej losowej N.

12.25. Załóżmy, że $\xi_1, \ldots, \xi_n, \ldots$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie $E\,(\lambda)$. Niech $\eta_0 = 0$ i $\eta_n = \sum_{i=1}^n \xi_i$ dla $n = 1, 2, \ldots$ Niech ζ będzie zmienną losową niezależną od zmiennych $\xi_1, \ldots, \xi_n, \ldots$ o rozkładzie $G\,(2, \beta)$, gdzie $\beta > 0$ jest ustaloną liczbą. Niech

$$N = \max\{n \geq 0 : \eta_n \leq \zeta\}.$$

Podać rozkład prawdopodobieństwa zmiennej N.

12.26. Załóżmy, że $\xi_0, \xi_1, \ldots, \xi_n, \ldots$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie $E\,(\lambda)$. Niech

$$N = \min\left\{k \geq 0 : \sum_{i=1}^{k} \xi_i > a\right\},$$

gdzie a jest ustaloną liczbą dodatnią. Podać rozkład prawdopodobieństwa zmiennej losowej N.

12.27. Zmienna losowa ξ ma rozkład wykładniczy $E\,(0.2)$. Zmienna losowa η ma rozkład jednostajny na pewnym odcinku, przy czym $E\eta = 5$ oraz $D^2\eta = 25/3$. Zmienne losowe ξ i η są niezależne. Obliczyć $P(\xi + \eta < 6)$.

12.28. Jabłko upada od jabłoni w odległości, która jest zmienną losową o rozkładzie $E\,(1/2)$ (pomijamy średnicę pnia i średnicę jabłka). Jabłko może spadać w każdym kierunku z tym samym prawdopodobieństwem. Jaka jest wartość oczekiwana odległości dwóch jabłek, które spadły niezależnie pod warunkiem, że obydwa upadły w tej samej odległości od jabłoni?

12.29. W wyniku eksperymentu pewna populacja została pozbawiona zdolności rozmnażania. Wiadomo, że czas życia każdego przedstawiciela tej populacji jest zmienną losową o rozkładzie wykładniczym o średniej 30 dni. Po ilu dniach od eksperymentu liczebność tej populacji spadnie o połowę?

12.30. Czas życia żarówki jest zmienną losową o rozkładzie wykładniczym o średniej 1000 h. Jaki jest rozkład czasu pracy układu złożonego z dwóch równolegle połączonych żarówek? Jakie jest prawdopodobieństwo, że układ przepracuje co najmniej 1500 h?

12.31. Czas życia żarówki jest zmienną losową o rozkładzie wykładniczym o średniej 1000 h. Jaki jest rozkład czasu pracy układu złożonego z dwóch szeregowo połączonych żarówek? Jakie jest prawdopodobieństwo, że układ przepracuje co najmniej 1500 h?

13. Rozkład normalny

Zmienna losowa ξ ma **rozkład normalny** $N(\mu, \sigma^2)$, jeżeli funkcja gęstości jej rozkładu prawdopodobieństwa wyraża się wzorem

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}.$$

Standaryzacja. Jeżeli ξ ma rozkład $N(\mu, \sigma^2)$, to zmienna losowa $\eta = (\xi - \mu)/\sigma$ ma rozkład $N(0,1)$ zwany **standardowym rozkładem normalnym**.

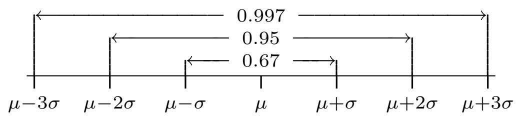
Gęstość i dystrybuantę rozkładu $N(0,1)$ oznaczamy odpowiednio przez $\phi(\cdot)$ oraz $\Phi(\cdot)$.

Dla zmiennej losowej o rozkładzie $N\left(\mu, \sigma^2\right)$ zachodzi **prawo trzech sigm**

$P(|X-\mu| < \sigma) \approx 0.67$

$P(|X-\mu| < 2\sigma) \approx 0.95$

$P(|X-\mu| < 3\sigma) \approx 0.997$



Zadania do samodzielnego rozwiązania

13.1. Narysować wykres funkcji gęstości rozkładu $N\left(\mu, \sigma^2\right)$.

13.2. Pokazać, że $E\xi = \mu$.

13.3. Pokazać, że

$$E(\xi - E\xi)^k = \begin{cases} 0, & \text{dla } k \text{ nieparzystych,} \\ 1 \cdot 3 \cdots (k-1)\sigma^k, & \text{dla } k \text{ parzystych.} \end{cases}$$

13.4. Udowodnić, że rozkład normalny jest jednomodalny. Wyznaczyć modę tego rozkładu.

13.5. Udowodnić, że $\Phi(x) - 0.5$ jest funkcją nieparzystą.

13.6. Pokazać, że $P(|\xi - \mu| \le a\sigma) = 2\Phi(a) - 1$ dla każdego $a > 0$. Udowodnić prawo trzech sigm.

13.7. Pokazać, że

$$P(\xi \in (a,b)) = \Phi\left(\tfrac{b-\mu}{\sigma}\right) - \Phi\left(\tfrac{a-\mu}{\sigma}\right).$$

13.8. Jeżeli $\xi_1, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o rozkładach $N\left(\mu_i, \sigma_i^2\right)$, $i = 1, \ldots, n$, to zmienna losowa $\eta = \sum_{i=1}^n \xi_i$ ma rozkład

$$N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

13.9. Jeżeli $\xi_1, \ldots, \xi_n$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0,1)$, to zmienna losowa $\xi = \sum_{i=1}^n \xi_i^2$ ma rozkład $\chi^2(n)$.

13.10. Niech ξ będzie zmienną losową o rozkładzie $N(-5, 100)$. Obliczyć:

$$P(\xi \le -9),\ P(\xi \in (-7, 1)),\ P(\xi \ge -7),\ P(|\xi + 5| \le 10).$$

13.11. Niech ξ będzie zmienną losową o rozkładzie $N(10, 25)$. Obliczyć:

$$P(\xi \le 8),\ P(\xi \in (9, 13)),\ P(\xi \ge 9),\ P(|\xi - 10| \le 5).$$

13.12. Zmienne losowe $\xi_1, \xi_2, \xi_3, \xi_4$ są niezależne i mają jednakowy rozkład $N\left(0, \sigma^2\right)$. Obliczyć $P(\xi_1^2 - 5\xi_2^2 < 5\xi_3^2 - \xi_4^2)$.

13.13. Zakładamy, że $\xi_1, \ldots, \xi_{20}$ są niezależnymi zmiennymi losowymi o rozkładzie $N\left(\mu, \sigma^2\right)$. Niech $\eta = \xi_1 + \cdots + \xi_{15}$ i $\zeta = \xi_6 + \cdots + \xi_{20}$. Obliczyć $E(\zeta|\eta)$.

13.14. Zmienne losowe ξ_1, ξ_2, ξ_3 mają łączny rozkład normalny taki, że $E\xi_i = 0$, $D^2\xi_i = 1$ dla $i = 1, 2, 3$. Załóżmy, że $Cov(\xi_1, \xi_2) = Cov(\xi_2, \xi_3) = Cov(\xi_1+\xi_2, \xi_2+\xi_3) = 0$. Udowodnić, że $P(\xi_1 = -\xi_3) = 1$.

13.15. Zmienne losowe $\xi_1, \ldots, \xi_{20}$ są niezależne o rozkładzie $N(10, 0.01)$. Wyznaczyć taką liczbę a, że $P(\max\{\xi_1, \ldots, \xi_{20}\} \le a) = 0.99$.

13.16. Niech $\xi_1, \ldots, \xi_n$ będą niezależnymi zmiennymi losowymi o identycznym rozkładzie $N(0, 1)$. Niech $\eta = (\xi_1 + \cdots + \xi_n)^2$. Obliczyć wariancję zmiennej losowej η.

13.17. Zmienne losowe $\xi_1, \ldots, \xi_{25}$ są niezależne o jednakowym rozkładzie $N\left(\mu, \sigma^2\right)$. Niech $\eta_{10} = \sum_{i=1}^{10} \xi_i$ i $\eta_{25} = \sum_{i=1}^{25} \xi_i$. Wyznaczyć $E(\eta_{10}|\eta_{25})$.

13.18. Wzrost kobiety jest zmienną losową o rozkładzie $N(158, 100)$. Obliczyć jaki jest procent kobiet o wzroście pomiędzy 148 a 168.

13.19. Ciężar jajek dostarczanych do skupu ma rozkład normalny ze średnią 2 dag i wariancją 0.01. Jajko kwalifikuje się do pierwszego gatunku, jeżeli jego waga wynosi co najmniej 2.096 dag. Jaki procent jajek dostarczanych do skupu można uznać za jajka pierwszego gatunku?

13.20. Średni czas żarzenia się liści tytoniu wynosi 17 sekund. Liście tlące się krócej niż 12 sekund są dyskwalifikowane. Jaki jest procent liści przydatnych do produkcji, jeżeli wariancja wspomnianej cechy wynosi 6.25?

13.21. Przyjmując, że przeciętna waga (w kilogramach) noworodka jest zmienną losową o rozkładzie $N(3, 0.25)$ określić procent noworodków o wadze z przedziału $(3, 3.5)$.

13.22. Stwierdzono, że 80% ludzi o IQ powyżej 90 jest w stanie nauczyć się posługiwać pewnym urządzeniem. Jaki jest to procent całej populacji jeśli wiadomo, że iloraz inteligencji jest cechą o rozkładzie $N(100, 100)$?

13.23. Aby zdać egzamin ze statystyki należy prawidłowo rozwiązać co najmniej 70% zadań egzaminacyjnych. Przyjmując, że wyniki testu Studentów zdających w pierwszym terminie mają rozkład normalny ze średnią 76% i odchyleniem standardowym 8.0%, obliczyć jaki procent Studentów zda egzamin w pierwszym terminie.

13.24. Automat tokarski produkuje nity, których średnica ma rozkład normalny z odchyleniem standardowym 0.04 mm. Wartość średnia tej zmiennej losowej może być dowolnie regulowana przez odpowiednie ustawienie automatu. Nit uważa się za dobry, jeżeli jego średnica mieści się w przedziale $(2.9, 3.1)$. Jakie jest prawdopodobieństwo wyprodukowania braku, gdy automat tokarski ustawiony jest tak, że średnia średnica jest równa 3.05 mm? Jak powinien być ustawiony automat, by wadliwość procesu produkcyjnego była najmniejsza?

13.25. Automat produkuje nity, których średnica jest zmienną losową o rozkładzie $N(2, 0.01)$. Dla jakiego ε średnica nitu z prawdopodobieństwem 0.95 znajduje się w przedziale $(2 - \varepsilon, 2 + \varepsilon)$?

13.26. Obliczyć odchylenie standardowe przyrządu pomiarowego o którym wiadomo, że z prawdopodobieństwem 0.95 daje błąd nie przekraczający trzech jednostek. Zakładamy, że rozkład błędu jest normalny z wartością średnią zero.

13.27. Cecha ξ o rozkładzie normalnym przyjmuje wartości z przedziału $\langle 28, 46\rangle$ (przedział ten ustalony został na mocy prawa trzech sigm). Wartości $40-46$ zaliczamy do pierwszej klasy, wartości $28 - 34$ do trzeciej klasy. Jakie jest prawdopodobieństwo tego, że losowo wybrana wartość cechy ξ należy do pierwszej lub trzeciej klasy?

13.28. Błąd pomiaru pewnym urządzeniem ma rozkład normalny o wartości średniej zero (pomiar nieobciążony) i odchyleniu standardowym 20 mm. Ile co najmniej należy wykonać pomiarów, by z prawdopodobieństwem co najmniej 0.99 błąd przynajmniej jednego z nich nie przekroczył 4 mm.

13.29. Zawartość tłuszczu w mleku pewnej rasy krów ma rozkład normalny o wartości średniej 5% i wariancji 4. Mleko uważa się za bardzo tłuste, jeżeli zawartość tłuszczu przekracza 7%. Obliczyć prawdopodobieństwo, że przynajmniej jedna z trzech niezależnych próbek mleka będzie uznana za bardzo tłustą.

13.30. Wzrost dzieci jest zmienną losową o rozkładzie normalnym o wartości średniej 110 cm i wariancji 400. Obliczyć prawdopodobieństwo, że przynajmniej jedno z trójki losowo wybranych dzieci będzie miało wzrost większy od przeciętnej.

13.31. Średnica nitu ma rozkład normalny o wartości oczekiwanej 2 mm i wariancji 0.01. Obliczyć prawdopodobieństwo, że wśród trzech losowo pobranych nitów wszystkie okażą się brakami, jeżeli za brak uważany jest nit o średnicy mniejszej niż 1.8 mm lub większej niż 2.2 mm.

13.32. Ciężar jajka kurzego jest zmienną losową o rozkładzie $N(10, 4)$. Obliczyć prawdopodobieństwo, że z trzech losowo wybranych jajek wszystkie będą miały ciężar poniżej przeciętnej.

14. Inne rozkłady prawdopodobieństwa

Rozkład Laplace'a.

Zmienna losowa ξ ma **rozkład Laplace'a** (**podwójnie wykładniczy**) $Lap(\alpha, \lambda)$ z parametrami $\alpha \in \mathbb{R}$ oraz $\lambda > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{1}{2\lambda} \exp\left\{-\frac{|x-\alpha|}{\lambda}\right\}, \ x \in \mathbb{R}.$$

14.1. Narysować wykres funkcji gęstości zmiennej losowej ξ.

14.2. Wyznaczyć dystrybuantę zmiennej losowej ξ.

14.3. Udowodnić, że

$$E\xi^{(2k+1)} = \alpha^{2k+1}(2k+1)!, \ E\xi^{2k} = \left[\frac{\alpha^{2k}}{(2k)!} + \frac{\alpha^{2(k-1)}\lambda^2}{(2(k-1))!} + \cdots + \lambda^{2k}\right].$$

14.4. Udowodnić, że $D^2\xi = 2\lambda^2$.

14.5. Pokazać, że jeżeli ξ_1 oraz ξ_2 są niezależnymi zmiennymi losowymi o rozkładach $E\,(\lambda)$, to zmienna losowa $\zeta = \xi_1 - \xi_2 + \alpha$ ma rozkład $Lap\,(\alpha, \lambda)$.

Rozkład Weibulla.

Zmienna losowa ξ ma **rozkład Weibulla** $Wei(\alpha, \lambda)$ z parametrami $\alpha > 0$ oraz $\lambda > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{\alpha}{\lambda} x^{\alpha-1} \exp\left\{-\frac{x^\alpha}{\lambda}\right\} \mathbf{1}_{(0,\infty)}(x).$$

14.6. Narysować wykres funkcji gęstości zmiennej losowej ξ.

14.7. Wyznaczyć dystrybuantę zmiennej losowej ξ.

14.8. Udowodnić, że $E\xi^k = \lambda^{\frac{k}{\alpha}} \Gamma\left(\frac{k}{\alpha} + 1\right)$.

14.9. Udowodnić, że $D^2\xi = \lambda^{\frac{2}{\alpha}} \left[\frac{2}{\alpha}\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha^2}\left(\Gamma\left(\frac{1}{\alpha}\right)\right)\right]$.

Rozkład Pareto.

Zmienna losowa ξ ma **rozkład Pareto** $Par(\alpha, \beta)$ z parametrami $\alpha > 0$ oraz $\beta > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{\beta}{\alpha}\left(\frac{\alpha}{x}\right)^{\beta+1}\mathbf{1}_{(\alpha,\infty)}(x).$$

Funkcja gęstości rozkładu $Par\,(0, \beta)$ określona jest wzorem

$$f(x) = \beta x^{-(\beta+1)}\mathbf{1}_{(0,\infty)}(x).$$

14.10. Narysować wykres funkcji gęstości zmiennej losowej ξ.

14.11. Wyznaczyć dystrybuantę zmiennej losowej ξ.

14.12. Udowodnić, że $E\xi^k = \frac{\beta}{\beta-k}\alpha^k$, $k < \beta$.

14.13. Udowodnić, że $D^2\xi = \frac{\beta}{(\beta-1)(\beta-2)}\alpha^2$, dla $\beta > 2$ oraz $D^2\xi = +\infty$ dla $\beta \le 2$.

Rozkład logarytmiczno-normalny.

Zmienna losowa ξ ma **rozkład logarytmiczno-normalny** $LNor(\mu, \sigma^2)$ z parametrami $\mu \in \mathbb{R}$ oraz $\sigma^2 > 0$, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right\} \mathbf{1}_{(0,\infty)}(x).$$

14.14. Narysować wykres funkcji gęstości zmiennej losowej ξ.

14.15. Udowodnić, że $E\xi^k = \exp\left\{\frac{1}{2}k^2\sigma^2 + k\mu\right\}$.

14.16. Udowodnić, że $D^2\xi = \exp\left\{\sigma^4 + m\right\}\left[\exp\{\sigma^4\} - 1\right]$.

14.17. Pokazać, że zmienna losowa ξ ma rozkład logarytmiczno-normalny, jeżeli zmienna losowa $\eta = e^\xi$ ma rozkład normalny.

14.18. Zmienna losowa ξ ma rozkład $LNor\,(\mu, 1)$. Wiadomo, że $P(\xi \le q) = 0.6$ oraz $P(\xi \le r) = 0.4$. Obliczyć wartość oczekiwaną zmiennej losowej ξ.

Rozkład t (Studenta).

Zmienna losowa ξ ma **rozkład t (Studenta)** $t(r)$ z $r>0$ stopniami swobody, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x)=\frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{r\pi}\,\Gamma\left(\frac{r}{2}\right)}\left(1+\frac{x^2}{2}\right)^{-\frac{r+1}{2}},\ x\in\mathbb{R}.$$

Wartością krytyczną rzędu α rozkładu $t\,(r)$ nazywamy kwantyl rzędu $1-\alpha$.

Dwustronną wartością krytyczną rzędu α rozkładu $t\,(r)$ nazywamy kwantyl rzędu $1-(\alpha/2)$. Oznaczenie: $t(\alpha;r)$.

14.19. Narysować wykres funkcji gęstości zmiennej losowej ξ.

14.20. Udowodnić, że $E\xi^{2k-1}=0$ oraz $E\xi^{2k}=\frac{r^k}{\sqrt{\pi}}\frac{\Gamma\left(\frac{r}{2}-k\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)}$ dla $2k<r$.

14.21. Udowodnić, że $D^2\xi=\frac{r}{r-2}$ dla $r>2$ oraz $D^2\xi=+\infty$ dla $r\le 2$.

14.22. Pokazać, że jeżeli zmienna losowa ξ ma rozkład $N\,(0,1)$, zmienna losowa η ma rozkład $\chi^2\,(r)$ oraz zmienne te są niezależne, to zmienna losowa $\zeta=\xi/\sqrt{\eta/r}$ ma rozkład $t\,(r)$.

14.23. Pokazać, że jeżeli zmienna losowa ξ ma rozkład $t\,(r)$, to $P\{|\xi|>t(\alpha;r)\}=\alpha$.

Rozkład F (Snedecora).

Zmienna losowa ξ ma **rozkład F (Snedecora)** $F(r,s)$ z $r > 0$ i $s > 0$ stopniami swobody, jeżeli jej funkcja gęstości rozkładu prawdopodobieństwa określona jest wzorem

$$f(x) = \frac{\Gamma\left(\frac{r+s}{2}\right) r^{\frac{r}{2}} s^{\frac{s}{2}}}{\Gamma\left(\frac{r}{2}\right)\Gamma\left(\frac{s}{2}\right)} x^{\frac{s}{2}-1}(s+rx)^{-\frac{r+s}{2}} \mathbf{1}_{(0,\infty)}(x).$$

Wartością krytyczną rzędu α rozkładu $F(r,s)$ nazywamy kwantyl rzędu $1-\alpha$. Oznaczenie: $F(\alpha; r, s)$.

14.24. Narysować wykres funkcji gęstości zmiennej losowej ξ.

14.25. Udowodnić, że $E\xi^k = \frac{\Gamma\left(\frac{r}{2}+k\right)\Gamma\left(\frac{s}{2}-k\right)s^k}{\Gamma\left(\frac{r}{2}\right)r^k\Gamma\left(\frac{s}{2}\right)}$ dla $2k < s$.

14.26. Udowodnić, że $E\xi = \frac{s}{s-2}$ dla $s > 2$.

14.27. Udowodnić, że $D^2\xi = \frac{2s^2(r+s-2)}{r(s-2)^2(s-4)}$ dla $s > 4$.

14.28. Pokazać, że jeżeli zmienna losowa ξ ma rozkład $\chi^2(r)$ stopniami swobody, zmienna losowa η ma rozkład $\chi^2(s)$ stopniami swobody i zmienne te są niezależne, to zmienna losowa $\zeta = (\xi/r)/(\eta/s)$ ma rozkład $F(r,s)$.

14.29. Pokazać, że jeżeli $\xi \sim F(2r, 2s)$, to $\frac{1}{1+\xi} \sim Bet(s,r)$.

Rozkład z Fishera.

Zmienna losowa ξ ma **rozkład z Fishera** $Fish(r,s)$ z (r,s) stopniami swobody, jeżeli jej funkcja gęstości wyraża się wzorem

$$f(x)=\frac{2r^{\frac{r}{2}}s^{\frac{s}{2}}\Gamma\left(\dfrac{r+s}{2}\right)e^{rx}}{\Gamma\left(\dfrac{r}{2}\right)\Gamma\left(\dfrac{s}{2}\right)(s+re^{2x})^{\frac{r+s}{2}}},\ x\in R$$

14.30. Narysować wykres funkcji gęstości zmiennej losowej ξ.

14.31. Udowodnić, że $E\xi=0$.

14.32. Udowodnić, że $D^2\xi=\frac{1}{2}\frac{r+s}{rs}$.

14.33. Pokazać, że jeżeli $\xi\sim F\left(r,s\right)$, to $\zeta=\frac{1}{2}\ln\xi$ ma rozkład $Fish\left(r,s\right)$.

Rozkład wielomianowy.

Wektor losowy $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ ma **rozkład wielomianowy** z parametrami n oraz $(p_1, \ldots, p_k)$, jeżeli

$$P\{\boldsymbol{\xi} = \boldsymbol{m}\} = P\{\xi_1 = m_1, \ldots, \xi_k = m_k\} = \frac{n!}{m_1! \cdots m_k!} p_1^{m_1} \cdots p_k^{m_k}$$

gdzie $\boldsymbol{m} = (m_1, \ldots, m_k)$, $\sum_{i=1}^k m_k = n$, $0 < p_i < 1$, $\sum_{i=1}^k p_k = 1$.

Rozkład wielomianowy jest wielowymiarowym analogiem rozkładu dwumianowego.

14.34. Udowodnić, że $E\boldsymbol{\xi} = n(p_1, \ldots, p_k)$.

14.35. Udowodnić, że $D^2\xi_i = np_i(1 - p_i)$, $Cov(\xi_i, \xi_j) = -np_ip_j$, dla $i \neq j$.

14.36. Pokazać, że zmienna losowa ξ_i ma rozkład $Bin\,(n, p_i)$.

14.37. Pokazać, że jeżeli $\boldsymbol{\xi}^{(1)}, \ldots, \boldsymbol{\xi}^{(k)}$ są niezależnymi k–wymiarowymi wektorami losowymi o rozkładach wielomianowych z parametrami $(n_1, \boldsymbol{p}), \ldots, (n_k, \boldsymbol{p})$, to $\boldsymbol{\xi} = \sum \boldsymbol{\xi}^{(i)}$ ma rozkład wielomianowy z parametrami $(\sum n_i, \boldsymbol{p})$.

Rozkład Dirichleta.

Wektor losowy $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ ma **rozkład Dirichleta** $Dir(\boldsymbol{\alpha})$ z parametrem $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ $(\alpha_i > 0)$, jeżeli funkcja gęstości wyraża się wzorem

$$f(\boldsymbol{x}) = \begin{cases} \frac{\Gamma(\alpha_1 + \cdots + \alpha_k)}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_k)} x_1^{\alpha_1} \cdots x_k^{\alpha_k}, & x \in \mathcal{S} \\ 0, & x \notin \mathcal{S} \end{cases}$$

gdzie $\mathcal{S} = \{\boldsymbol{x} \in R^k : \sum_{i=1}^k x_k = 1, x_i > 0\}$.

Rozkład Dirichleta jest wielowymiarowym analogiem rozkładu beta

14.38. Udowodnić, że

$$E\boldsymbol{\xi} = \frac{\boldsymbol{\alpha}}{\alpha_0}, \quad Cov(\xi_i, \xi_j) = -\frac{\alpha_i \alpha_j}{\alpha_0^2(1+\alpha_0)} (i \neq j), \quad D^2\xi_i = \frac{\alpha_i}{\alpha_0} \quad \left(\alpha_0 = \sum_{i=1}^k \alpha_i\right).$$

.

Zadania do samodzielnego rozwiązania

14.39. Zmienne losowe ξ i η są niezależne. Zmienna losowa ξ ma rozkład normalny $N(0, 0.5)$. Zmienna losowa η ma rozkład wykładniczy $E(1)$. Obliczyć $P(\eta > \xi^2)$.

14.40. Zmienne losowe $\xi_1, \ldots, \xi_n, \ldots$ są niezależne i mają jednakowy rozkład wykładniczy $E(1)$. Zmienna losowa N jest niezależna od nich i ma rozkład geometryczny $Ge(q)$. Niech $S = \sum_{i=1}^{N} \xi_i$ będzie sumą losowej liczby zmiennych losowych (przyjmujemy, że $S = 0$, gdy $N = 0$). Udowodnić, że $D^2(N|S = s) = E(N|S = s) - 1$.

14.41. Niech ξ i η będą zmiennymi losowymi takimi, że ξ ma rozkład $E(1)$, warunkowy (pod warunkiem $\xi = x$) rozkład zmiennej losowej η jest rozkładem $Po(x)$. Udowodnić, że zmienne losowe ξ i $\eta - \xi$ są nieskorelowane.

14.42. Zmienna losowa ξ ma rozkład warunkowy (pod warunkiem $\Lambda = \lambda$) jest rozkładem $E(1/\lambda)$. Rozkład brzegowy zmiennej losowej Λ jest rozkładem $G(2, 0.5)$. Wyznaczyć medianę rozkładu bezwarunkowego zmiennej ξ.

14.43. Niech $\xi = N \exp(t\zeta)$, gdzie N jest zmienną losową o rozkładzie $Po(\lambda)$, ζ jest zmienną losową o rozkładzie $N(\mu, \sigma^2)$, niezależną od N, natomiast t jest stałą. Obliczyć $\frac{D^2\xi}{(E\xi)^2}$.

14.44. Załóżmy, że ξ i η są niezależnymi zmiennymi losowymi o rozkładzie $N(0, 1)$. Zmienna losowa ζ jest równa $\zeta = \frac{|\xi|}{\sqrt{\xi^2+\eta^2}}$. Wyznaczyć funkcję gęstości zmiennej losowej η.

14.45. Zmienne losowe ξ i η są niezależne i każda z nich ma rozkład prawdopodobieństwa o gęstości

$$f(x) = \begin{cases} \frac{4}{(1+x)^5}, & \text{dla } x > 0, \\ 0, & \text{poza tym.} \end{cases}$$

Rozważamy zmienną losową $\zeta = \frac{\ln \xi}{\ln[(1+\xi)(1+\eta)]}$. Udowodnić, że zmienna losowa ζ ma rozkład $U(0, 1)$.

14.46. Niech $N_1 = \sum_{i=1}^{N} \xi_i$ i $N_0 = N - N_1$, gdzie N jest zmienną losową o rozkładzie $Po(\lambda)$, zaś $\xi_1, \ldots, \xi_n, \ldots$ są zmiennymi losowymi niezależnymi od N i od siebie nawzajem. Zakładamy, że każda ze zmiennych ξ_i ma rozkład $D(p)$. Obliczyć $E\left[\frac{N_1}{N_0+1}\right]$.

14.47. Rozważamy kolektywny model ryzyk. Zakładamy, że $\eta = \eta_N = \sum_{i=1}^{N} \xi_i$, gdzie N oraz $\xi_1, \xi_2, \ldots$ są niezależnymi zmiennymi losowymi, przy czym N ma rozkład $Po(\lambda)$, zaś każda ze zmiennych ξ_n ma rozkład taki, że $P(\xi_n = 1) = \frac{2}{3} = 1 - P(\xi_n = 2)$. Obliczyć warunkową wartość oczekiwaną $E(N|\eta = 3)$.

14.48. Załóżmy, że zmienne losowe $\xi_1, \xi_2, \ldots$ są niezależne i mają jednakowy rozkład wykładniczy $E(1/\lambda)$. Zmienne losowe $\bar{\xi}_1, \bar{\xi}_2, \ldots$ określamy wzorem

$$\bar{\xi}_i = \begin{cases} \xi_i - aE\xi_i, & \text{gdy } \xi_i > aE\xi_i, \\ 0, & \text{gdy } \xi_i \le aE\xi_i. \end{cases}$$

Zmienna losowa N ma rozkład $Po\,(\lambda)$ i jest niezależna od $\xi_1, \xi_2, \ldots$ Niech $S = \sum_{i=1}^{N} \xi_i$ oraz $\bar{S} = \sum_{i=1}^{N} \bar{\xi}_i$. Dobrać liczbę a tak, by $D^2\bar{S} = 0.36D^2S$.

14.49. Niech $N, \xi_1, \xi_2, \ldots, \xi_n, \ldots$ będą niezależnymi zmiennymi losowymi, przy czym zmienna losowa N ma rozkład $Ge\,(q)$, gdzie $q \in (0,1)$ jest ustaloną liczbą, a zmienne losowe $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ mają ten sam rozkład $E\,(\lambda)$. Niech

$$S_N = \begin{cases} \xi_1 + \cdots + \xi_N, & \text{gdy } N > 0, \\ 0, & \text{gdy } N = 0. \end{cases}$$

Wyznaczyć prawdopodobieństwo $P(S_N \leq x)$ dla $x > 0$.

14.50. Niech $\xi_1, \xi_2, \ldots, \xi_n, \ldots,\ I_1, I_2, \ldots, I_n, \ldots$ oraz N będą niezależnymi zmiennymi losowymi. Zmienne $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ mają rozkład $E\,(1/\mu)$. Zmienne losowe $I_1, I_2, \ldots, I_n, \ldots$ mają rozkład $D\,(p)$, gdzie $p \in (0,1)$ jest ustaloną liczbą. Zmienna N ma rozkład $NB\,(r,q)$, gdzie $r > 0$ i $q \in (0,1)$ są ustalone. Niech

$$T_n = \begin{cases} \sum_{i=1}^{N} \xi_i, & \text{gdy } N > 0, \\ 0, & \text{gdy } N = 0, \end{cases} \qquad S_n = \begin{cases} \sum_{i=1}^{N} I_i\xi_i, & \text{gdy } N > 0, \\ 0, & \text{gdy } N = 0, \end{cases}$$

Wyznaczyć kowariancję $Cov(T_N, S_N)$.

14.51. Załóżmy, że dla danej wartości $\Theta = \theta$ zmienne losowe $\xi_1, \ldots, \xi_n, \ldots$ są warunkowo niezależne i mają rozkład $D\,(p)$. Zmienna losowa Θ ma rozkład $U\,(0,1)$. Niech $N = \min\{n : \xi_n = 1\}$. Obliczyć $P(N = n+1|N > n)$ dla $n = 0, 1, 2, \ldots$

14.52. Załóżmy, że $\xi_1, \ldots, \xi_n, \ldots$ są niezależnymi zmiennymi losowymi o rozkładzie $U\,(0,1)$, zaś N jest zmienną o rozkładzie $Po\,(\lambda)$, niezależną od $\xi_1, \ldots, \xi_n, \ldots$ Niech

$$M = \begin{cases} \max\{\xi_1, \ldots, \xi_n\}, & \text{gdy } N > 0, \\ 0, & \text{gdy } N = 0. \end{cases}$$

Obliczyć EM.

14.53. Niech $\xi_1, \xi_2, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie $U\,(1,2)$. Niech N będzie zmienną losową o rozkładzie $NB\,(3,p)$ niezależną od zmiennych losowych $\xi_1, \xi_2, \ldots$ Niech

$$M_N = \begin{cases} \max\{\xi_1, \ldots, \xi_N\}, & \text{gdy } N > 0, \\ 0, & \text{gdy } N = 0. \end{cases}$$

Obliczyć EM_N.

14.54. Niech $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie $E\,(1/a)$. Niech N i M będą niezależnymi zmiennymi losowymi o rozkładach Poissona i niezależnymi od zmiennych losowych $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, przy czym $EN = \lambda$ i $EM = \mu$. Niech

$$\eta_n = \begin{cases} \max\{\xi_1, \ldots, \xi_n\}, & \text{gdy } n > 0, \\ 0, & \text{gdy } n = 0. \end{cases}$$

Obliczyć $P(\eta_{M+N} > \eta_M)$.

14.55. Niech $\xi_0, \xi_1, \ldots, \xi_n, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie $E\,(1)$. Niech N będzie zmienną losową o rozkładzie $Po\,(\lambda)$, niezależną od zmiennych $\xi_1, \ldots, \xi_n, \ldots$. Niech $M_N = \min\{\xi_0, \xi_1, \ldots, \xi_N\}$. Wyznaczyć $Cov(M_N, N)$.

14.56. Niech $\xi_1, \ldots, \xi_n, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie $E\,(1)$. Niech N będzie zmienną losową o rozkładzie $NB\left(2, e^{-1}\right)$ niezależną od zmiennych $\xi_1, \ldots, \xi_n, \ldots$ Niech

$$M_N = \begin{cases} \min\{\xi_1, \ldots, \xi_N\}, & \text{gdy } N > 0, \\ 0, & \text{gdy } N = 0. \end{cases}$$

Wyznaczyć EM_N.

14.57. Niech $\xi_1, \ldots, \xi_n, \ldots$ będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $E\,(\alpha)$. Niech N będzie zmienną losową niezależną od $\xi_1, \ldots, \xi_n, \ldots$ o rozkładzie $Po\,(\lambda)$. Niech

$$\eta = \begin{cases} \min\{\xi_1, \ldots, \xi_n\}, & \text{gdy } N > 0, \\ 0, & \text{gdy } N = 0. \end{cases}$$

Obliczyć $E(N|\eta = y)$ przy założeniu, że $y > 0$.

15. Wielowymiarowy rozkład normalny

Definicja 15.1 Wektor losowy $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_p]^T$ ma p-wymiarowy rozkład normalny $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, jeżeli jego funkcja gęstości prawdopodobieństwa wyraża się wzorem

$$f_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

dla $\boldsymbol{x} \in \mathbb{R}^p$.

Wektor $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_p]^T$ jest wektorem wartości oczekiwanych wektora losowego $\boldsymbol{\xi}$, tzn. $E\xi_i = \mu_i$ dla $i = 1, \ldots, p$.

Macierz $\boldsymbol{\Sigma} = [\sigma_{ij}]_{i,j=1,\ldots,p}$ jest macierzą kowariancji wektora losowego $\boldsymbol{\xi}$, tzn. $D^2\xi_i = \sigma_{ii}$ oraz $Cov(\xi_i, \xi_j) = \sigma_{ij}$. Zwyczajowe oznaczenie $D^2\xi_i = \sigma_i^2$

Twierdzenie 15.1. *Jeżeli $\boldsymbol{A}$ jest $r \times p$ macierzą, to $\boldsymbol{A\xi} \sim N_r\left(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T\right)$.*

Wniosek. *Każdy rozkład brzegowy jest normalny.*

Rozkłady warunkowe. Niech

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Rozkład wektora $\boldsymbol{\xi}_1$ pod warunkiem $\boldsymbol{\xi}_2 = \boldsymbol{x}_2$ jest rozkładem normalnym

$$N_m\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T(\boldsymbol{x}_2 - \mu_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T\right)$$

Warunkowa wartość oczekiwana.

$$E(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T(\boldsymbol{\xi}_2 - \mu_2)$$

Twierdzenie 15.3. *$\boldsymbol{\xi}_1$ oraz $\boldsymbol{\xi}_2$ są niezależne wtedy i tylko wtedy, gdy*

$$\boldsymbol{\Sigma}_{12} = \mathbf{0}.$$

W szczególności ξ_i oraz ξ_j są niezależne wtedy i tylko wtedy, gdy $\sigma_{ij} = 0$

Zadania do samodzielnego rozwiązania

15.1. Udowodnić, że jeżeli $\boldsymbol{\xi} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ oraz $\boldsymbol{A}$ jest $r \times p$ macierzą, to $\boldsymbol{A\xi} \sim N_r\left(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^T\right)$.

15.2. Wyznaczyć rozkład warunkowy wektora $\boldsymbol{\xi}_1$ pod warunkiem $\boldsymbol{\xi}_2 = \boldsymbol{x}_2$.

15.3. Niech $\boldsymbol{\xi} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ oraz niech $\boldsymbol{A}$ będzie macierzą symetryczną. Udowodnić, że $E(\boldsymbol{\xi}^T\boldsymbol{A}\boldsymbol{\xi}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}$.

15.4. Niech $\boldsymbol{\xi} \sim N_p\left(\boldsymbol{\mu}, \sigma^2\boldsymbol{I}\right)$ oraz niech $\boldsymbol{A}$ będzie macierzą symetryczną. Udowodnić, że $D^2(\boldsymbol{\xi}^T\boldsymbol{A}\boldsymbol{\xi}) = 2\sigma^4\mathrm{tr}\boldsymbol{A}^2 + 4\sigma^2\boldsymbol{\mu}^T\boldsymbol{A}^2\boldsymbol{\mu}$.

15.5. Niech $\boldsymbol{\xi} \sim N_p\left(\boldsymbol{\mu}, \sigma^2\boldsymbol{I}\right)$ oraz niech $\boldsymbol{A}$ będzie macierzą symetryczną. Udowodnić, że $(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{A}(\boldsymbol{\xi}-\boldsymbol{\mu})$ ma rozkład $\chi^2\left(\mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma})\right)$ stopniami swobody wtedy i tylko wtedy, gdy $\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}$.

15.6. Niech $\boldsymbol{\xi} \sim N_p(\boldsymbol{\mu}, \boldsymbol{I})$ oraz niech $\boldsymbol{A}$ będzie $p \times p$ macierzą rzędu r. Udowodnić, że forma kwadratowa $(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{A}(\boldsymbol{\xi}-\boldsymbol{\mu})$ ma rozkład $\chi^2(r)$ wtedy i tylko wtedy, gdy $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$.

15.7. (*Twierdzenie Cochrana*) Niech $\boldsymbol{\xi} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$ i niech $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_k$ będą takimi macierzami symetrycznymi rzędów $n_1, \ldots, n_k$, że $\boldsymbol{\xi}^T\boldsymbol{\xi} = \sum_{i=1}^k \boldsymbol{\xi}^T\boldsymbol{A}_i\boldsymbol{\xi}$. Formy kwadratowe $\boldsymbol{\xi}^T\boldsymbol{A}_i\boldsymbol{\xi}$ są niezależnymi zmiennymi losowymi o rozkładach $\chi^2(n_i)$ stopniami swobody wtedy i tylko wtedy, gdy $\sum_{i=1}^k n_i = n$.

15.8. Rozpatrzmy zmienne losowe ξ i η o łącznym rozkładzie normalnym. Wiadomo, że $D^2\eta = 9$, $E(\eta|\xi) = \frac{1}{2}\xi + 7$, $D^2(\eta|\xi) = 8$. Wyznaczyć $Cov(\eta|\xi)$.

15.9. Załóżmy, że zmienne losowe ξ i η mają łączny rozkład normalny, $E\xi = E\eta = 0$, $D^2\xi = D^2\eta = 1$ i $Cov(\xi, \eta) = \varrho$. Obliczyć $Cov(\xi^2, \eta^2)$.

15.10. Załóżmy, że ξ, η są zmiennymi losowymi o łącznym rozkładzie normalnym, $E\xi = E\eta = 0$, $D^2\xi = D^2\eta = 1$ i $Cov(\xi, \eta) = \varrho$. Obliczyć $D^2(\xi\eta)$.

15.11. Niech ξ i η będą zmiennymi losowymi o łącznym rozkładzie normalnym takim, że $E\xi = E\eta = 0$, $D^2\xi = 1$, $D^2\eta = 5$, $Cov(\xi, \eta) = -2$. Obliczyć $E(\eta^2|\xi = x)$.

15.12. Załóżmy, że ξ, η i ζ są niezależnymi zmiennymi losowymi rozkładzie $N(0, 1)$. Znaleźć liczbę a taką, że

$$P\left(\frac{|\xi|}{\sqrt{\xi^2 + \eta^2 + \zeta^2}} \le a\right) = 0.6.$$

15.13. Wektor losowy $[\xi_1, \xi_2, \xi_3]^T$ ma trójwymiarowy rozkład normalny z wartością oczekiwaną $[0, 0, 0]^T$ i macierzą kowariancji $\begin{bmatrix} 4 & 1.5 & 1 \\ 1.5 & 1 & 0.5 \\ 1 & 0.5 & 1 \end{bmatrix}$. Występująca w równaniu $\xi_1 = a\xi_2 + b\xi_3 + \varepsilon$ zmienna losowa ε jest nieskorelowana ze zmiennymi losowymi (ξ_2, ξ_3). Wyznaczyć stałą a.

15.14. Wektor losowy $[\xi,\eta,\zeta]^T$ ma rozkład normalny z wartością oczekiwaną $E\xi = E\eta = 1$, $E\zeta = 0$ i macierzą kowariancji

$$\begin{bmatrix} 4 & 1 & 2 \\ 1 & 1 & 1 \\ 2 & 1 & 4 \end{bmatrix}.$$

Obliczyć $D^2((\xi-\eta)\zeta)$.

15.15. Wiemy, że $\eta = 2\xi + \zeta$, gdzie ξ i ζ są niezależnymi zmiennymi losowymi, ξ ma rozkład $N(0,9)$, a ζ ma rozkład $N(0,4)$. Dla jakiego a zachodzi związek $\xi = a\eta + \gamma$ i zmienne η i γ są niezależne.

15.16. Załóżmy, że zmienne losowe $\xi_1,\ldots,\xi_5,\xi_6,\ldots,\xi_{20}$ są niezależne, o jednakowym rozkładzie $N(\mu,\sigma^2)$. Niech $\eta_5 = \xi_1 + \cdots + \xi_5$, $\eta_{20} = \xi_1 + \cdots + \xi_{20}$. Wyznaczyć warunkową wartość oczekiwaną $E(\eta_5|\eta_{20})$.

15.17. Załóżmy, że ξ_0 oraz $\eta_1,\ldots,\eta_{10}$ są niezależnymi zmiennymi losowymi, przy tym każda ze zmiennych $\eta_1,\ldots,\eta_{10}$ ma jednakowy rozkład $N(5,1)$. Niech

$$\xi_{n+1} = \frac{1}{2}\xi_n + \eta_{n+1}, \quad \text{dla} \quad n = 0,1,\ldots,9.$$

Wiadomo, że zmienne losowe ξ_0 i ξ_{10} mają rozkład normalny o jednakowych parametrach. Wyznaczyć parametry tego rozkładu.

15.18. Załóżmy, że $\xi_1,\ldots,\xi_m,\xi_{m+1},\ldots,\xi_n$ są niezależnymi zmiennymi losowymi o rozkładzie $N(\mu,\sigma^2)$. Niech $\bar{\xi}_m = \frac{1}{m}\sum_{i=1}^m \xi_i$ oraz $\bar{\xi}_n = \frac{1}{n}\sum_{i=1}^n \xi_i$. Obliczyć

$$E\left[\frac{\sum_{i=1}^m (\xi_i - \bar{\xi}_m)^2}{\sum_{i=1}^n (\xi_i - \bar{\xi}_n)^2}\right].$$

15.19. Niech $[\xi_1,\eta_1]^T,\ldots,[\xi_n,\eta_n]^T$ będą niezależnymi wektorami losowymi o dwuwymiarowym rozkładzie normalnym o nieznanych parametrach:

$$E\xi_i = E\eta_i = \mu, \quad D^2\xi_i = D^2\eta_i = \sigma^2, \quad Cov(\xi_i,\eta_i) = \varrho\sigma^2.$$

Niestety, obserwacje zmiennych losowych ξ i η zostały oddzielone, zmienne η pomieszane, po czym zagubiono informacje o przynależności do par. Możemy to sformalizować przyjmując, iż mamy nadal niezmieniony ciąg zmiennych ξ oraz ciąg $\zeta_1,\ldots,\zeta_n$ stanowiący losową permutację ciągu $\eta_1,\ldots,\eta_n$. Obliczyć $Cov(\xi_i,\zeta_i)$.

15.20. Załóżmy, że $\xi_1,\ldots,\xi_n$ i $\eta_1,\ldots,\eta_m$ są niezależnymi niezależnymi zmiennymi losowymi o rozkładzie $N(\mu,\sigma^2)$. Niech $\bar{\xi} = \frac{1}{n}\sum_{i=1}^n \xi_i$ oraz $\bar{\eta} = \frac{1}{m}\sum_{i=1}^m \eta_i$. Obliczyć $P(|\bar{\xi}-\mu| > |\bar{\eta}-\mu|)$ dla $n = 100$ i $m = 385$.

16. Twierdzenia graniczne

Rodzaje zbieżności.

Definicja 16.1 Ciąg zmiennych losowych $(\xi_n)_{n=1}^{\infty}$ jest zbieżny do zmiennej losowej ξ **prawie na pewno**, jeżeli

$$P\left(\left\{\omega: \lim_n \xi_n(\omega) = \xi(\omega)\right\}\right) = 1.$$

Oznaczenie: $\xi_n \xrightarrow{\text{p.n.}} \xi$

Definicja 16.2 Ciąg zmiennych losowych $(\xi_n)_{n=1}^{\infty}$ jest zbieżny do zmiennej losowej ξ **według prawdopodobieństwa**, jeśli

$$(\forall \varepsilon > 0) \ \lim_n P(|\xi_n - \xi| > \varepsilon) = 0.$$

Oznaczenie: $\xi_n \xrightarrow{\text{P}} \xi$

Definicja 16.3 Ciąg zmiennych losowych $(\xi_n)_{n=1}^{\infty}$ jest zbieżny do zmiennej losowej ξ **według rozkładu**, jeśli ciąg dystrybuant $(F_{\xi_n})_{n=1}^{\infty}$ jest zbieżny do dystrybuanty F_ξ w każdym punkcie jej ciągłości.

Oznaczenie: $\xi_n \xrightarrow{\text{D}} \xi$

Zachodzi:

$$\xi_n \xrightarrow{\text{p.n.}} \xi \Rightarrow \xi_n \xrightarrow{\text{P}} \xi \Rightarrow \xi_n \xrightarrow{\text{D}} \xi$$

Prawa Wielkich Liczb.

Niech $\xi_1, \xi_2, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie takim, że $E\xi_1 = \mu$, $D^2\xi_1 = \sigma^2 \in (0, \infty)$ oraz niech $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$ i $\bar{\xi}_n = \frac{\xi_1 + \xi_2 + \cdots + \xi_n}{n} = \frac{S_n}{n}$.

Dla ciągu $\xi_1, \xi_2, \ldots$ zachodzi **mocne prawo wielkich liczb**, jeżeli $\bar{\xi}_n \xrightarrow{\text{p.n.}} \mu$, tzn.

$$P\left(\lim_{n\to\infty} \bar{\xi}_n = \mu\right) = 1$$

Dla ciągu $\xi_1, \xi_2, \ldots$ zachodzi **słabe prawo wielkich liczb**, jeżeli $\bar{\xi}_n \xrightarrow{\text{P}} \mu$, tzn.

$$(\forall \varepsilon > 0) \ \lim_{n\to\infty} P\left(\left|\bar{\xi}_n - \mu\right| < \varepsilon\right) = 1$$

Nierówność Czebyszewa.

$$(\forall \varepsilon > 0) \quad P\left(|\bar{\xi}_n - \mu| \geq \varepsilon\right) \leq \frac{E|\xi_1 - \mu|}{n\varepsilon}$$

Nierówność Czebyszewa-Bienayme.

$$(\forall \varepsilon > 0) \quad P\left(|\bar{\xi}_n - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Przykład. *Zastosujemy mocne prawo wielkich liczb do oszacowania liczby π za pomocą komputerowego generatora liczb losowych*

Rozwiązanie.

Niech $\eta \sim U(0,1)$. Niech $\xi = 4\sqrt{1-\eta^2}$. Jak łatwo sprawdzić

$$E\xi = \int_{-\infty}^{\infty} 4\sqrt{1-x^2} f_\eta(x)dx = 4\int_0^1 \sqrt{1-x^2}dx = \pi$$

Z komputerowego generatora liczb losowych generujemy n realizacji $\eta_1, \ldots, \eta_n$ zmiennej losowej η. Następnie obliczamy $\xi_1, \ldots, \xi_n$. Z Prawa Wielkich Liczb: $\bar{\xi}_n \approx \pi$

Jak dużo symulacji, by z dużym prawdopodobieństwem γ błąd był nie większy niż zadane ε?

Z nierówności Czebyszewa-Bienayme: $\frac{\sigma^2}{n\varepsilon^2} \leq 1-\gamma$. Zatem $n \geq \frac{\sigma^2}{\varepsilon^2(1-\gamma)}$

□

Centralne Twierdzenie Graniczne (CTG).

Twierdzenie 16.1. *(CTG) Niech $\xi_1, \xi_2, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie, o wartości oczekiwanej μ i wariancji $0 < \sigma^2 < \infty$. Wtedy*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) - \Phi(x)\right| = 0.$$

Centralne twierdzenie graniczne mówi, że

$$\frac{\bar{\xi}_n - \mu}{\sigma}\sqrt{n} \xrightarrow{\mathrm{D}} N(0,1)$$

W skrócie zapisujemy:

$$(S_n)_{n=1}^{\infty} \sim AN(n\mu, n\sigma^2)$$

Twierdzenie Berry-Esséna'a.

Twierdzenie 16.2. *(Berry-Esséna'a) Jeżeli $\xi_1, \xi_2, \ldots$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie oraz $E|\xi_1|^3 < \infty$, to*

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) - \Phi(x) \right| \le C \frac{E|\xi_1 - \mu|^3}{\sigma^3 \sqrt{n}},$$

gdzie $1/\sqrt{2\pi} \le C < 0.8$.

Rozkład dwumianowy.

Twierdzenie 16.3. *(de Moivre-Laplace'a) Niech $\zeta_n \sim Bin(n, p)$. Wtedy*

$$\lim_{n\to\infty} \sup_{x \in \mathbb{R}} \left| P\left(\frac{\zeta_n - np}{\sqrt{npq}} \le x\right) - \Phi(x) \right| = 0.$$

Twierdzenie Berry-Esséna'a dla zmiennych dwumianowych: niech $\zeta_n \sim Bin(n, p)$. Wtedy

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{\zeta_n - np}{\sqrt{npq}} \le x\right) - \Phi(x) \right| \le C \frac{p^2 + (1-p)^2}{\sqrt{np(1-p)}}$$

Przybliżenie poissonowskie rozkładu dwumianowego

Niech $\zeta_n \sim Bin(n, p)$. Wtedy dla każdego zbioru $M \subseteq \mathbb{N}$ mamy

$$\left| P(\zeta_n \in M) - \sum_{k \in M} \frac{(np)^k}{k!} e^{-np} \right| \le \frac{(np)^2}{n}$$

Dystrybuanta empiryczna. Niech $\xi_1, \xi_2, \ldots$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie o dystrybuancie F. **Dystrybuantą empiryczną** nazywamy

$$F_n : \mathbb{R} \ni x \to \frac{\#\{1 \le j \le n : \xi_j \le x\}}{n} \in [0, 1].$$

Z Mocnego Prawa Wielkich Liczb mamy:

$$(\forall x \in \mathbb{R}) \ \lim_{n\to\infty} F_n(x) = F(x)$$

Z Centralnego Twierdzenie Granicznego wynika:

$$(\forall x \in \mathbb{R}) \ F_n(x) \sim AN\left(F(x), \frac{F(x)(1 - F(x))}{n}\right)$$

Zadania do samodzielnego rozwiązania

16.1. Obliczyć w przybliżeniu prawdopodobieństwo, że partia pięciuset elementów, z których każdy ma czas pracy T_i $(i = 1, 2, \ldots, 500)$ wystarczy na zapewnienie pracy urządzenia przez łącznie tysiąc godzin, gdy wiadomo, że $ET_i = 2$ oraz $D^2 T_i = 1$.

16.2. Tygodniowe wypłaty z pewnego funduszu są niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z tą samą wartością oczekiwaną 2000 zł. Obliczyć prawdopodobieństwo, że łączna wypłata z tego funduszu w okresie roku, tzn. 52 tygodni, przekroczy 100000 zł.

16.3. W grupie studenckiej przeprowadzono test z analizy, w którym można uzyskać od 0 do 100 punktów. Przeciętna liczba punktów uzyskiwanych przez studenta wynosi 40, a odchylenie standardowe 20. Zakładając, że wyniki studentów są niezależnymi zmiennymi losowymi o tym samym rozkładzie prawdopodobieństwa, obliczyć prawdopodobieństwo tego, że średnia liczba punktów w 150 osobowej grupie studenckiej będzie większa od 45.

16.4. Komputer dodaje 15000 liczb rzeczywistych, z których każdą zaokrągla do najbliższej całkowitej, a liczbę $n + 0.5$ do najbliższej parzystej (n jest liczbą całkowitą). Zakładając, że błędy zaokrągleń są niezależnymi zmiennymi losowymi o rozkładach $U(-0.5, 0.5)$, obliczyć prawdopodobieństwo tego, że błąd w obliczeniu sumy przekroczy 10.

16.5. Licznik Geigera–Müllera i źródło promieniowania umieszczono względem siebie tak, że prawdopodobieństwo zarejestrowania przez licznik wypromieniowanej cząsteczki wynosi $\frac{1}{1000}$. Załóżmy, że w czasie obserwacji preparat radioaktywny wypromieniował 30000 cząstek. Jakie jest prawdopodobieństwo, że licznik zarejestruje co najwyżej trzy cząstki?

16.6. Załóżmy, że przy składaniu książki w drukarni każda litera ma prawdopodobieństwo 0.0001, że zostanie złożona błędnie. Po złożeniu książki szpalty czyta korektor, który znajduje każdy błąd z prawdopodobieństwem 0.5. Znaleźć prawdopodobieństwo, że w książce o stu tysiącach znaków drukarskich pozostaną po korekcie nie więcej niż dwa błędy.

16.7. Wśród ziaren pszenicy znajduje się 0.2% ziaren chwastów. Jakie jest prawdopodobieństwo, że wśród wybranych losowo tysiąca ziaren znajdują się co najmniej trzy ziarna chwastów?

16.8. Wielu botaników dokonywało doświadczeń nad krzyżowaniem żółtego groszku (hybryda). Według znanej hipotezy Mendla, prawdopodobieństwo pojawienia się zielonego groszku przy takich krzyżówkach równa się 1/4. Zakładając słuszność hipotezy Mendla obliczyć prawdopodobieństwo, że dla 3400 doświadczeń nad krzyżówką, w co najmniej dziewięciuset przypadkach otrzymano zielony groszek.

16.9. Pewna partia polityczna ma dwudziestoprocentowe poparcie w społeczeństwie. Jakie jest prawdopodobieństwo, że w okręgu wyborczym składającym się z 10000 wyborców, partia uzyska co najmniej 1000 głosów przy siedemdziesięcioprocentowej frekwencji.

16.10. Z dotychczasowych obserwacji wynika, że zainteresowanie mężczyzn pewnym kierunkiem studiów jest niewielkie i można je szacować na poziomie 1%. Jakie jest prawdopodobieństwo, że wśród 200 osób przyjmowanych w kolejnej rekrutacji będzie co najmniej trzech mężczyzn?

16.11. Niech $\xi_{(1)}, \ldots, \xi_{(400)}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie ciągłym o wariancji σ^2, ustawionymi w porządku niemalejącym, tzn. $\xi_{(1)} \leq \cdots \leq \xi_{(400)}$. Niech m będzie medianą rozważanego rozkładu. Na podstawie Centralnego Twierdzenie Granicznego wyznaczyć przybliżoną wartość prawdopodobieństwa $P(\xi_{(220)} \leq m)$.

16.12. Załóżmy, że $\xi_1, \ldots, \xi_{735}$ oraz $\eta_1, \ldots, \eta_{880}$ są niezależnymi zmiennymi losowymi o rozkładach $D\,(3/7)$ oraz $D\,(0.5)$, odpowiednio. Korzystając z Centralnego Twierdzenia Granicznego obliczyć $P\left(\sum_{i=1}^{735} \xi_i < \sum_{i=1}^{880} \eta_i\right)$.

16.13. Zmienne losowe $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ są niezależne o jednakowym rozkładzie

$$P(\xi_n = 0) = P(\xi_n = 1) = P(\xi_n = 2) = P(\xi_n = 3) = \frac{1}{4}.$$

Niech $\eta_0 = 3$ oraz niech dla $n = 1, 2, \ldots$ zachodzi

$$\eta_n = \begin{cases} 3, & \text{gdy } \xi_n = 3, \\ \min\{\eta_{n-1}, \xi_n\}, & \text{gdy } \xi_n < 3. \end{cases}$$

Obliczyć $\lim\limits_{n\to\infty} P(\eta_n \leq 1)$.

16.14. Załóżmy, że $\eta_1, \ldots, \eta_n, \ldots$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie:

$$P(\eta_n = 0) = P(\eta_n = 1) = \cdots = P(\eta_n = 9) = \frac{1}{10}.$$

Niech $\xi_0 = 0$ oraz niech dla $n = 1, 2, \ldots$

$$\xi_n = \begin{cases} \max\{\xi_{n-1}, \eta_n\}, & \text{gdy } \eta_n > 0, \\ 0, & \text{gdy } \eta_n = 0. \end{cases}$$

Obliczyć granicę $\lim\limits_{n\to\infty} P(\xi_n \geq 3)$.

16.15. Załóżmy, że $\xi_1, \ldots, \xi_n, \ldots$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie $U\,(0, 1)$. Rozważmy ciąg średnich geometrycznych $\sqrt[n]{\xi_1 \cdots \xi_n}$. Udowodnić, że $\lim\limits_{n\to\infty} P\left(\sqrt[n]{\xi_1 \cdots \xi_n} \leq \frac{1}{3}\right) = 0$.

16.16. Zmienne losowe $\xi_1, \ldots, \xi_n, \ldots$ są niezależne i mają identyczny rozkład $U\,(0, 2)$. Niech $\eta_n = \xi_1 \cdots \xi_n$. Obliczyć $\lim\limits_{n\to\infty} P(\eta_n \leq 0.5)$.

16.17. Zmienne losowe $\xi_1, \ldots, \xi_n, \ldots$ są niezależne i mają identyczny rozkład $U\,(0, 2)$. Niech $\eta_n = \xi_1 \cdots \xi_n$. Udowodnić, że $\lim\limits_{n\to\infty} P(\eta_n \leq (2/e)^n) = 0.5$.

16.18. Niech $\xi_1, \ldots, \xi_n$ będą niezależnymi zmiennymi losowymi z rozkładu o gęstości

$$f(x) = \begin{cases} 2x, & \text{dla } x \in (0, 1), \\ 0, & \text{dla } x \notin (0, 1). \end{cases}$$

Niech $\eta_n = \prod_{i=1}^{n} \xi_i^{\frac{1}{n}}$. Udowodnić, że $\lim\limits_{n\to\infty} P\{|\eta_n - e^{-0.5}|\sqrt{n} > e^{-0.5}\} = 0.046$.

16.19. W urnie I znajdują się dwie kule i w urnie II znajdują się dwie kule. Na te cztery kule w sumie składają się dwie kule białe i dwie czarne. Przeprowadzamy następujące doświadczenie losowe.

a. najpierw losujemy jedną kulę z urny I i przekładamy ją do urny II,

b. następnie losujemy jedną kulę z urny II i przekładamy ją do urny I.

Sekwencję losowań a) i b) powtarzamy wielokrotnie. Przed każdym losowaniem dokładnie mieszamy kule w urnie. Niech $p_n(1)$ oznacza prawdopodobieństwo tego, że po n powtórzeniach (czyli po $2n$ losowaniach) w urnie I znajduje się jedna biała i jedna czarna kula. Obliczyć $\lim_{n\to\infty} p_n(1)$.

16.20. Na początku doświadczenia w urnie I znajdują się trzy kule białe, zaś w urnie II trzy kule czarne. Losujemy po jednej kuli z każdej urny, po czym wylosowaną kule z urny I wrzucamy do urny II, a tę wylosowaną z urny II wrzucamy do urny I. Czynność te powtarzamy wielokrotnie. Obliczyć granicę (przy $n \to \infty$) prawdopodobieństwa, iż obie kule wylosowane w n-tym kroku są jednakowego koloru.

16.21. Urządzenie zawiera dwa podzespoły A i B. Obserwujemy działanie urządzenia w chwilach $t = 0, 1, 2, \ldots$ Każdy z podzespołów w ciągu jednostki czasu może ulec awarii z prawdopodobieństwem $1 - p$, niezależnie od drugiego. Jeśli w chwili t oba podzespoły są niesprawne, następuje naprawa i w chwili $t + 1$ oba są już sprawne. Jeśli w chwili t tylko jeden podzespół jest niesprawny, to nie jest naprawiany. W chwili 0 oba zespoły są sprawne. Obliczyć granicę prawdopodobieństwa, iż podzespół A jest sprawny w chwili t, przy $t \to \infty$.

16.22. Rozważmy ciąg $\xi_1, \ldots, \xi_n, \ldots$ niezależnych zmiennych losowych o jednakowym rozkładzie $N(0, 1)$. Niech

$$S_n = \xi_1\xi_2 + \xi_2\xi_3 + \cdots + \xi_{n-1}\xi_n + \xi_n\xi_{n+1}.$$

Udowodnić, że $\lim_{n\to\infty} P\left(\frac{S_n}{\sqrt{n}} \le a\right) = \Phi(a)$ dla każdego a

16.23. Zmienne losowe $I_1, \ldots, I_n, \ldots$ i $\xi_1, \ldots, \xi_n, \ldots$ są niezależne. Każda ze zmiennych I_i ma jednakowy rozkład $D(p)$. Każda ze zmiennych ξ_i ma jednakowy rozkład prawdopodobieństwa taki, że $E\xi_i = \mu$ i $D^2\xi_i = \sigma^2$. Niech $S_n = \sum_{i=1}^n I_i\xi_i$ oraz $K_n = \sum_{i=1}^n I_i$. Zbadać zbieżność rozkładów prawdopodobieństwa zmiennych losowych $\frac{S_n - K_n\mu}{\sqrt{n}}$ przy $n \to \infty$.

16.24. Zmienne losowe $\zeta_1, \ldots, \zeta_n$ i $(\xi_1, \eta_1), \ldots, (\xi_n, \eta_n)$ są niezależne. Każda ze zmiennych losowych ζ_i ma jednakowy rozkład $D(p)$. Każda ze zmiennych losowych (ξ_i, η_i) ma jednakowy rozkład prawdopodobieństwa taki, że $E\xi_i = E\eta_i = m$, $D^2\xi_i = D^2\eta_i = \sigma^2$ i współczynnik korelacji $Corr(\xi_i, \eta_i) = \varrho$. Niech $S_n = \sum_{i=1}^n \zeta_i\xi_i$ i $T_n = \sum_{i=1}^n \zeta_i\eta_i$. Zbadać zbieżność rozkładów prawdopodobieństwa zmiennych

$$\frac{S_n - T_n}{\sqrt{n}} \qquad \text{przy} \qquad n \to +\infty.$$

16.25. Załóżmy, że $\xi_1, \ldots, \xi_n, \ldots$ jest ciągiem zmiennych losowych takim, że zmienna ξ_1 ma rozkład $E(1/\lambda)$; warunkowo, dla danych $\xi_1, \ldots, \xi_n$ zmienna ξ_{n+1} ma gęstość wykładniczą (dla $w_{n+1} > 0$)

$$f(w_{n+1}|w_1, \ldots, w_n) = \begin{cases} \lambda \exp(-\lambda w_{n+1}), & \text{gdy } w_n \le c, \\ \mu \exp(-\mu w_{n+1}), & \text{gdy } w_n > c. \end{cases}$$

Niech $c = \ln 2$, $\lambda = 1$, $\mu = 2$. Obliczyć $\lim\limits_{n\to\infty} E\xi_n$.

16.26. Załóżmy, że $W_1, \ldots, W_n, \ldots$ jest ciągiem zmiennych takim, że zmienna losowa W_1 ma rozkład o gęstości $f(w_1) = \frac{4}{(1+w_1)^5}$ dla $w_1 > 0$, natomiast W_{n+1} ma rozkład, przy danych $w_1, \ldots, w_n$ o funkcji gęstości

$$f(w_{n+1}|w_1, \ldots, w_n) = \begin{cases} \frac{4}{(1+w_{n+1})^5}, & \text{gdy } w_n \le 1, \\ \frac{3}{(1+w_{n+1})^4}, & \text{gdy } w_n > 1. \end{cases}$$

Wyznaczyć $\lim\limits_{n\to\infty} E(W_n)$.

16.27. Niech $\chi^2_{0.1}(n)$ oznacza kwantyl rzędu 0.1 rozkładu $\chi^2(n)$. Obliczyć (z dokładnością do 0.01)

$$g = \lim_{n\to\infty} \frac{\chi^2_{0.1}(n) - n}{\sqrt{n}}.$$

16.28. Obserwujemy działanie urządzenia w kolejnych chwilach $t = 0, 1, 2, \ldots$ Działanie tego urządzenia zależy od pracy dwóch podzespołów A i B. Każdy z nich może ulec awarii w jednostce czasu z prawdopodobieństwem 0.1 niezależnie od drugiego. Jeżeli jeden z podzespołów ulega awarii, to urządzenie nie jest naprawiane i działa dalej wykorzystując drugi podzespół. Jeżeli oba podzespoły są niesprawne w chwili t, to następuje ich naprawa i w chwili $t + 1$ oba są sprawne. Wyznaczyć prawdopodobieństwo, że podzespół B jest sprawny w chwili t, przy t dążącym do nieskończoności (z dokładnością do 0.001).

16.29. Niech ξ będzie dowolną zmienną losową, przy czym $E\xi = 0$, $D^2\xi = \sigma^2$; niech $F(x)$ będzie dystrybuantą zmiennej losowej ξ. Udowodnić, że dla $x < 0$ mamy $F(x) \ge x^2/(\sigma^2 + x^2)$. Pokazać na przykładzie, że w pewnych przypadkach nierówności te mogą być równościami.

16.30. Jeżeli ograniczyć się do pewnych klas rozkładów, to udaje się niekiedy wzmocnić nierówność Czebyszewa. I tak w 1821 roku Gauss udowodnił, że dla jednomodalnych rozkładów typu ciągłego, tzn. takich rozkładów, których gęstość ma dokładnie jedno maksimum, dla dowolnego $\varepsilon > 0$ mamy

$$P\{|\xi - x_0| \ge \varepsilon\tau\} \le \frac{4}{9\varepsilon^2},$$

gdzie x_0 jest wartością modalną, a $\tau^2 = D^2\xi + (x_0 - E\xi)^2$ jest drugim momentem względem wartości modalnej. Jeżeli posłużyć się miarą asymetrii

$$S = \frac{E\xi - x_0}{\sqrt{D^2\xi}},$$

wprowadzoną przez K. Pearsona, to z poprzedniej nierówności wynika, że dla $|\varepsilon| > s$

$$P\{|\xi - E\xi| \geq \varepsilon\sqrt{D^2\xi}\} \leq \frac{4(1+s^2}{9(\varepsilon - |s|)^2}.$$

Udowodnić obie te nierówności.

Wskazówka. Pokazać najpierw, że jeżeli $g(x)$ jest funkcją nierosnącą dla $x > 0$, to dla dowolnego $\varepsilon > 0$

$$\varepsilon^2 \int_\varepsilon^\infty g(x)dx \leq \frac{4}{9} \int_0^\infty x^2 g(x)dx.$$

16.31. Uogólnienie nierówności Czebyszewa.

a. Udowodnić, że jeżeli zmienna losowa ξ ma tę własność, że istnieje $Ee^{a\xi}$ ($a > 0$ jest stałą), to

$$P\{\xi \geq \varepsilon\} \leq \frac{Ee^{a\xi}}{e^{a\varepsilon}}.$$

b. Niech $f(x) > 0$ będzie funkcją niemalejącą. Udowodnić, że jeżeli istnieje $Ef(|\xi - E\xi|)$, to

$$P\{|\xi - E\xi| \geq \varepsilon\} \leq \frac{Ef(|\xi - E\xi|)}{f(\varepsilon)}.$$

17. Statystyka matematyczna - preliminaria

Model statystyczny.

Modelem statystycznym nazywamy trójkę

$$(\mathcal{X}, \mathcal{B}, \{P_\theta,\ \theta \in \Theta\}),$$

gdzie $\mathcal{X}$ oznacza przestrzeń obserwacji, tzn. zbiór wartości obserwowanej zmiennej losowej lub wektora losowego. $\mathcal{B}$ oznacza σ-ciało podzbiorów przestrzeni $\mathcal{X}$ (zdarzeń losowych), natomiast $\{P_\theta,\ \theta \in \Theta\}$ oznacza rodzinę miar probabilistycznych. Zbiór Θ nazywamy **przestrzenią parametrów**.

Niech F_θ oraz p_θ oznaczają odpowiednio dystrybuantę oraz gęstość rozkładu P_θ.

Definicja 17.1 **Próbą** nazywamy ciąg niezależnych zmiennych losowych $X_1, \ldots, X_n$ o jednakowym rozkładzie P_θ.

Niech $F_n(\cdot)$ będzie dystrybuantą empiryczną.

Twierdzenie 17.1. *(Podstawowe twierdzenie statystyki matematycznej; twierdzenie Gliwienki-Cantelliego) Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu o dystrybuancie F_θ, to zmienna losowa*

$$D_n = \sup_{-\infty < t < \infty} |F_n(t) - F_\theta(t)|$$

dąży do zera z prawdopodobieństwem 1.

Estymator.

Definicja 17.2 **Statystyką** nazywamy odwzorowanie $T : \mathcal{X} \to \mathbb{R}^k$.

Definicja 17.3 **Estymatorem** parametru θ nazywamy odwzorowanie $\hat{\theta} : \mathcal{X} \to \Theta$.

Kryteria estymacji.

Definicja 17.4 **Funkcją straty** estymatora $\hat{\theta}$ nazywamy odwzorowanie

$$L : \hat{\theta}(\mathcal{X}) \times \Theta \to \mathbb{R}_+.$$

Definicja 17.5 **Ryzykiem** estymatora $\hat{\theta}$ nazywamy funkcję

$$\theta \to R_\theta(\hat{\theta}) = E_\theta L(\hat{\theta}(X), \theta)$$

.

Typową funkcją straty jest **kwadratowa** funkcja straty:

$$L(\hat{\theta}(x), \theta) = (\hat{\theta}(x) - \theta)^2.$$

Definicja 17.6 Estymator $\hat{\theta}$ nazywamy **słabo zgodnym**, jeżeli dla każdego $\theta \in \Theta$ zbiega on (wraz ze wzrostem liczności próby) do wartości θ według prawdopodobieństwa P_θ:

$$\hat{\theta} \xrightarrow{\mathrm{P}} \theta, \quad \forall \theta \in \Theta.$$

Definicja 17.7 Estymator $\hat{\theta}$ nazywamy **mocno zgodnym**, jeżeli dla każdego $\theta \in \Theta$ zbiega on (wraz ze wzrostem liczności próby) do wartości θ z prawdopodobieństwem P_θ równym 1:

$$\hat{\theta} \xrightarrow{\mathrm{p.n}} \theta, \quad \forall \theta \in \Theta.$$

Dostateczność.

Definicja 17.8 Statystyka T jest **dostateczna**, jeżeli rozkład warunkowy $P_\theta\{\cdot | T = t\}$ nie jest zależny od θ dla wszystkich $\theta \in \Theta$.

Definicja 17.9 Statystyka T jest **minimalną statystyką dostateczną**, jeżeli dla każdej statystyki dostatecznej U istnieje funkcja h taka, że $T = h(U)$.

Definicja 17.10 Statystyka T jest **zupełna**, jeżeli dla wszystkich $\theta \in \Theta$ zachodzi

$$E_\theta h(T) = 0, \text{ to } h \equiv 0.$$

Twierdzenie 17.2. *(O faktoryzacji) Statystyka T jest dostateczna dla rodziny rozkładów $\{P_\theta : \theta \in \Theta\}$ wtedy i tylko wtedy, gdy gęstość $p_\theta(x)$ może być przedstawiona w postaci*

$$p_\theta(x) = g_\theta\{T(x)\} h(x),$$

gdzie funkcja h nie zależy od θ.

Rodziny wykładnicze.

Definicja 17.11 Rodzina rozkładów prawdopodobieństwa $\{P_\theta : \theta \in \Theta\}$ nazywa się **rodziną wykładniczą**, jeżeli każdy rozkład P_θ ma gęstość p_θ o postaci

$$p_\theta(x) = \exp\left\{\sum_{j=1}^{k} c_j(\theta)T_j(x) - b(\theta)\right\} \cdot h(x).$$

Funkcje $T_1(x), T_2(x), \ldots, T_k(x)$ są liniowo niezależne oraz $\{(c_1(\theta), \ldots, c_k(\theta)) : \theta \in \Theta\}$ jest pewnym k-wymiarowym podzbiorem w $\mathbb{R}^k$.

Twierdzenie 17.3. *Jeżeli $\{P_\theta : \theta \in \Theta\}$ oraz $\Theta \subset \mathbb{R}^k$ jest rodziną wykładniczą rozkładów z gęstościami*

$$p_\theta(x) = \exp\left\{\sum_{j=1}^{k} c_j(\theta)T_j(x) - b(\theta)\right\} \cdot h(x),$$

to $(T_1(x), T_2(x), \ldots, T_k(x))$ jest (k-wymiarową) minimalną statystyką dostateczną zupełną.

Zadania do samodzielnego rozwiązania

17.1. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $D(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0, 1)$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru θ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.2. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Bin(n, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0, 1)$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru θ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.3. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Ge(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0, 1)$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru θ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.4. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $NB(r, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0, 1)$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru θ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.5. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Po(\lambda)$ z nieznanym parametrem $\lambda > 0$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru λ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.6. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(0, \theta)$ z nieznanym parametrem $\theta > 0$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru θ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.7. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(\theta, 0)$ z nieznanym parametrem $\theta < 0$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru θ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.8. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(\theta_1, \theta_2)$ z nieznanymi parametrami $\theta_1 < \theta_2$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru (θ_1, θ_2).

c. Podać model statystyczny dla statystyki dostatecznej.

17.9. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E(\lambda)$ z nieznanym parametrem $\lambda > 0$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru λ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.10. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $G(\alpha, \lambda)$ z nieznanym parametrem $\lambda > 0$ i znanym parametrem α.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru λ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.11. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $G(\alpha, \lambda)$ z nieznanymi parametrami $\alpha > 0$ i $\lambda > 0$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru (α, λ).

c. Podać model statystyczny dla statystyki dostatecznej.

17.12. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$ z nieznaną wartością oczekiwaną μ i znaną wariancją $\sigma^2 > 0$.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru μ.

c. Podać model statystyczny dla statystyki dostatecznej.

17.13. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$ z nieznanymi parametrami.

a. Podać model statystyczny próby.

b. Wyznaczyć statystykę dostateczną dla parametru (μ, σ^2).

c. Podać model statystyczny dla statystyki dostatecznej.

17.14. Wykonujemy n doświadczeń Bernoulliego, 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.

17.15. 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 $E(\alpha)$, a czas życia elementów typu B jest zmienną losową o rozkładzie $E(\beta)$. Obserwuje się czas życia T całego urządzenia. Sformułować model statystyczny obserwacji.

17.16. Wykonujemy ciąg niezależnych doświadczeń Bernoulliego, z których każde kończy się sukcesem z nieznanym prawdopodobieństwem θ. Doświadczenia wykonujemy dopóty, dopóki nie uzyskamy m sukcesów. Sformułować model statystyczny tego eksperymentu.

17.17. Przeprowadza się $n = \sum_{j=1}^{k} n_j$ eksperymentów w taki sposób, że n_j eksperymentów wykonuje się na poziomie x_j, $j = 1, 2, \ldots, k$. Prawdopodobieństwo sukcesu w eksperymencie przeprowadzanym na poziomie x jest równe

$$p(x) = \frac{1}{1 + e^{-(\alpha + \beta x)}}, \qquad \alpha \in \mathbb{R}^1,\ \beta > 0,$$

gdzie (α, β) jest nieznanym parametrem. Sformułować model statystyczny tego eksperymentu.

17.18. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $Po(\lambda)$. Wyznaczyć rozkład warunkowy próby pod warunkiem, że $T = t$, gdzie $T = X_1 + X_2 + \cdots + X_n$. Wykazać, że T jest statystyką dostateczną.

17.19. Dla rodziny rozkładów jednostajnych $\{U\left(\theta - \frac{1}{2}, \theta + \frac{1}{2}\right) : \theta \in \mathbb{R}\}$ wyznaczyć statystykę dostateczną z próby $X_1, X_2, \ldots, X_n$.

17.20. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E(\beta; \theta)$. Wykazać, że minimalną statystyką dostateczną jest $(X_{1:n}, \sum_{i=1}^{n}(X_i - X_{1:n}))$.

17.21. Wektor losowy $\boldsymbol{X}$ ma rozkład znany z dokładnością do dwóch parametrów rzeczywistych θ_1 i θ_2. Niech t_1 będzie statystyką dostateczną dla θ_1, gdy θ_2 jest znane oraz niech t_2 będzie statystyką dostateczną dla θ_2, gdy θ_1 jest znane. Udowodnić, że (t_1, t_2) jest statystyką dostateczną dla (θ_1, θ_2).

18. Estymatory nieobciążone o minimalnej wariancji

Niech $(\mathcal{X}, \mathcal{B}, \{P_\theta,\ \theta \in \Theta\})$ będzie modelem statystycznym oraz niech $g : \Theta \to \mathbb{R}^1$ będzie funkcją rzeczywistą. Chcemy oszacować nieznaną wartość $g(\theta)$.

Definicja 18.1 **Błędem średniokwadratowym** estymatora $\hat{g}$ nazywamy

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

Jak łatwo sprawdzić

$$\begin{aligned} R_\theta(\hat{g}) &= E_\theta(\hat{g}(X) - g(\theta))^2 \\ &= E_\theta(\hat{g}(X) - E_\theta\hat{g}(X))^2 + (E_\theta\hat{g}(X) - g(\theta))^2 = D_\theta^2\hat{g}(X) + (E_\theta\hat{g}(X) - g(\theta))^2 \end{aligned}$$

Definicja 18.2 Estymator $\delta(X_1, X_2, \ldots, X_n)$ wielkości $g(\theta)$ nazywamy **nieobciążonym**, jeżeli $(\forall\theta \in \Theta)$ $E_\theta\delta(X_1, X_2, \ldots, X_n) = g(\theta)$.

Estymator wielkości $g(\theta)$ oznaczamy przez $\widehat{g(\theta)}$ lub, krócej, przez $\hat{g}$.

Jeżeli $E_\theta\hat{g} = g(\theta)$ $(\forall\theta \in \Theta)$, to $R_\theta(\hat{g}) = D_\theta^2\hat{g}$

Definicja 18.3 Estymator $\hat{g}$ nazywamy **estymatorem nieobciążonym o minimalnej wariancji** wielkości $g(\theta)$ $(ENMW[g(\theta)])$, jeżeli $D_\theta^2\hat{g} \le D_\theta^2\tilde{g}$ $(\forall\theta \in \Theta)$ dla wszystkich nieobciążonych estymatorów $\tilde{g}$ wielkości $g(\theta)$.

Twierdzenie 18.1. *(Rao-Blackwella) Jeżeli $\hat{g}$ jest estymatorem nieobciążonym i jeżeli T jest statystyką dostateczną zupełną, to $E_\theta(\hat{g}|T)$ jest również estymatorem nieobciążonym o jednostajnie nie większej wariancji niż wariancja estymatora $\hat{g}$.*

Twierdzenie 18.2. *Jeżeli T jest statystyką dostateczną zupełną i jeżeli dla danej funkcji g istnieje funkcja $\hat{g}$ taka, że*

$$(\forall\theta \in \Theta)\ E_\theta\hat{g}(T) = g(\theta),$$

to $\hat{g}(T)$ jest $ENMW[g(\theta)]$.

Definicja 18.4 Wielkość $I_\theta = E_\theta[\{\partial \log p_\theta(X)/\partial\theta\}^2]$ nazywamy ilością informacji o parametrze θ zawartą w X.

Twierdzenie 18.3. *(Nierówność Cramera-Rao) Niech $\{P_\theta : \theta \in \Theta\}$ będzie rodziną rozkładów, niech θ będzie parametrem liczbowym i niech Θ będzie przedziałem na prostej. Zakładamy, że dla każdego θ rozkład P_θ ma gęstość p_θ. Jeżeli spełnione są pewne warunki regularności, to nierówność*

$$D^2_\theta \theta^* \geq I_\theta^{-1}$$

spełniona jest dla każdego estymatora nieobciążonego θ^* parametru θ.

Definicja 18.5 Liczbę $\mathrm{eff}(\theta^*) = \frac{I_\theta^{-1}}{D^2_\theta \theta^*}$ nazywamy *efektywnością* estymatora θ^*.

Twierdzenie 18.4. *Jeżeli spełnione są warunki regularności, to*

$$I_\theta = E_\theta \left[-\frac{\partial^2}{\partial \theta^2} \log p_\theta(X) \right]$$

Przykład. *Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $D(\theta)$ z nieznanym parametrem. Skonstruować $ENMW$ dla wariancji $\theta(1-\theta)$.*

Rozwiązanie. *Modelem statystycznym pojedynczej obserwacji X jest*

$$(\{0,1\}, \{D(\theta),\ \theta \in [0,1]\}),$$

natomiast modelem dla próby $X_1, X_2, \ldots, X_n$:

$$(\{0,1\}, \{D(\theta),\ \theta \in [0,1]\})^n.$$

Zauważmy, że rozkład prawdopodobieństwa próby jest z rodziny wykładniczej:

$$\begin{aligned} P_\theta(X_1 = x_1, \ldots, X_n = x_n) &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\ &= \exp\left\{ \log \frac{\theta}{1-\theta} \sum_{i=1}^n x_i + \log(1-\theta) \right\}. \end{aligned}$$

Statystyką dostateczną jest $T = \sum X_i$, a jej modelem statystycznym jest

$$(\{0,1,\ldots,n\}, \{Bin(n,\theta),\ \theta \in [0,1]\}).$$

W celu znalezienia $ENMW$ wariancji $\theta(1-\theta)$ skorzystamy z twierdzenia. Czyli, chcemy wyznaczyć taką funkcję $\hat{g}$ statystyki T, że dla wszystkich $\theta \in \Theta$

$$E_\theta \hat{g}(T) = \sum_{t=0}^n \hat{g}(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = \theta(1-\theta).$$

Dokonując odpowiednich przekształceń otrzymujemy:

$$(1-\theta)^n \sum_{t=0}^{n} \hat{g}(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t = \theta(1-\theta).$$

$$\sum_{t=0}^{n} \hat{g}(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t = \left(\frac{\theta}{1-\theta}\right)\left(1+\left(\frac{\theta}{1-\theta}\right)\right)^{n-2}.$$

Podstawiając $v = \theta/(1-\theta)$ otrzymujemy

$$\sum_{t=0}^{n} \hat{g}(t) \binom{n}{t} v^t = v(1+v)^{n-2}.$$

$$\sum_{t=0}^{n} \hat{g}(t) \binom{n}{t} v^t = \sum_{t=0}^{n-2} \binom{n-2}{t} v^{t+1}.$$

Porównując odpowiednie współczynniki obu wielomianów otrzymujemy:

$$ENMW[\theta(1-\theta)] = \frac{T(n-T)}{n(n-1)}.$$

□

Przykład. *Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Po(\theta)$ z nieznanym parametrem. Skonstruować $ENMW$ dla $P_\theta\{X=0\} = e^{-\theta}$.*

Rozwiązanie. *Modelem statystycznym pojedynczej obserwacji X jest*

$$\left(\{0,1,2,\ldots\}, \{Po(\theta),\ \theta \in \mathbb{R}_+\}\right),$$

natomiast modelem dla próby $X_1, X_2, \ldots, X_n$:

$$\left(\{0,1,2\ldots\}, \{Po(\theta),\ \theta \in \mathbb{R}_+\}\right)^n.$$

Zauważmy, że rozkład prawdopodobieństwa próby jest z rodziny wykładniczej:

$$\begin{aligned} P_\theta(X_1 = x_1, \ldots, X_n = x_n) &= \frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} e^{-n\theta} \\ &= \exp\left\{\log\theta \sum_{i=1}^n x_i - n\theta\right\} \frac{1}{\prod_{i=1}^n x_i!}. \end{aligned}$$

Statystyką dostateczna jest $T = \sum X_i$, a jej modelem statystycznym jest

$$\left(\{0,1,\ldots,n\}, \{Po(n\theta),\ \theta \in [0,1]\}\right).$$

W celu znalezienia ENMW prawdopodobieństwa $e^{-\theta}$ skorzystamy z twierdzenia. Wprowadźmy oznaczenie $\lambda \overset{\text{ozn}}{=\!=} e^{-\theta}$. Chcemy wyznaczyć estymator nieobciążony λ^* wielkości λ. Następnie wyznaczymy $\hat{\lambda} = E_\theta(\lambda^*|T)$ uzyskując ENMW.
Niech

$$Y_j = \begin{cases} 1, & \text{jeżeli } X_j = 0, \\ 0, & \text{jeżeli } X_j > 0 \end{cases}$$

oraz niech $\lambda^* = \frac{1}{n}\sum_{j=1}^{n} Y_j$. Estymator λ^* jest nieobciążony, gdyż

$$E_\theta Y_j = 1 \cdot P_\theta\{X = 0\} + 0 \cdot P_\theta\{X > 0\}.$$

Wyznaczamy teraz $E_\theta(\lambda^*|T)$. Zauważmy, że:

$$\begin{aligned} P_\theta\{X_1 = x_1, \ldots, X_n = x_n | T = t\} &= \frac{P_\theta\{X_1 = x_1, \ldots, X_n = x_n, T = t\}}{P_\theta\{T = t\}} \\ &= \begin{cases} 0, & \text{jeżeli } \sum x_i \neq t \\ \frac{\theta^{x_1 + \cdots x_n} e^{-n\theta}}{x_1! \ldots x_n!} \cdot \frac{t!}{(n\theta)^t e^{-n\theta}}, & \text{jeżeli } \sum x_i = t, \end{cases} \\ &= \begin{cases} 0, & \text{jeżeli } \sum x_i \neq t \\ \frac{t!}{n^t x_1! \ldots x_n!}, & \text{jeżeli } \sum x_i = t. \end{cases} \end{aligned}$$

A zatem

$$\begin{aligned} E_\theta(\lambda^*|T = t) &= E_\theta\left(\frac{1}{n}\sum_{j=1}^{n} Y_j | T = t\right) \\ &= E_\theta\left(Y_1 | T = t\right) \\ &= P_\theta\{X_1 = 0 | T = t\} \\ &= \sum_{x_2 + \cdots + x_n = t} P_\theta\{X_1 = 0, X_2 = x_2, \ldots, X_n = x_n | T = t\} \\ &= \sum_{x_2 + \cdots + x_n = t} \frac{t!}{n^t x_2! \ldots x_n!} \\ &= \frac{(n-1)^t}{n^t}. \end{aligned}$$

Ostatnia równość wynika z faktu, że

$$\sum_{x_1 + \cdots + x_n = t} \frac{t!}{x_1! \ldots x_n!} = n^t.$$

(Dla przypomnienia: $(\alpha_1 + \cdots + \alpha_n)^t = \sum_{x_1 + \cdots + x_n = t} \frac{t!}{x_1! \ldots x_n!} \alpha_1^{x_1} \cdots \alpha_n^{x_n}$.) Otrzymujemy następujący ENMW prawdopodobieństwa $e^{-\theta}$:

$$ENMW[e^{-\theta}] = \left(1 - \frac{1}{n}\right)^T.$$

□

Zadania do samodzielnego rozwiązania

18.1. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $D(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0, 1)$. Wyznaczyć $ENMW$ dla parametru θ.

18.2. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Bin\,(m, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć $ENMW$ dla parametru θ.

18.3. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Ge\,(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć $ENMW$ dla parametru θ.

18.4. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $NB\,(r, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć $ENMW$ dla parametru θ.

18.5. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Po\,(\lambda)$ z nieznanym parametrem $\lambda > 0$. Wyznaczyć $ENMW$ dla parametru λ.

18.6. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U\,(0, \theta)$ z nieznanym parametrem $\theta > 0$. Wyznaczyć $ENMW$ dla parametru θ.

18.7. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U\,(\theta, 0)$ z nieznanym parametrem $\theta < 0$. Wyznaczyć $ENMW$ dla parametru θ.

18.8. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E\,(\lambda)$ z nieznanym parametrem $\lambda > 0$. Wyznaczyć $ENMW$ dla parametru λ.

18.9. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $G\,(\alpha, \lambda)$ z nieznanym parametrem $\lambda > 0$ i znanym parametrem α. Wyznaczyć $ENMW$ dla parametru λ.

18.10. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznaną wartością oczekiwaną μ i znaną wariancją $\sigma^2 > 0$. Wyznaczyć $ENMW$ dla parametru μ.

18.11. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami. Wyznaczyć $ENMW$ dla parametru (μ, σ^2).

18.12. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu skoncentrowanego na dodatniej półosi i niech gęstość tego rozkładu wyraża się wzorem $\theta^2 x e^{-\theta x}$, $\theta > 0$. Wyznaczyć $ENMW$ parametru θ.

18.13. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $Po\,(\theta)$ o nieznanej średniej θ. Znaleźć $ENMW$ parametru $e^{-\theta}$ (prawdopodobieństwa, że zmienna losowa przyjmie wartość zero). Obliczyć wariancję tego estymatora.

18.14. Niech $X_1, X_2, \ldots, X_n$, gdzie $n > 3$, będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ o nieznanych μ i σ^2. Znaleźć $ENMW$ dla μ^2/σ^2.

18.15. Zmienne losowe $X_1, \ldots, X_n$ mają taką samą wartość oczekiwaną μ. Wiadomo, że

$$Cov(X_i, X_j) = \begin{cases} \sigma^2, & \text{dla } i = j, \\ \frac{\sigma^2}{2}, & \text{dla } i \neq j. \end{cases}$$

Niech $S^2(c) = c\,\mathrm{var}X$. Dla jakiej wartości c, statystyka $S^2(c)$ jest nieobciążonym estymatorem parametru σ^2?

18.16. Dana jest n-elementowa próba z rozkładu skoncentrowanego na dodatniej półosi, o gęstości

$$\frac{x+1}{\theta(\theta+1)}e^{-x/\theta},$$

gdzie θ jest nieznanym parametrem dodatnim. Wyznaczyć taki estymator wielkości $(3+2\theta)(2+\theta)/(\theta+1)$, którego wariancja osiąga dolne ograniczenie Cramera-Rao.

18.17. Niech $X_1,\ldots,X_n$ będzie próbą z rozkładu $Wei\,(2,\theta)$, gdzie $\theta>0$ jest nieznanym parametrem. Rozważamy nieobciążony estymator parametru θ postaci $T_n=aY$, gdzie $Y=\min\{X_1^2,\ldots,X_n^2\}$ i a jest odpowiednio dobraną stałą (być może zależną od liczebności próby n). Pokazać, że

$$(\forall\ \theta>0)\ (\forall\ 0<\varepsilon<1)\ P_\theta(|T_n-\theta|>\varepsilon)=1-\exp(-1)\left(\exp\left(\frac{\varepsilon}{\theta}\right)-\exp\left(-\frac{\varepsilon}{\theta}\right)\right).$$

18.18. Dla $t=1,\ldots,T$ obserwujemy niezależne realizacje zmiennej losowej X_t, o których zakładamy, iż pochodzą z rozkładu o parametrach $EX_t=n_t\mu$ i $D^2X_t=n_t\sigma^2$, gdzie $n_1,\ldots,n_T$ są znanymi liczbami dodatnimi, natomiast parametry μ oraz σ^2 są nieznane. Wybieramy estymator parametru σ^2 z klasy estymatorów postaci $c\sum_{t=1}^T(X_t-n_t\bar{X})^2$, gdzie $\bar{X}=\frac{1}{n}\sum_{t=1}^T X_t$ oraz $n=\sum_{t=1}^T n_t$, zaś c jest pewną liczbą rzeczywistą (parametrem konkretnego estymatora). Dla jakiej wartości c otrzymamy estymator nieobciążony.

18.19. Rozpatrzmy następujący model regresji liniowej bez wyrazu wolnego:

$$Y_i=\beta x_i+\varepsilon_i,\qquad i=1,\ldots,n,$$

gdzie x_i są znanymi liczbami, β jest nieznanym parametrem, zaś ε_i są błędami losowymi. Zakładamy, że

$$E\varepsilon_i=0,\quad D^2\varepsilon_i=x_i^2\sigma^2,\qquad i=1,\ldots,n.$$

Skonstruować estymator $\hat{\beta}$ parametru β o następujących własnościach:

$\hat{\beta}$ jest liniową funkcją obserwacji, tzn. $\hat{\beta}=\sum_{i=1}^n c_iY_i$,

$\hat{\beta}$ jest nieobciążony, tzn. $E\hat{\beta}=\beta$,

$\hat{\beta}$ ma najmniejszą wariancję spośród estymatorów liniowych i nieobciążonych.

18.20. Niech (X,Y) będzie dwuwymiarową zmienną losową, o wartości oczekiwanej (μ_X,μ_Y), wariancji każdej ze współrzędnych równej σ^2 oraz kowariancji równej $\varrho\sigma^2$. Staramy się obserwować niezależne realizacje tej zmiennej, ale nie w pełni to wychodzi – czasem udaje się zaobserwować jedynie pierwszą lub jedynie drugą ze współrzędnych. Przyjmijmy ważne założenie, iż do „zgubienia” obserwacji (całkowitego, jej pierwszej współrzędnej lub jej drugiej współrzędnej) dochodzi całkowicie niezależnie od wartości tych obserwacji. Załóżmy, iż otrzymaliśmy próbę, zawierającą 20 obserwacji wyłącznie pierwszej współrzędnej, 60 obserwacji całej pary oraz 20 obserwacji wyłącznie drugiej współrzędnej. Niech teraz $\overline{X}$ oznacza średnią z próby (osiemdziesięciu) obserwacji na zmiennej X, $\overline{Y}$ oznacza średnią z próby (osiemdziesięciu) obserwacji na zmiennej Y oraz niech $\overline{X-Y}$ oznacza średnią z próby (sześćdziesięciu) obserwacji na różnicy zmiennych $X-Y$. Niech $\overline{X}-\overline{Y}$ oraz $\overline{X-Y}$ oznaczają dwa alternatywne estymatory różnicy $\mu_X-\mu_Y$. Dla jakiej wartości współczynnika ϱ estymatory te mają jednakową wariancję?

18.21. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi, przy czym $EX_i = i\mu$ oraz $D^2X_i = i^2\mu^2$, $i = 1, \ldots, n$. Niech $\tilde{\mu}$ będzie estymatorem parametru μ minimalizującym błąd średniokwadratowy w klasie estymatorów postaci $\tilde{\mu} = \sum_{i=1}^n a_i X_i$, gdzie $a_1, \ldots, n$ są liczbami rzeczywistymi. Wyznaczyć błąd średniokwadratowy $E_\mu(\tilde{\mu} - \mu)^2$.

18.22. Niech $X_1, \ldots, X_n$, gdzie $n > 1$, będzie próbą z rozkładu o gęstości

$$f_c(x) = \begin{cases} \frac{4c^4}{x^5}, & \text{gdy } x > c, \\ 0, & \text{poza tym,} \end{cases}$$

gdzie $c > 0$ jest nieznanym parametrem. Rozważamy dwa estymatory parametru c: $T_1 = a\min\{X_1, \ldots, X_n\}$ i $T_2 = b\overline{X}$, gdzie a, b są dobrane tak, aby estymatory były nieobciążone. Wyznaczyć różnicę ryzyk estymatorów, czyli

$$R = E_c(T_2 - c)^2 - E_c(T_1 - c)^2.$$

18.23. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi, przy czym zmienna losowa X_i ma rozkład $N\left(\mu, i\mu^2\right)$, $i = 1, \ldots, n$, gdzie $\mu \neq 0$ jest nieznanym parametrem. Rozważamy estymatory parametru μ postaci $\hat{\mu} = \sum_{i=1}^n a_i X_i$. Znaleźć współczynniki a_i minimalizujące błąd średniokwadratowy estymatora.

18.24. Niech $Y_1, \ldots, Y_n$ będą niezależnymi zmiennymi losowymi, przy czym Y_k ma rozkład $N\left(k\mu, \sigma^2\right)$, dla $k = 1, \ldots, n$. Znaleźć $ENMW$ parametru μ w klasie estymatorów postaci $\hat{\mu} = \sum_{k=1}^n a_k Y_k$.

18.25. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \gamma^2\sigma^2\right)$, gdzie $\mu \in R$ jest nieznanym parametrem, zaś γ^2 znanym współczynnikiem. Dla jakich współczynników $c_1 \ldots, c_n$ estymator postaci $\hat{\mu} = \sum_{i=1}^n c_i X_i$ parametru μ ma jednostajnie najmniejszy błąd średniokwadratowy $E_\mu(\hat{\mu} - \mu)^2$.

18.26. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu normalnego o nieznanych parametrach (μ, σ^2) i niech $n > 1$ oraz $\sigma^2 > 0$. Niech

$$S^2 = \frac{1}{n-1}\mathrm{var}X, \quad t(\mu_0) = \frac{\overline{X} - \mu_0}{\sqrt{S^2}},$$

gdzie μ_0 jest ustaloną liczbą. Niech $t(\alpha; n-1)$ będzie dwustronną wartością krytyczną rozkładu $t\,(n-1)$. Rozważmy następujący estymator $\hat{\mu}$ parametru μ:

$$\hat{\mu} = \begin{cases} \mu_0, & \text{jeżeli } |t(\mu_0)| < t(\alpha; n-1), \\ \bar{X}, & \text{w przeciwnym przypadku.} \end{cases}$$

Dla jakich μ obciążenie $E_\mu\hat{\mu} - \mu$ estymatora jest dodatnie?

18.27. Wykonano n pomiarów pewnej nieznanej wielkości μ jednym przyrządem pomiarowym, a następnie m pomiarów innym przyrządem. Zakładamy, że wyniki pomiarów $X_1, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}$ są niezależnymi zmiennymi losowymi, przy czym każda ze zmiennych $X_1, \ldots, X_n$ ma rozkład $N\,(\mu, 0.01)$, a każda ze zmiennych $X_{n+1}, \ldots, X_{n+m}$ ma rozkład $N\,(\mu, 0.04)$. Dobrać współczynniki $c_1, \ldots, c_{n+m}$ tak, żeby estymator $\hat{\mu} = \sum_{i=1}^{n+m} c_i X_i$ był $ENMW$ parametru μ.

18.28. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$. Pokazać, że estymator $\frac{1}{n}\mathrm{var}X$ parametru σ^2 ma mniejszy błąd średniokwadratowy niż $ENMW$ $\frac{1}{n-1}\mathrm{var}X$.

18.29. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu normalnego o nieznanej średniej i wariancji. Rozpatrzmy klasę estymatorów wariancji określonych wzorem $S(c) = c\mathrm{var}X$, gdzie c jest pewną liczbą rzeczywistą. Wyznaczyć wartość c, przy której błąd średniokwadratowy estymatora $S(c)$ osiąga minimum.

18.30. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$. Niech $S^2 = \frac{1}{n-1}\mathrm{var}X$. Interesuje nas względny błąd estymacji: $R = \frac{S^2 - \sigma^2}{\sigma^2}$. Wyznaczyć rozmiar n próby, dla którego $E(R^2) = 0.01$.

18.31. Niech X_{ij}, $i = 1, \ldots, k$, $j = 1, \ldots, n$, będzie próbą z rozkładu normalnego o nieznanej wartości oczekiwanej μ i wariancji σ^2. Niech $Y_i = \sum_{j=1}^n X_{ij}$. Szacujemy wariancję σ^2 używając estymatora postaci $c\sum_{i=1}^k (Y_i - \overline{Y})^2$. Wyznaczyć stałą c tak, by estymator był nieobciążony.

18.32. Niech $X_1, \ldots, X_n, \ldots, X_{n+m}$, $m, n > 1$, będzie próbą z rozkładu $N(\mu, \sigma^2)$. Bezpośrednio dostępne są tylko obserwacje $X_1, \ldots, X_n$, ale znamy średnią $\bar{X}_{n+m} = \frac{1}{n+m}\sum_{i=1}^{n+m} X_i$. Dla jakiej liczby c estymator $c\sum_{i=1}^n (X_i - \bar{X}_{n+m})^2$ wariancji σ^2 jest nieobciążony?

18.33. Niech $x_1, \ldots, x_{25}$ będzie realizacją próby z rozkładu $N(\mu, \sigma^2)$, zaś $x_{26}, \ldots, x_{50}$ - realizacją próby losowej z rozkładu $N(\nu, \tau^2)$, gdzie $\mu, \nu, \sigma^2, \tau^2$ są nieznanymi parametrami. Wiadomo, że

$$\bar{x}_{25} = \frac{1}{25}\sum_{i=1}^{25} x_i = 10.4, \quad \bar{x}_{50} = \frac{1}{50}\sum_{i=1}^{50} x_i = 10.0,$$

$$s_{25}^2 = \frac{1}{24}\sum_{i=1}^{25}(x_i - \bar{x}_{25})^2 = 3.333, \quad s_{50}^2 = \frac{1}{49}\sum_{i=1}^{50}(x_i - \bar{x}_{50})^2 = 2.000.$$

Obliczyć na tej podstawie wartość nieobciążonego estymatora wariancji τ^2.

18.34. Mamy dwie niezależne obserwacje X_1 oraz X_2 z rozkładu normalnego, przy czym jedna z nich pochodzi z rozkładu o parametrach (μ, σ^2), a druga z rozkładu o parametrach $(2\mu, 2\sigma^2)$. Niestety zgubiliśmy informację, która z obserwacji z którego z rozkładów pochodzi. Parametry (μ, σ^2) są nieznane. W tej sytuacji wybieramy estymator parametru σ^2 z klasy estymatorów postaci

$$\hat{\sigma}^2 = a(X_1 - X_2)^2 + b(X_1 + X_2)^2,$$

gdzie (a, b) to para liczb rzeczywistych (parametry konkretnego estymatora). Dla jakich (a, b) otrzymamy estymator nieobciążony?

18.35. Rozważmy dwie niezależne próby $X_{1,1},\ldots,X_{1,n_1}$ z rozkładu $N\left(\mu,\sigma^2\right)$ oraz $X_{2,1},\ldots,X_{2,n_2}$ z rozkładu $N\left(\mu,2\sigma^2\right)$. Niech $\bar{X}_1=\frac{1}{n_1}\sum_{i=1}^{n_1}X_{1,i}$, $\bar{X}_2=\frac{1}{n_2}\sum_{i=1}^{n_2}X_{2,i}$ oraz $\bar{X}=\frac{n_1\bar{X}_1+n_2\bar{X}_2}{n_1+n_2}$. Dla jakiego c estymator parametru σ^2 postaci

$$c\left\{\sum_{i=1}^{n_1}(X_{1,i}-\bar{X})^2+\sum_{i=1}^{n_2}(X_{2,i}-\bar{X})^2\right\}$$

jest nieobciążony?

18.36. Niech $X_1,\ldots,X_n,X_{n+1},\ldots,X_{n+m}$ będą niezależnymi zmiennymi losowymi, przy czym zmienne losowe $X_1,\ldots,X_n$ mają rozkład $N\left(\mu_1,\sigma^2\right)$, zaś $X_{n+1}\ldots,X_{n+m}$ mają rozkład $N\left(\mu_2,k\sigma^2\right)$, gdzie $k>0$ jest znaną liczbą. Niech

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad \bar{X}_2=\frac{1}{m}\sum_{i=n+1}^{n+m}X_i,\quad \bar{X}=\frac{1}{n+m}\sum_{i=1}^{n+m}X_i.$$

Dobrać liczby α i β tak, żeby statystyka

$$\hat{\sigma}^2=\alpha\sum_{i=1}^{n+m}(X_i-\bar{X})^2+\beta(\bar{X}_1-\bar{X}_2)^2$$

była nieobciążonym estymatorem parametru σ^2.

18.37. Niech $X_1,\ldots,X_n$ będzie próbą z rozkładu $N\left(\mu,\sigma^2\right)$ z nieznanymi parametrami. Niech $S^2=\frac{1}{n-1}\mathrm{var}X$ będzie nieobciążonym estymatorem wariancji. Obliczyć $P(S^2\le\sigma^2)$.

18.38. $X_1,\ldots,X_n,X_{n+1},\ldots,X_m$ jest próbą z rozkładu $N\left(\mu,\sigma^2\right)$ z nieznanymi parametrami μ oraz σ^2. Obserwujemy zmienne $X_1,\ldots,X_n$ i ponadto znamy średnią ze wszystkich zmiennych $\bar{X}_m=\frac{1}{m}\sum_{i=1}^{m}X_i$. Znaleźć stałą $c_{n,m}$ taką, żeby statystyka

$$\frac{1}{c_{n,m}}\sum_{i=1}^{n}(X_i-\bar{X}_m)^2$$

była nieobciążonym estymatorem wariancji σ^2.

18.39. Niech $X_1,\ldots,X_n$ będzie próbą z rozkładu $N\left(\mu,\sigma^2\right)$, gdzie $\mu\in R$, $\sigma^2>0$ są nieznanymi parametrami. Wyznaczyć $ENMW$ parametru μ/σ.

18.40. Załóżmy, że $X_1,\ldots,X_n$ jest próbą z rozkładu $N\left(\mu,\sigma^2\right)$ z nieznanymi parametrami. Rozważmy nieobciążony estymator wielkości μ^2 dany wzorem

$$\overline{\mu^2}=(\overline{X})^2-\frac{S^2}{n},$$

gdzie $S^2=\frac{1}{n-1}\mathrm{var}X$. Obliczyć $D^2(\overline{\mu^2})$.

18.41. Niech $(X_1, Y_1), \ldots, (X_n, Y_n)$ będzie próbą z rozkładu $N_2\left((\mu_X, \mu_Y), \begin{bmatrix} 1 & \varrho \\ \varrho & 1 \end{bmatrix}\right)$. Dla jakich wartości ϱ statystyka $\min\left\{\overline{X}, \overline{Y}\right\}$ jest nieobciążonym estymatorem parametru $\min\{\mu_X, \mu_Y\}$?

18.42. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $E\,(1/\theta)$, $\theta > 0$. Pokazać, że zmienna losowa przyjmująca wartość 1, gdy $X_1 \geq k$ i wartość 0, gdy $X_1 < k$, jest estymatorem nieobciążonym dla $g(\theta) = e^{-k\theta}$. Na tej podstawie pokazać, że przy odpowiednim wyborze statystyki T, zmienna losowa

$$\hat{g}(T) = \begin{cases} 0, & \text{gdy } T < k, \\ \left[\frac{T-k}{k}\right]^{n-1}, & \text{gdy } T \geq k \end{cases}$$

jest $ENMW$ dla g.

18.43. Niech $X_1, \ldots, X_n$, gdzie $n > 1$, będzie próbą z rozkładu $E\,(\mu)$. Niech

$$\hat{\mu}_1 = \overline{X}, \quad \hat{\mu}_2 = n \min\{X_1, \ldots, X_n\}$$

będą dwoma estymatorami parametru μ. Pokazać, że oba estymatory są nieobciążone oraz, że estymator $\hat{\mu}_1$ ma zawsze mniejszą wariancję niż $\hat{\mu}_2$.

18.44. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E\,(1/\lambda)$, gdzie $\lambda > 0$ jest nieznanym parametrem oraz $n > 1$. Niech $a > 0$ będzie daną liczbą. Interesuje nas estymacja parametru $p = e^{-\lambda a} = P_\lambda(X_1 > a)$. Niech N_a oznacza liczbę obserwacji większych od a, zaś $S = \sum_{i=1}^{n} X_i$. Rozważmy trzy estymatory parametru p:

$$\hat{p}_1 = \exp\left(-a\frac{n}{S}\right), \quad \hat{p}_2 = \frac{N_a}{N}, \quad \hat{p}_3 = \begin{cases} 0, & \text{gdy } S \leq a, \\ \left(\frac{S-a}{S}\right)^{n-1}, & \text{gdy } S > a \end{cases}.$$

Które z estymatorów są nieobciążone? Wyznaczyć wariancje tych estymatorów.

18.45. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E\,(1/\lambda)$. Parametr $\lambda > 0$ jest nieznany. Znaleźć taką liczbę c, żeby $c(\overline{X})^2$ był nieobciążonym estymatorem wariancji pojedynczej zmiennej X_i.

18.46. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E\,(\mu; c)$, gdzie $c \in R$ i $\mu > 0$ są nieznanymi parametrami. Wyznaczyć nieobciążony estymator parametru μ.

18.47. Niech $W_1, \ldots, W_n$ $(n > 1)$ będzie próbą z rozkładu $E\,(\mu)$. Rozważmy estymatory parametru μ postaci

$$\hat{\mu} = aS, \quad \text{gdzie} \quad S = \sum_{i=1}^{n} W_i.$$

Znaleźć liczbę a, dla której błąd średniokwadratowy estymatora jest najmniejszy.

18.48. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $U(0, \theta)$. Pokazać, że statystyką dostateczną dla θ jest $\max(X_1, X_2, \ldots, X_n)$. Pokazać, że istnieje estymator nieobciążony tego parametru postaci $k \max(X_1, X_2, \ldots, X_n)$. Wyznaczyć ten estymator nieobciążony. Porównać jego wariancję z dolnym ograniczeniem Cramera-Rao dla wariancji estymatorów nieobciążonych i uzasadnić, dlaczego nie można w tym przypadku stosować nierówności Cramera-Rao.

18.49. Niech X będzie obserwacją z rozkładu $U(\theta - 0.5, \theta + 0.5)$ z nieznanym parametrem θ. Wiemy, że θ jest liczbą rzeczywistą. Za pomocą estymatora $|X|$ estymujemy wartość bezwzględną parametru θ. Wyznaczyć maksymalne obciążenie $E_\theta|X| - \theta$ tego estymatora.

18.50. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(\varphi_0, \varphi_1)$ z nieznanymi oboma parametrami i niech $n > 1$. Interesuje nas szerokość przedziału $\varphi_1 - \varphi_0$. Dobrać parametr a tak, aby estymator $a(\max\{X_1, \ldots, X_n\} - \min\{X_1, \ldots, X_n\})$ szerokości przedziału był nieobciążony.

18.51. Niech $X_1, X_2, \ldots, X_n, \ldots$ będzie próbą z rozkładu $U(0, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Rozważamy estymator nieobciążony parametru θ postaci $T_n(X_1, X_2, \ldots, X_n) = T_n = a \min\{X_1, X_2, \ldots, X_n\}$ i a jest pewną stałą. Udowodnić, że

$$(\exists\ \varepsilon > 0)(\exists\ \theta > 0)\ \lim_{n\to\infty} P_\theta(|T_n - \theta| > \varepsilon) = 1 + \exp\left(-1 - \frac{\varepsilon}{\theta}\right) - \exp\left(-1 + \frac{\varepsilon}{\theta}\right).$$

18.52. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(0, \varphi)$ z nieznanym $\varphi > 0$. Udowodnić, że estymator $\frac{n+1}{n} \max\{X_1, \ldots, X_n\}$ jest nieobciążony. Wyznaczyć wariancję tego estymatora.

18.53. Załóżmy, że K oznacza liczbę sukcesów w n próbach Bernoulliego z nieznanym prawdopodobieństwem sukcesu θ. Rozważmy estymator parametru θ postaci

$$\hat{\theta} = \frac{a + K}{b + n}.$$

Niech $n = 16$. Przypuśćmy, że dodatnie liczby a i b dobrane zostały tak, że funkcja ryzyka estymatora, $R(\theta) = E_\theta(\hat{\theta} - \theta)^2$ jest funkcją stałą, czyli $R(\theta) = R$ dla każdej wartości parametru θ. Podać liczbę R.

18.54. $X_1, \ldots, X_n$ jest próbą z rozkładu $Ge(p)$, gdzie $p \in (0, 1)$ i liczebność próby przekracza 1. W klasie estymatorów parametru p danych wzorem:

$$\frac{a}{a + \sum_{i=1}^n X_i}$$

dobrać parametr a tak, aby otrzymać estymator nieobciążony.

18.55. Entomolog pobierał próbę z dużej populacji pewnych owadów. Notował płeć chwytanych osobników męskich i przerwał pobieranie próbki, gdy otrzymał M ($M > 1$) osobników męskich. Otrzymał próbę o liczebności X. Niech θ będzie frakcją osobników męskich w populacji. Znaleźć $ENMW$ parametru θ.

18.56. Bada się liczbę owadów na liściach pewnej rośliny. Stwierdzono, że X_i zbadanych liści miało dokładnie i owadów ($i = 1, 2, \ldots, \sum X_i = N$). Zakłada się, że liczba owadów na liściu jest zmienną losową o rozkładzie Poissona z tą jednak różnicą, że na wielu liściach nie stwierdzono ani jednego owada nie w wyniku losowych fluktuacji liczby owadów, ale dlatego, że liście te nie nadawały się na pokarm dla owadów. W związku z tym „pustych” liści nie brano pod uwagę. Pokazać, że

$$\sum_{i=2}^{\infty} \frac{iX_i}{N}$$

jest nieobciążonym estymatorem średniej μ rozkładu Poissona i wyznaczyć efektywność tego estymatora.

18.57. Pobieramy próbę $X_1, \ldots, X_n$ z rozkładu $Po(\lambda)$ o nieznanym parametrze λ. Szacujemy $p_0 = e^{-\lambda}$ za pomocą estymatora $\hat{p}_0 = e^{-\overline{X}}$. Wyznaczyć znak obciążenia $E_\lambda \hat{p}_0 - p_0$ tego estymatora.

18.58. Niech $N_1, \ldots, N_n$ będzie próbą z rozkładu $Po(\lambda)$ o nieznanym parametrze λ. Dla jakiej wartości C estymator $C^{\overline{N}}$ parametru $e^{-\lambda}$ będzie nieobciążony?

18.59. Niech $N_1, \ldots, N_{10}$ będzie próbą z rozkładu $Po(\lambda)$ z nieznanym parametrem λ. Interesuje nas drugi moment obserwacji, czyli wielkość $m_2(\lambda) = E_\lambda(N_1^2)$. Skonstruować taki estymator wielkości $m_2(\lambda)$, który jest nieobciążony i który jest funkcją zmiennej $S = N_1 + \cdots + N_{10}$.

18.60. Pobieramy próbę z rozkładu $Po(\lambda)$ o nieznanym parametrze $\lambda > 0$. Niestety sposób obserwacji uniemożliwia odnotowanie realizacji o wartości 0. Pobieranie próby kończymy w momencie, gdy liczebność odnotowanych realizacji wynosi n. Tak więc, każda z naszych kolejnych odnotowanych realizacji $K_1, \ldots, K_n$ wynosi co najmniej 1 i nie wiemy, ile w międzyczasie pojawiło się obserwacji o wartości 0. Estymujemy parametr λ za pomocą estymatora postaci

$$\hat{\lambda} = \sum_{i=2}^{\infty} iN_i,$$

gdzie N_i jest liczbą obserwacji o wartości i. Obliczyć wariancję estymatora $\hat{\lambda}$.

18.61. Proces pojawiania się szkód jest procesem Poissonowskim z parametrem intensywności λ, tzn. prawdopodobieństwo pojawienia się n szkód na odcinku czasu $(0, T]$ jest równe $\frac{(\lambda T)^n}{n!} e^{-\lambda T}$. Obserwujemy proces od momentu 0. Niech $T_1, T_2, T_3, \ldots$ oznaczają momenty pojawiania się kolejnych szkód. Ustalamy z góry liczbę n taką, że obserwację procesu przerwiemy w momencie T_n pojawienia się n-tej szkody. Dla jakiej wartości C estymator $\frac{C}{T_n}$ parametru λ jest estymatorem nieobciążonym?

18.62. Sygnały pojawiają się zgodnie z procesem Poissona, a oczekiwana liczba sygnałów na jednostkę czasu wynosi λ. Obserwujemy proces od momentu T_0 do momentu T_n pojawienia się n-tego sygnału, przy czym n jest z góry ustaloną liczbą całkowitą równą co najmniej 2. Wyznaczyć nieobciążony estymator parametru λ.

18.63. Niech $X = (X_1, \ldots, X_k)$ będzie zmienną losową o rozkładzie wielomianowym z parametrami $(n, p_1, \ldots, p_k)$, gdzie wektor $\boldsymbol{p} = (p_1, \ldots, p_k)$ ($p_i \geq 0$ dla $i = 1, \ldots, k$ oraz $\sum_{i=1}^{k} p_i = 1$) jest wektorem nieznanych parametrów. Rozważamy problem estymacji wektora $\boldsymbol{p}$ przy kwadratowej funkcji straty $L(\hat{\boldsymbol{p}}, \boldsymbol{p}) = \frac{1}{k} \sum_{i=1}^{k} (\hat{p}_i - p_i)^2$. Wśród estymatorów wektora $\boldsymbol{p}$ postaci $\hat{\boldsymbol{p}} = (aX_1 + b, \ldots, aX_k + b)$ (gdzie a i b są liczbami rzeczywistymi) o ryzyku (tzn. $EL(\hat{\boldsymbol{p}}, \boldsymbol{p})$) stałym, niezależnym od $\boldsymbol{p}$, wyznaczyć estymator o najmniejszym ryzyku.

19. Metoda największej wiarogodności

Niech $(\mathcal{X}, \mathcal{B}, \{P_\theta,\ \theta \in \Theta\})$ będzie modelem statystycznym.

Definicja 19.1 Dla ustalonego $x \in \mathcal{X}$ wielkość $L(\theta; x) = p_\theta(x)$ nazywamy **wiarogodnością parametru** θ, gdy zaobserwowano x.

Definicja 19.2 Jeżeli przy każdym ustalonym $x \in \mathcal{X}$ istnieje $\hat{\theta} = \hat{\theta}(x) \in \Theta$ takie, że

$$L(\hat{\theta}; x) \geq L(\theta; x) \quad (\forall \theta \in \Theta)\,,$$

to odwzorowanie $\hat{\theta} : \mathcal{X} \to \Theta$ nazywamy **estymatorem największej wiarogodności**.

Estymator największej wiarogodności parametru θ oznaczamy przez $ENW[\theta]$

Twierdzenie 19.1. *Jeżeli $h : \Theta \to \Theta$ jest funkcją różnowartościową oraz $\hat{\theta}$ jest $ENW[\theta]$, to $h(\hat{\theta})$ jest $ENW[h(\theta)]$.*

Definicja 19.3 Jeżeli $h : \Theta \to \Theta$, to $ENW[h(\theta)]$ określamy jako $h(\hat{\theta})$, gdzie $\hat{\theta}$ jest $ENW[\theta]$.

W wielu zagadnieniach zamiast szukać maksimum funkcji wiarogodności $L(\cdot; x)$ wygodniej jest analizować $\mathcal{L}(\cdot; x) = \log L(\cdot; x)$.

Niech $p_\theta(x)$ spełniają pewne warunki regularności.

Twierdzenie 19.2. *Jeżeli $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi o tym samym rozkładzie o gęstości $p_\theta(x)$, to równanie wiarogodności*

$$\frac{\partial}{\partial\theta}[p_\theta(x_1)\cdots p_\theta(x_n)] = 0$$

ma rozwiązanie $\hat{\theta}_n(x_1, \ldots, x_n)$ takie, że

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P_{\theta_0}\left\{\left|\hat{\theta}_n(X_1, \ldots, X_n) - \theta_0\right| < \varepsilon\right\} = 1$$

oraz

$$(*) \qquad \sqrt{n}\left(\hat{\theta}_n(X_1, \ldots, X_n) - \theta_0\right) \to N\left(0, \frac{1}{I_\theta}\right).$$

Definicja 19.4 Estymator $\hat{\theta}_n$ spełniający warunek $(*)$ nazywamy **asymptotycznie efektywnym**.

Przykład. *Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami. Skonstruować ENW dla obu nieznanych parametrów.*

Rozwiązanie. *Powiedzmy, że zaobserwowano $X_1 = x_1, \ldots, X_n = x_n$. Wiarogodność parametru (μ, σ^2) jest równa*

$$L(\mu, \sigma^2; x_1, \ldots, x_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\}.$$

Chcemy znaleźć $\hat{\mu}$ i $\hat{\sigma^2}$ maksymalizujące wiarogodność. Ponieważ funkcja wiarogodności jest ciągłą i różniczkowalną funkcją swoich argumentów, więc można posłużyć się standardowymi metodami poszukiwania ekstremum. Niemniej jednak, wygodniej będzie analizować logarytm funkcji wiarogodności:

$$\begin{aligned}\mathcal{L}(\mu, \sigma^2; x_1, \ldots, x_n) &= \log L(\mu, \sigma^2; x_1, \ldots, x_n)\\ &= -n\log\sqrt{2\pi} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.\end{aligned}$$

W celu znalezienia poszukiwanych estymatorów rozwiązujemy układ równań:

$$\begin{cases} \dfrac{\partial\mathcal{L}(\mu, \sigma^2; x_1, \ldots, x_n)}{d\mu} = 0,\\ \dfrac{\partial\mathcal{L}(\mu, \sigma^2; x_1, \ldots, x_n)}{d\sigma^2} = 0.\end{cases}$$

Łatwe rachunki dają:

$$\begin{cases} \dfrac{\partial\mathcal{L}}{\partial\mu} = -\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i - \mu) = 0,\\ \dfrac{\partial\mathcal{L}}{\partial\sigma^2} = -\dfrac{n}{2}\dfrac{1}{\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i - \mu)^2 = 0.\end{cases}$$

Rozwiązując powyższy układ równań względem μ i σ^2 otrzymujemy

$$\mu_* = \overline{X} \qquad \text{oraz} \qquad \sigma^2_* = \frac{1}{n}\mathrm{var}X.$$

W celu sprawdzenia, że logarytm funkcji wiarogodności w wyznaczonym punkcie osiąga maksimum, wystarczy sprawdzić, że Hesjan

$$\begin{bmatrix} \frac{\partial^2\mathcal{L}}{\partial\mu^2} & \frac{\partial^2\mathcal{L}}{\partial\mu\partial\sigma^2}\\ \frac{\partial^2\mathcal{L}}{\partial\mu\partial\sigma^2} & \frac{\partial^2\mathcal{L}}{\partial(\sigma^2)^2}\end{bmatrix}$$

w punkcie (μ_*, σ_*^2) *jest macierzą ujemnie określoną.*

Ostatecznie, estymatorami największej wiarogodności są:

$$ENW[\mu] = \overline{X} \qquad \textit{oraz} \qquad ENW[\sigma^2] = \frac{1}{n}\mathrm{var}X$$

□

Zadania do samodzielnego rozwiązania

19.1. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $D(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć ENW dla parametru θ.

19.2. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Bin(m, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć ENW dla parametru θ.

19.3. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Ge(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć ENW dla parametru θ.

19.4. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $NB(r, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć ENW dla parametru θ.

19.5. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Po(\lambda)$ z nieznanym parametrem $\lambda > 0$. Wyznaczyć ENW dla parametru λ.

19.6. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(0, \theta)$ z nieznanym parametrem $\theta > 0$. Wyznaczyć ENW dla parametru θ.

19.7. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(\theta, 0)$ z nieznanym parametrem $\theta < 0$. Wyznaczyć ENW dla parametru θ.

19.8. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E(\lambda)$ z nieznanym parametrem $\lambda > 0$. Wyznaczyć ENW dla parametru λ.

19.9. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $G(\alpha, \lambda)$ z nieznanym parametrem $\lambda > 0$ i znanym parametrem α. Wyznaczyć ENW dla parametru λ.

19.10. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$ ze znaną wariancją σ^2. Wyznaczyć ENW dla parametru μ.

19.11. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$ ze znaną średnią μ i nieznaną wariancją σ^2. Wyznaczyć ENW dla parametru σ^2.

19.12. Wykonano n doświadczeń Bernoulliego z prawdopodobieństwem sukcesu $p = 1/3$. Liczba n jest nieznanym parametrem. Okazało się, że liczba porażek jest o cztery większa od liczby sukcesów. Wyznaczyć wartość $ENW[n]$.

19.13. W urnie znajduje się razem N kul: białych i czarnych. Wylosowano n kul, wśród których było k kul białych. Wyznaczyć ENW liczby kul białych w urnie.

19.14. Mieszkańcy wyspy Kalythos (na morzu Egejskim) cierpią na dziedziczną chorobę oczu, której objawy potęgują się wraz z wiekiem. Zbadano pięćdziesięcioelementowe próbki ludności w różnym wieku i zarejestrowano liczbę niewidomych w każdej próbce:

Wiek	20	35	45	55	70
Liczba niewidomych	6	17	26	37	44

Przypuszcza się, że prawdopodobieństwo ślepoty w wieku x, oznaczone przez $P(x)$, wyraża się wzorem

$$P(x) = \left\{1 + e^{-(\alpha+\beta x)}\right\}^{-1}.$$

Sporządzić odpowiedni wykres wyników i skomentować sensowność takiej hipotezy. Oszacować α i β na podstawie tego wykresu i następnie obliczyć wartość $ENW[\alpha, \beta]$. Oszacować również wiek, w którym bardziej prawdopodobne jest oślepnięcie niż pozostanie w zdrowiu.

19.15. Niech T oznacza liczbę pełnych okresów przeżytych przez pacjenta po pewnej operacji. Załóżmy, że T jest zmienną losową o rozkładzie $Ge\,(\theta)$, przy czym $\theta \in (0, 1)$ jest nieznanym parametrem. Obserwujemy losową grupę n niezależnych pacjentów, przy czym dla tych pacjentów, dla których $T \le t$ (t jest ustalone), znamy T dokładnie, a jeżeli pacjent żyje co najmniej $t + 1$ okresów, to jego czas życia jest nieznany, zatem dla każdego z pozostałych pacjentów wiemy tylko, że $T \ge t + 1$. Estymujemy θ na podstawie tych obserwacji. Wyznaczyć $ENW[\theta]$ wiedząc, że dla m pacjentów zaobserwowano dokładną liczbę przeżytych okresów.

19.16. Przyjmujemy, że liczby wypadków $N_1, \ldots, N_k$ zgłoszonych w kolejnych k latach są niezależnymi zmiennymi losowymi. Zakładamy, że zmienna N_i ma rozkład $Po\,(\lambda m_i)$, gdzie m_i jest znaną liczbą samochodów ubezpieczonych w i-tym roku, zaś λ nieznanym parametrem. Wyznaczyć $ENW[\lambda]$.

19.17. Zmienna losowa N ma rozkład $Po\,(\lambda)$ z parametrem λ, który chcemy oszacować. Niestety możemy obserwować jedynie zmienną losową M, która przyjmuje wartość zero, jeśli N równa się zero, a wartość jeden, jeśli N jest większa od zera. Średnią arytmetyczną z próbki niezależnych obserwacji zmiennej M oznaczmy przez $\bar{m}$. Wyznaczyć $ENW[\lambda]$.

19.18. Pobieramy próbę o rozkładzie $Po\,(\lambda)$. Niestety nasz sposób obserwacji uniemożliwia odnotowanie realizacji o wartości zero. Pobieranie próby kończymy w momencie, gdy liczebność odnotowanych realizacji wynosi T. Tak więc każda z naszych kolejnych odnotowanych realizacji $k_1, \ldots, k_T$ wynosi co najmniej 1, i nic nie wiemy o tym, ile w międzyczasie pojawiło się (i umknęło z naszego pola widzenia) obserwacji zerowych. Wyznaczyć $ENW[\lambda]$.

19.19. W urnie jest r czarnych kul. O liczbie r wiemy tylko tyle, że jest większa od zera. Powtarzamy trzy razy następujące czynności: losujemy jedną kulę z urny i odkładamy ją na bok (nie zwracamy), a następnie wrzucamy do urny jedną kulę białą. Wynikiem doświadczenia jest sekwencja trzech liter – C lub B na przykład CBB oznacza, iż wylosowaliśmy po kolei kulę czarną, potem białą i znowu białą. Wyznaczyć wartość estymatora największej wiarogodności nieznanej liczby r, gdy zaobserwowano ciąg CBC.

19.20. Przeprowadzamy wśród wylosowanych osób ankietę na delikatny temat. Ankietowana osoba rzuca kostką do gry i w zależności od wyniku rzutu kostką (wyniku tego nie zna ankieter) podaje odpowiednio zakodowaną odpowiedź na pytanie: *Czy zdarzyło się Panu/Pani w roku 1999 dać łapówkę w klasycznej formie pieniężnej, przekraczającą kwotę* 100 *zł?* Przyjmijmy, iż interesująca nas cecha X przyjmuje wartość 1, jeśli odpowiedź brzmi „TAK" i 0, jeśli odpowiedź brzmi „NIE". Pierwszych 100 osób udziela odpowiedzi $Z_1, \ldots, Z_{100}$ zgodnie z regułą: jeśli wynik rzutu kostką, to liczba oczek równa 1, 2, 3 lub 4, to $Z_i = X_i$, natomiast jeśli wynik rzutu kostką, to liczba oczek równa 5 lub 6, to $Z_i = 1 - X_i$. Następnych 100 osób udziela odpowiedzi $Z_{101}, \ldots, Z_{200}$ zgodnie z regułą: jeśli wynik rzutu kostką, to liczba oczek równa 1 lub 2, to $Z_i = X_i$, natomiast jeśli wynik rzutu kostką to liczba oczek równa 3, 4, 5 lub 6, to $Z_i = 1 - X_i$. Dla uproszczenia zakładamy, że dwieście ankietowanych osób to próba z (hipotetycznej) populacji o nieskończonej liczebności, a podział na podpróby jest także całkowicie losowy. Interesujący nas parametr tej populacji to oczywiście $q_X = P(X = 1)$. W wyniku przeprowadzonej ankiety dysponujemy średnimi z podprób: $\bar{Z}_1 = \frac{1}{100}\sum_{i=1}^{100} Z_i$ i $\bar{Z}_2 = \frac{1}{100}\sum_{i=101}^{200} Z_i$. $ENW[q_X]$ ma postać $\hat{q}_X = a_0 + a_1\hat{Z}_1 + a_2\hat{Z}_2$. Dla jakich a_0, a_1, a_2 estymator ten jest nieobciążony?

19.21. Wektor losowy (X, Y) ma łączny rozkład prawdopodobieństwa dany następującą tabelką:

	$Y = y_1$	$Y = y_2$
$X = x_1$	$\varepsilon(1-\theta)$	$\varepsilon\theta$
$X = x_2$	$(1-\varepsilon)\theta$	$(1-\varepsilon)(1-\theta)$

gdzie $\theta \in (0, 1)$ jest nieznanym parametrem, natomiast $\varepsilon \in (0, 1)$ jest znane. Na podstawie n elementowej próby z tego rozkładu wyznaczono $ENW[\theta]$. Obliczyć wariancję $D^2\hat{\theta}$ estymatora.

19.22. W pewnej populacji prawdopodobieństwo tego, że osobnik przeżyje rok jest równe $(1 - \theta)$. Jeżeli osobnik przeżył rok, to (warunkowe) prawdopodobieństwo tego, że przeżyje następny rok też jest równe $(1 - \theta)$. W próbce liczącej n osobników z tej populacji zanotowano n_0 przypadków, kiedy osobnik nie przeżył roku, n_1 przypadków, kiedy osobnik przeżył rok, ale nie przeżył drugiego oraz n_2 przypadków, kiedy osobnik przeżył dwa lata. Wyznaczyć $ENW[\theta]$.

19.23. Zakładamy, że każda pojedyncza szkoda, niezależnie od pozostałych, z prawdopodobieństwem θ jest likwidowana w roku, w którym została zgłoszona, w drugim roku po zgłoszeniu – z prawdopodobieństwem $\theta(1 - \theta)$, w trzecim roku lub później – z prawdopodobieństwem $(1-\theta)^2$. Dane, którymi dysponujemy dotyczą szkód. Wiemy, że spośród nich n_1 zostało zlikwidowanych w roku, w którym zostały zgłoszone, n_2 zostało zlikwidowanych w drugim roku po zgłoszeniu oraz n_3 zostało zlikwidowanych w trzecim roku lub później, gdzie $n_1 + n_2 + n_3 = n$. Wyznaczyć $ENW[\theta]$.

19.24. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(-\theta, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Niech $\hat{\theta}$ oznacza $ENW[\theta]$. Niech $\gamma \in (0,1)$. Wyznaczyć takie a, że

$$P_\theta(\hat{\theta} < \theta < a\hat{\theta}) = \gamma.$$

19.25. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(0, \theta)$. Obserwujemy zmienne losowe $Y_i = \min\{X_i, M\}$, gdzie M jest daną liczbą. Wyznaczyć $ENW[\theta]$.

19.26. Pewna komórka zawiera granulki, które można traktować jako kulki o jednakowym, ale nieznanym promieniu r i o których zakłada się, że są rozmieszczone w komórce losowo. W celu oszacowania r obserwuje się pod mikroskopem pewien przekrój komórki. Obserwowany przekrój zawiera n kołowych przekrojów granulek o promieniach $X_1, X_2, \ldots, X_n$. Wyznaczyć na tej podstawie $ENW[r]$. Znaleźć rozkład tego estymatora dla dużych n.

19.27. Pewien produkt elektrotechniczny wytwarzany jest w bardzo wielu fabrykach. Wadliwość (frakcja produktów wadliwych) p jest różna w różnych fabrykach i ma (na przestrzeni fabryk) w przybliżeniu rozkład $Bet(\alpha, \beta)$ przy czym α i β nie są znane. Przypuśćmy, że wylosowano s fabryk i w każdej z nich zbadano n produktów. Niech m_i $(i = 1, 2, \ldots, n)$ oznacza liczbę produktów wadliwych zaobserwowaną w i–tej fabryce. Podać szczegółowo sposób wyznaczania $ENW[\alpha, \beta]$. Pokazać, że gdy $n = 1$, parametry α i β nie są identyfikowalne.

19.28. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E(\lambda)$. Nie obserwujemy dokładnych wartości zmiennych X_i, tylko wartości $Z_i = \lfloor X_i \rfloor + 1$. Wyznaczyć $ENW[\lambda]$ oparty na obserwacjach $Z_1, \ldots, Z_n$.

19.29. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $E(\theta_1)$ oraz niech $Y_1, Y_2, \ldots, Y_n$ będzie niezależną próbą z rozkładu $E(\theta_2)$. Wyznaczyć $ENW[\theta_1, \theta_2]$. W jawny sposób wyznaczyć $ENW[\theta_1, \theta_2]$ przy dodatkowym warunku $\theta_1 = \theta_2$, i zilustrować na tym przykładzie ogólną teorię warunkowych estymatorów największej wiarogodności.

19.30. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E(\mu)$. Estymujemy μ na podstawie częściowej informacji o próbce: wiadomo, że m obserwacji miało wartości poniżej ustalonej wartości k i znana jest ich średnia arytmetyczna. Wyznaczyć $ENW[\mu]$.

19.31. $X_1, \ldots, X_n$ jest próbą z rozkładu $E(1; \theta)$. Znaleźć $ENW[\theta]$.

19.32. Zakładając, że $X_1, \ldots, X_{10}$ jest próbą z rozkładu $E(\mu)$ przeprowadzono estymację parametru μ metodą największej wiarogodności i otrzymano wartość $ENW[\mu]$ równą 50. Największa zaobserwowana w próbie wartość, tzn. wartość zmiennej losowej $\max\{X_1, \ldots, X_{10}\}$, wyniosła 100, a dziewięć pozostałych było ściśle mniejszych od 100. Okazało się jednak, że w istocie zaobserwowane przez nas wartości $X_1, \ldots, X_{10}$ stanowią próbę z uciętego rozkładu wykładniczego $X_i = \min\{Y_i, 100\}$, gdzie zmienne losowe Y_i pochodzą z $E(\mu)$. Obliczyć wartość $ENW[\mu]$ po uwzględnieniu modyfikacji założeń.

19.33. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E(\alpha;\theta)$. Wyznaczono $ENW[\theta,\alpha]$ w sytuacji, gdy oba parametry są nieznane ($\alpha > 0$). Znaleźć taką liczbę c, żeby $c\hat{\alpha}$ był nieobciążonym estymatorem parametru α.

19.34. Rozważmy losową liczbę zmiennych losowych $X_1, \ldots, X_N$. Zakładamy, że zmienne X_i są wzajemnie niezależne i niezależne od zmiennej losowej N. Wiemy, że każda ze zmiennych X_i ma rozkład $E(\alpha)$. Zmienna N ma rozkład $Po(\lambda)$. Zarówno $\lambda > 0$ jak i $\alpha > 0$ są nieznane. Obserwujemy tylko te spośród zmiennych $X_1, \ldots, X_N$, które przekraczają ustaloną wartość k. Nie wiemy, ile jest pozostałych zmiennych ani jakie są ich wartości. Wyznaczyć $ENW[\lambda,\alpha]$.

19.35. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi, przy czym zmienne X_i ($i = 1, \ldots, m < n$) mają rozkład $E(1/\lambda)$, a pozostałe zmienne mają rozkład $G(2, 1/\lambda)$. Wyznaczyć $ENW[\lambda]$.

19.36. Na podstawie próby $X_1, \ldots, X_n$ z rozkładu $G(2, 1/\theta)$ estymujemy parametr θ wykorzystując ENW $\hat{\theta}$. Wyznaczyć w przybliżeniu rozmiar próby n taki, że

$$P_\theta\left(\frac{|\hat{\theta}-\theta|}{\theta} \le 0.05\right) \approx 0.95.$$

Posłużyć się aproksymacją rozkładem normalnym.

19.37. Wiemy, że zmienna losowa X ma rozkład $N(\mu, 4)$. Na podstawie czteroelementowej próby estymujemy μ^2. Zaobserwowano $x_1 = 2$, $x_2 = 3.5$, $x_3 = 3$, $x_4 = 7$. Wyznaczyć różnicę między wartością estymatora największej wiarogodności a wartością estymatora nieobciążonego o minimalnej wariancji.

19.38. Niech $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ będzie ciągiem par wyników pewnych pomiarów, przy czym wszystkie $2n$ wyniki są niezależne i mają rozkład normalny o wariancji σ^2. Wartość oczekiwana X_i jest równa ξ_i, wartość oczekiwana Y_i jest równa η_i, a wszystkie pary (ξ_i, η_i), $i = 1, 2, \ldots, n$, leżą na okręgu o środku w punkcie (ξ, η) i o promieniu ϱ. Należy oszacować ξ, η i ϱ. Skonstruować $ENW[\xi, \eta, \varrho]$ i opracować szczegółowo algorytm obliczeniowy.

19.39. Przy wyznaczaniu oporu pewnego kryształu otrzymano ciąg niezależnych pomiarów (X_i, Y_i), $(i = 1, 2, \ldots, n)$, pomiarów natężenia prądu x i napięcia y. Pomiary te obciążone są błędami (ε_i, η_i) tak, że

$$X_i = \mu_i + \varepsilon_i, \qquad Y_i = \nu_i + \eta_i,$$

gdzie μ_i oraz ν_i sa prawdziwymi wartościami natężenia i napięcia prądu w i–tym pomiarze oraz $\nu_i = \alpha\mu_i$, gdzie α jest oporem kryształu. Zakładamy, że błędy są niezależne i ε_i ma rozkład $N\left(0, \sigma_1^2\right)$, a η_i ma rozkład $N\left(0, \lambda\sigma_1^2\right)$, przy czym λ jest znane. Pokazać, że $\hat{\alpha} = ENW[\alpha]$ jest rozwiązaniem równania

$$\hat{\alpha}^2 S_{xy} + \hat{\alpha}(\lambda S_{xx} - S_{yy}) - \lambda S_{xy} = 0,$$

gdzie

$$S_{xy} = \frac{1}{n}\sum X_iY_i \qquad S_{xx} = \frac{1}{n}\sum X_i^2 \qquad S_{yy} = \frac{1}{n}\sum Y_i^2.$$

Pokazać, że gdy ciąg $\sum \mu_i^2/n$ jest zbieżny dla $n \to \infty$, wtedy $\hat{\alpha}$ jest estymatorem zgodnym. Pokazać, że metoda największej wiarogodności nie jest zadowalająca, gdy λ nie jest znane. Wyjaśnić, dlaczego w tym przypadku nie można stosować standardowych twierdzeń o estymatorach największej wiarogodności.

19.40. Zakładamy, że $X_1,\ldots,X_n$, $Y_1,\ldots,Y_n$ są niezależnymi zmiennymi losowymi o rozkładach normalnych, przy czym $EX_i = EY_i = \mu$, $D^2X_i = \sigma^2$, $D^2Y_i = 4\sigma^2$ dla $i = 1,\ldots,n$. Parametry μ i σ są nieznane. Niech $\hat{\sigma}^2$ będzie $ENW[\sigma^2]$ w tym modelu. Wyznaczyć stałą a, tak aby $\tilde{\sigma}^2 = a\hat{\sigma}^2$ był estymatorem nieobciążonym parametru σ^2.

19.41. Niech $X_1,\ldots,X_n$ będzie próbą z rozkładu $N\left(\mu,\sigma^2\right)$, gdzie oba parametry są nieznane. Estymując parametr μ^2 wyznaczono dwa estymatory: $T_1 = ENW[\mu^2]$ oraz $T_2 = ENMW[\mu^2]$. Wyznaczyć różnicę ryzyk estymatorów T_1 i T_2 przy kwadratowej funkcji straty.

19.42. Niech $X_1,\ldots,X_{m+n}$ będą niezależnymi zmiennymi losowymi. Zmienne losowe X_i, $i = 1,\ldots,m$, mają rozkład $Wei\,(1/2,\theta)$, a X_i, $i = m+1,\ldots,m+n$, mają rozkład $Wei\,(1/2,\theta/2)$, gdzie $\theta > 0$ jest nieznanym parametrem. Dla przypadku $m = n = 5$, obliczyć błąd średniokwadratowy estymatora $ENW[\theta]$ wyznaczonego na podstawie zmiennych losowych $X_1,\ldots,X_{m+n}$.

19.43. Zmienna losowa N ma rozkład $Po\,(\lambda)$. Rozważamy losową liczbę zmiennych losowych $X_1,\ldots,X_N$, przy czym zmienne losowe $X_1,\ldots,X_N$ są niezależne wzajemnie i niezależne od zmiennej losowej N. Każda ze zmiennych losowych ma rozkład $Wei\,(2,\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Obserwujemy tylko te spośród zmiennych $X_1,\ldots,X_N$, które są większe od 10. Nie wiemy ile jest pozostałych zmiennych ani jakie są ich wartości. Przypuśćmy, że zaobserwowaliśmy cztery wartości większe od 10 i suma ich kwadratów jest równa 1200. Na podstawie tych danych wyznaczyć wartości estymatorów największej wiarogodności parametrów θ i λ.

19.44. Zakładając, że obserwacje $X_1,\ldots,X_{10}$ stanowią próbę z rozkładu $Par\,(3,\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem, wyznaczono wartość $ENW[\theta]$), która wyniosła 2. W próbie były dwie obserwacje o wartości sześć, a pozostałe osiem obserwacji miało wartości mniejsze od 6. Okazało się, że w rzeczywistości zaobserwowane wartości stanowiły próbę z uciętego rozkładu Pareto, czyli były realizacjami zmiennych losowych $X_i = \min\{Y_i, 6\}$, gdzie Y_i, $i = 1,\ldots,10$, są niezależnymi zmiennymi losowymi z rozkładu o gęstości f_θ. Wyznaczyć wartość estymatora największej wiarogodności parametru θ po uwzględnieniu modyfikacji założeń.

19.45. $X_1,\ldots,X_n$ jest próbą z rozkładu o dystrybuancie

$$F_\alpha(x) = \frac{1}{(1+e^{-x})^\alpha}, \qquad x \in R,\ \alpha > 0.$$

Wyznaczyć $ENW[\alpha]$.

19.46. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu o gęstości

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} x^{\frac{1}{\theta}-1}, & \text{dla } x \in (0,1), \\ 0, & \text{poza tym.} \end{cases}$$

Niech $\hat{\theta}$ będzie estymatorem największej wiarogodności nieznanego parametru $\theta > 0$. Obliczyć wariancję oraz błąd średniokwadratowy tego estymatora.

19.47. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Pow\,(\theta)$, a $Y_1, \ldots, Y_m$ będzie próbą z rozkładu $Pow\,(2\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Wszystkie zmienne losowe są niezależne. Dobrać stałą a tak, aby

$$P_\theta\left(\frac{T}{\theta} > a\right) = 0.9,$$

wiedząc, że T jest estymatorem największej wiarogodności parametru θ otrzymanym na podstawie zmiennych losowych $X_1, \ldots, X_n$ i $Y_1, \ldots, Y_m$.

19.48. Zmienne losowe $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ są niezależne o tym samym rozkładzie z gęstością

$$f_\theta(x) = \begin{cases} \frac{\theta}{(1+x)^{\theta+1}}, & \text{dla } x > 0, \\ 0, & \text{poza tym,} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Wyznaczono $\hat{\theta}_1 = ENW[\theta]$ na podstawie próby $X_1, \ldots, X_n$ oraz $\hat{\theta}_2 = ENW[\theta]$ na podstawie próby $Y_1, \ldots, Y_m$. Wyznaczyć stałe a i b tak, aby

$$P_\theta\left(\frac{\hat{\theta}_1}{\hat{\theta}_2} < a\right) = P_\theta\left(\frac{\hat{\theta}_1}{\hat{\theta}_2} > b\right) = 0.05.$$

19.49. Niech $X_1, \ldots, X_n, \ldots$ będzie próbą z rozkładu $Wei\,(4, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Niech T_n oznacza ENW funkcji $g(\theta) = P_\theta(X_1 > 1) = e^{-\theta}$ wyznaczony w oparciu o próbę $X_1, \ldots, X_n$. Niech $\theta = 2$. Udowodnić, że

$$\lim_{n\to\infty} P\{|T_n - e^{-2}|\sqrt{n} > 2e^{-2}\} = 0.32.$$

19.50. Pewne radioaktywne ciało emituje losowo cząstki z intensywnością, która z upływem czasu maleje i po czasie t wynosi $\lambda e^{-\kappa t}$. Zaobserwowano, że n pierwszych cząstek zostało wyemitowanych w chwilach $t_1, t_2, \ldots, t_n$. Napisać równania dla estymatorów największej wiarogodności $\hat{\lambda}$ i $\hat{\kappa}$ i pokazać, że $\hat{\kappa}$ spełnia równanie

$$\frac{\hat{\kappa} t_n}{e^{\hat{\kappa} t_n} - 1} = 1 - \hat{\kappa}\bar{t},$$

gdzie $\bar{t} = \frac{1}{n}\sum_{i=1}^n t_i$. Podać proste oszacowanie dla $\hat{\kappa}$, gdy $\hat{\kappa} t_n$ jest małe.

20. Metoda najmniejszych kwadratów

Estymator najmniejszych kwadratów.

Obserwujemy zmienne losowe $Y_1, \ldots, Y_n$ takie, że $EY_i = g_i(\theta)$, dla $i = 1, \ldots, n$, gdzie $g_i : \Theta \to \mathbb{R}^1$ są znanymi funkcjami rzeczywistymi. Zakładamy, że $\Theta \subset \mathbb{R}^k$. Niech

$$S(\theta) = \sum_{i=1}^{n} (Y_i - g_i(\theta))^2$$

Definicja 20.1 **Estymatorem najmniejszych kwadratów** parametru θ nazywamy wielkość $\hat{\theta}$ minimalizującą $S(\theta)$.

Estymator najmniejszych kwadratów oznaczamy $EMNK[\theta]$

Definicja 20.2 Wielkość $S(\hat{\theta}) = \sum_{i=1}^{n} (Y_i - g_i(\hat{\theta}))^2$ nazywamy **resztową sumą kwadratów**.

Model liniowy.

Definicja 20.3 **Modelem liniowym** nazywamy model statystyczny, w którym obserwacje $Y_1, \ldots, Y_n$ mają postać

$$Y_i = \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i, \quad i = 1, \ldots, n$$

gdzie x_{ji} są ustalonymi liczbami, β_j są nieznanymi parametrami modelu, ε_i są niezależnymi „błędami losowymi” takimi, że $E\varepsilon_i = 0$ oraz $D^2\varepsilon_i = \sigma^2$.

W modelach liniowych stosowany jest zapis macierzowy:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

gdzie

$$\boldsymbol{Y}^T = (Y_1, \ldots, Y_n), \ \boldsymbol{\beta}^T = (\beta_0, \beta_1, \ldots \beta_p), \ \boldsymbol{\varepsilon}^T = (\varepsilon_1, \ldots, \varepsilon_n)$$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{p1} \\ x_{12} & \cdots & x_{p2} \\ \vdots & \cdots & \vdots \\ x_{1n} & \cdots & x_{pn} \end{bmatrix}$$

Zakładamy, że macierz $\boldsymbol{X}$ jest macierzą pełnego rzędu.

W modelu liniowym ($\|\cdot\|$ oznacza długość wektora):

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}\beta_j x_{ji}\right)^2 = (\boldsymbol{Y}-\boldsymbol{X\beta})^T(\boldsymbol{Y}-\boldsymbol{X\beta}) = \|\boldsymbol{Y}-\boldsymbol{X\beta}\|^2$$

Ponieważ

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y} - \boldsymbol{X}^T\boldsymbol{X\beta},$$

więc $EMNK[\boldsymbol{\beta}]$ jest rozwiązaniem równania

$$\boldsymbol{X}^T\boldsymbol{X\beta} = \boldsymbol{X}^T\boldsymbol{Y}.$$

Ten układ równań nazywany jest **układem równań normalnych**. Ponieważ założyliśmy, że macierz $\boldsymbol{X}$ jest pełnego rzędu, więc

$$EMNK[\boldsymbol{\beta}] = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \stackrel{\text{ozn}}{=\!=} \hat{\boldsymbol{\beta}}.$$

Twierdzenie 20.1. *Jeżeli $\boldsymbol{c} \in \mathbb{R}^p$, to $EMNK[\boldsymbol{c}^T\boldsymbol{\beta}] = \boldsymbol{c}^T\hat{\boldsymbol{\beta}}$.*

Twierdzenie 20.2. *$EMNK[\boldsymbol{\beta}]$ jest estymatorem nieobciążonym o macierzy kowariancji $\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}$*

Definicja 20.4 Funkcję parametryczną $\boldsymbol{c}^T\boldsymbol{\beta}$ nazywamy **estymowalną**, jeżeli istnieje jej estymator nieobciążony postaci $\boldsymbol{b}^T\boldsymbol{Y}$.

Twierdzenie 20.3. *Funkcja parametryczna $\boldsymbol{c}^T\boldsymbol{\beta}$ jest estymowalna wtedy i tylko wtedy, gdy $\boldsymbol{c} \in \mathrm{Im}\boldsymbol{X}^T$.*

Twierdzenie 20.4. *(Gaussa–Markowa) Jeżeli błędy losowe $\varepsilon_1, \ldots, \varepsilon_n$ są nieskorelowanymi zmiennymi losowymi o zerowej wartości oczekiwanej i takiej samej wariancji, to dla każdej estymowalnej funkcji parametrycznej $\boldsymbol{c}^T\boldsymbol{\beta}$ i dla każdego nieobciążonego estymatora liniowego $\boldsymbol{b}^T\boldsymbol{Y}$ tej funkcji zachodzi*

$$D^2_{\boldsymbol{\beta}}(\boldsymbol{c}^T\hat{\boldsymbol{\beta}}) \le D^2_{\boldsymbol{\beta}}(\boldsymbol{b}^T\boldsymbol{Y}), \quad \forall \boldsymbol{\beta} \in \mathbb{R}^p$$

Wielkość

$$S(\hat{\boldsymbol{\beta}}) = \|\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 = \boldsymbol{Y}^T(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T)\boldsymbol{Y}$$

nazywamy **resztową sumą kwadratów**. Oznaczana jest przez RSS.

Twierdzenie 20.5. *Nieobciążony estymator wariancji ma postać*

$$\hat{\sigma}^2 = \frac{1}{(n - \mathrm{rz}\boldsymbol{X})} \boldsymbol{Y}^T (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T) \boldsymbol{Y}$$

Dodatkowe ograniczenia.

Rozważamy model liniowy:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

Załóżmy, że wektor $\boldsymbol{\beta}$ parametrów spełnia pewne dodatkowe ograniczenia liniowe $\boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$, gdzie $\boldsymbol{A}$ jest znaną $q \times p$ macierzą pełnego rzędu, a $\boldsymbol{c} \in \mathbb{R}^q$ jest znanym wektorem. Estymatory najmniejszych kwadratów uzyskuje się stosując technikę mnożników Lagrange'a. Funkcja Lagrange'a przyjmuje postać

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}) = (\boldsymbol{Y} - \boldsymbol{X})\boldsymbol{\beta}^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\gamma}^T (\boldsymbol{A}\boldsymbol{\beta} - \boldsymbol{c}),$$

gdzie $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_q]^T$ są mnożnikami Lagrange'a. Różniczkując funkcję Lagrange'a i przyrównując odpowiednie pochodne do zera otrzymujemy estymator wektora $\boldsymbol{\beta}$ przy dodatkowych ograniczeniach:

$$\hat{\boldsymbol{\beta}}_A = \hat{\boldsymbol{\beta}} - (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{A}^T [\boldsymbol{A}(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{A}^T]^{-1} (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c}).$$

Resztowa suma kwadratów w modelu z dodatkowymi ograniczeniami ma postać

$$\begin{aligned} \mathrm{RSS}_A &= \|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_A\|^2 = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_A)^T (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_A) \\ &= \mathrm{RSS} + (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T [\boldsymbol{A}(\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{A}^T]^{-1} (\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c}). \end{aligned}$$

Twierdzenie 20.6. *Nieobciążony estymator wariancji w modelu z ograniczeniami ma postać*

$$\hat{\sigma}_A^2 = \frac{RSS_A}{(n - \mathrm{rz}\boldsymbol{X} + \mathrm{rz}\boldsymbol{A})}$$

Rozkłady prawdopodobieństwa estymatorów.

Zakładamy, że

$$\boldsymbol{\varepsilon}^T = (\varepsilon_1, \ldots, \varepsilon_n) \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$$

Korzystając z własności wielowymiarowego rozkładu normalnego łatwo można udowodnić poniższe twierdzenie.

Twierdzenie 20.7. *Jeżeli* $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$ *oraz* $\mathrm{rz}\boldsymbol{X} = p$, *to*

1. $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1})$

2. $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{X}^T \boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \sigma^2 \chi^2_p$

3. $\hat{\boldsymbol{\beta}}$ *jest niezależne od* s^2

4. $(n-p)s^2 \sim \sigma^2 \chi^2_{n-p}$

Przykład. *Niech* $E(Y|X = x) = \beta_0 + \beta_1 x$. *Na podstawie* n *par* $(Y_1, x_1), \ldots, (Y_n, x_n)$ *obserwacji znaleźć estymatory najmniejszych kwadratów współczynników* β_0 *oraz* β_1 *funkcji regresji. Zakładając, że wariancja zmiennej losowej* Y *wynosi* σ^2 *i nie jest zależna od* x, *znaleźć estymator najmniejszych kwadratów wariancji* σ^2. *Przy założeniu normalności rozkładu zmiennej losowej* Y, *znaleźć rozkłady prawdopodobieństwa wyznaczonych estymatorów.*

Rozwiązanie.

Każdą obserwację Y_i *zapisujemy w postaci*

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, \ldots, n.$$

Zakładamy, że ε_i *są niezależnymi zmiennymi losowymi o tym samym rozkładzie o wartości oczekiwanej zero i wariancji* σ^2. *W notacji macierzowej mamy*

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

lub w skrócie $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, *gdzie* $\boldsymbol{X}^T = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \ldots & x_n \end{bmatrix}$ *oraz* $\boldsymbol{\beta}^T = [\beta_0, \beta_1]$. *Stosując metodę najmniejszych kwadratów otrzymujemy*

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}.$$

Jeżeli wśród obserwacji są przynajmniej dwa różne punkty x_i, *to macierz* $\boldsymbol{X}$ *jest pełnego rzędu, a co za tym idzie, istnieje odwrotność macierzy* $\boldsymbol{X}^T\boldsymbol{X}$. *Wówczas składowe wektora* $\boldsymbol{\beta}$ *są estymowalne, tzn. każdy z parametrów funkcji regresji jest estymowalny. Wykonując odpowiednie obliczenia otrzymujemy*

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{x},$$
$$\hat{\beta}_1 = \frac{\mathrm{cov}(x, Y)}{\mathrm{var} x}.$$

Znajdziemy oszacowanie wariancji σ^2. Jak wiemy, nieobciążoną oceną wariancji jest RSS/$(n-2)$. Dokładniej

$$s^2 = \frac{1}{n-2}\Big(\mathrm{var}Y - \hat{\beta}_1 \,\mathrm{cov}(x,Y)\Big)$$

jest poszukiwanym estymatorem wariancji.

Otrzymaliśmy klasyczne estymatory współczynników liniowej funkcji regresji.

Rozkłady odpowiednich estymatorów są następujące:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \sim N_2\left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \frac{\sigma^2}{\mathrm{var}x}\begin{bmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{bmatrix}\right).$$

Zatem

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\frac{\overline{x^2}}{\mathrm{var}x}\right), \quad \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\mathrm{var}x}\right).$$

Zmienna losowa $(n-2)s^2/\sigma^2$ ma rozkład $\chi^2\,(n-2)$.

□

Zadania do samodzielnego rozwiązania

20.1. Pokazać, że $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ jest rzutem ortogonalnym $\boldsymbol{Y}$ na $\{\boldsymbol{X}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^p\}$.

20.2. Niech $Y_1, \ldots, Y_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$. Wyznaczyć $EMNK[\mu]$ oraz $EMNK[\sigma^2]$. Wyznaczyć rozkłady uzyskanych estymatorów.

20.3. Niech $Y_{11}, \ldots, Y_{1n_1}$ będzie próbą z rozkładu $N\left(\mu_1, \sigma^2\right)$ oraz niech $Y_{21}, \ldots, Y_{2n_2}$ będzie próbą z rozkładu $N\left(\mu_2, \sigma^2\right)$. Wyznaczyć $EMNK[\mu_1 - \mu_2]$ oraz $EMNK[\sigma^2]$. Wyznaczyć rozkłady uzyskanych estymatorów.

20.4. Niech $Y_{i1}, \ldots, Y_{in_i}$ będzie próbą z rozkładu $N\left(\mu_i, \sigma^2\right)$, $i = 1, \ldots, k$. Niech $\boldsymbol{c} \in \mathbb{R}^k$ będzie danym wektorem. Wyznaczyć $EMNK[\boldsymbol{c}^T\boldsymbol{\mu}]$, gdzie $\boldsymbol{\mu}$ jest wektorem wartości oczekiwanych, tzn. $\boldsymbol{\mu}^T = (\mu_1, \ldots, \mu_k)$. Wyznaczyć $EMNK[\sigma^2]$ oraz rozkłady uzyskanych estymatorów.

20.5. Wyniki obserwacji $Y_{i1}, \ldots, Y_{in_i}$ mogą być przedstawione w postaci

$$Y_{ij} = \beta_{i0} + \beta_{i1}a_{ij} + \varepsilon_{ij} \qquad (i = 1, 2,\ j = 1, \ldots, n_i),$$

gdzie a_{ij} $(i = 1, 2,\ j = 1, \ldots, n_i)$ są znanymi wartościami zmiennej towarzyszącej, ε_{ij} $(i = 1, 2,\ j = 1, \ldots, n_i)$ są niezależnymi zmiennymi o rozkładach $N\left(0, \sigma^2\right)$, a parametry $\boldsymbol{\beta}^T = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})$ są nieznanymi parametrami. Wyznaczyć $EMNK[\boldsymbol{\beta}]$. Wyznaczyć $EMNK[\boldsymbol{\beta}]$ przy założeniu, że $\beta_{11} = \beta_{21}$. W obu przypadkach wyznaczyć $EMNK[\sigma^2]$. Wyznaczyć rozkłady prawdopodobieństwa uzyskanych estymatorów.

20.6. Wyniki obserwacji $Y_{i1}, \ldots, Y_{in_i}$ mogą być przedstawione w postaci

$$Y_{ij} = \beta_{i0} + \beta_{i1} a_{ij} + \varepsilon_{ij} \qquad (i = 1, 2,\ j = 1, \ldots, n_i),$$

gdzie $a_{11} \leq \cdots \leq a_{1n_1} \leq a_{21} \leq \cdots \leq a_{2n_2}$ są znanymi wartościami zmiennej towarzyszącej, ε_{ij} $(i = 1, 2,\ j = 1, \ldots, n_i)$ są niezależnymi zmiennymi o rozkładach $N\left(0, \sigma^2\right)$, a parametry $\boldsymbol{\beta}^T = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})$ są nieznanymi parametrami. Wyznaczyć $EMNK[\boldsymbol{\beta}]$. Wyznaczyć $EMNK[\boldsymbol{\beta}]$ przy założeniu, że $\beta_{10} + a\beta_{11} = \beta_{20} + a\beta_{21}$, gdzie a jest znaną liczbą taką, że $a_{1n_1} \leq a \leq a_{21}$. W obu przypadkach wyznaczyć $EMNK[\sigma^2]$. Wyznaczyć rozkłady prawdopodobieństwa uzyskanych estymatorów.

20.7. Wyniki obserwacji $Y_1, \ldots, Y_n$ mogą być przedstawione w postaci

$$Y_i = \beta_0 + \beta_1 a_i + \beta_2 a_i^2 + \varepsilon_i \qquad (i = 1, \ldots, n),$$

gdzie $a_1, \ldots, a_n$ są znanymi wartościami zmiennej towarzyszącej, oraz $\varepsilon_1, \ldots, \varepsilon_n$ są nieskorelowanymi zmiennymi losowymi o zerowej wartości oczekiwanej i jednakowej wariancji. Pokazać, że parametr $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2)$ jest estymowalny wtedy i tylko wtedy, gdy wśród $a_1, \ldots, a_n$ znajdują się co najmniej trzy różne wartości. Wyznaczyć $EMNK[\boldsymbol{\beta}]$.

20.8. Model $Y_i = \beta_0 + \beta_1 a_i + \varepsilon_i$ $(i = 1, \ldots, n)$ może być zapisany w postaci

$$Y_i = \beta_0 + \beta_1 \overline{a} + \beta_1 (a_i - \overline{a}) + \varepsilon_i = \alpha + \beta_1 (a_i - \overline{a}) + \varepsilon_i.$$

Pokazać, że wybór parametrów α i β_1 zamiast β_0 i β_1 ułatwia rachunki związane z wyznaczeniem estymatorów najmniejszych kwadratów. Sprawdzić, że w ogólnym przypadku model

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

może być zawsze, po odpowiedniej zmianie parametrów, przedstawiony w postaci

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

w której $\boldsymbol{B}$ jest macierzą o ortogonalnych kolumnach; estymator parametru $\boldsymbol{\gamma}$ otrzymuje się za pomocą łatwych rachunków.

20.9. Obserwacje Y_{ij} $(i = 1, \ldots, r, j = 1, \ldots, n)$ mają postać $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$, gdzie ε_{ij} są nieskorelowanymi błędami o wspólnej wariancji. Pokazać, że parametry $\mu, \tau_1, \tau_2, \ldots, \tau_r$ nie są estymowalne, ale że stają się estymowalne po wprowadzeniu warunku $\tau_1 + \tau_2 + \cdots + \tau_r = 0$. Pokazać, że estymatory najmniejszych kwadratów mają przy tym warunku postać

$$\hat{\mu} = X_{\cdot\cdot} = \frac{1}{rn} \sum_{i,j} X_{ij} \qquad \hat{\tau}_i = X_{i\cdot} - X_{\cdot\cdot} \qquad \text{gdzie} \qquad X_{i\cdot} = \frac{1}{n} \sum_j X_{ij}.$$

20.10. Niech

$$\begin{aligned} Y_1 &= \theta_1 + \theta_2 + \varepsilon_1, \\ Y_2 &= \theta_1 - 2\theta_2 + \varepsilon_2, \\ Y_3 &= 2\theta_1 - \theta_2 + \varepsilon_3. \end{aligned}$$

Błędy losowe mają zerową wartość oczekiwaną. Znaleźć $EMNK[\theta_1]$ oraz $EMNK[\theta_2]$. Wyznaczyć estymatory najmniejszych kwadratów parametrów następującego modelu rozszerzonego:

$$\begin{aligned} Y_1 &= \theta_1 + \theta_2 + \theta_3 + \varepsilon_1, \\ Y_2 &= \theta_1 - 2\theta_2 + \theta_3 + \varepsilon_2, \\ Y_3 &= 2\theta_1 - \theta_2 + \theta_3 + \varepsilon_3. \end{aligned}$$

20.11. Niech Y_1, Y_2, Y_3, Y_4 będą wynikami lotniczych pomiarów kątów $\theta_1, \theta_2, \theta_3, \theta_4$ pewnego czworokąta na powierzchni ziemi. Założyć, że obserwacje obciążone są błędami, które są niezależne, mają wartość oczekiwaną zero i taką samą wariancję σ^2. Przy tych założeniach wyznaczyć estymatory najmniejszych kwadratów wielkości θ. Wyznaczyć nieobciążony estymator wariancji σ^2.

Przypuśćmy, że wiadomo, iż dany czworokąt jest równoległobokiem takim, że $\theta_1 = \theta_3$ oraz $\theta_2 = \theta_4$. Jaką postać mają wtedy estymatory najmniejszych kwadratów kątów i jak można oszacować wariancję σ^2.

20.12. Pewne trzy substancje mogą być ważone tylko w opakowaniu. Chcemy oszacować jednostkowy ciężar tych substancji mając do dyspozycji wagę sprężynową. Wykonano cztery pomiary według następującego schematu:

Nr	Ważenie	Wynik
1	substancja 1 w opakowaniu	Y_1
2	substancja 2 w opakowaniu	Y_2
3	substancja 3 w opakowaniu	Y_3
4	opakowanie	Y_4

Napisać model tego doświadczenia. Wyznaczyć estymatory najmniejszych kwadratów jednostkowych ciężarów poszczególnych substancji oraz macierz kowariancji tych estymatorów.

20.13. Pewien produkt chemiczny można wytwarzać bez użycia katalizatora, ale przypuszcza się, że w obecności katalizatorów wydajność procesu będzie większa. Dla zbadania tego zagadnienia przeprowadzono pięć doświadczeń według następującego schematu:

Doświadczenie	Warunki	Wydajność
1	bez katalizatora	Y_1
2	katalizator A w ilości a_1	Y_2
3	katalizator A w ilości $2a_1$	Y_3
4	katalizator B w ilości a_2	Y_4
5	katalizator B w ilości $2a_2$	Y_5

Zakładając liniową regresję między wydajnością procesu i ilością każdego z katalizatorów skonstruować estymatory najmniejszych kwadratów „poziomu bezkatalizatorowego” i obu współczynników regresji. Wyznaczyć, przy zwykłych założeniach o błędzie, macierz kowariancji tych estymatorów. Wywnioskować, że przy danym $a_1 + a_2$ estymator różnicy współczynników regresji ma najmniejszą wariancję, gdy $a_1 = a_2$.

20.14. Pewien deterministyczny proces $y_0, y_1, \ldots, y_n$ przebiega tak, że

$$y_{i+1} = ay_i \qquad (i = 0, 1, \ldots, n-1),$$

przy czym a jest znaną stałą. Wielkości y_i nie mogą być obserwowane bezbłędnie, lecz ich obserwacje $X_0, X_1, \ldots, X_n$ mają postać

$$X_i = y_i + \varepsilon_i \qquad (i = 0, 1, \ldots, n-1),$$

gdzie ε_i są nieskorelowanymi błędami o jednakowej wariancji. Wyznaczyć estymatory najmniejszych kwadratów $y_0, y_1, \ldots, y_n$. Jak można by szacować $y_0, y_1, \ldots, y_n$, gdyby stała a nie była znana? W każdym z powyższych przypadków skonstruować estymator wariancji błędów.

20.15. Rozważmy model regresji liniowej

$$Y_i = ax_i + \varepsilon_i, \quad i = 1, 2, 3, 4,$$

gdzie ε_i są niezależnymi zmiennymi losowymi o rozkładach $N\left(0, \sigma^2\right)$, natomiast x_i są nielosowymi punktami z przedziału $[0, 3]$. Wielkość a jest nieznanym współczynnikiem. Dla jakich x_1, x_2, x_3, x_4 wariancja estymatora $EMNK[a]$ jest najmniejsza?

20.16. Obserwujemy zmienną y_t oraz zmienne $[x_{t,1} \ldots, x_{t,K}]$, co w postaci macierzowej zapisujemy:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_T \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,K} \\ \vdots & \vdots & \vdots \\ x_{T,1} & \cdots & x_{T,K} \end{bmatrix}.$$

Zakładamy, że rząd macierzy $\boldsymbol{X}$ wynosi K, a rząd macierzy rozszerzonej $[\boldsymbol{y}|\boldsymbol{X}]$ wynosi $K+1$ oraz ilość obserwacji $T > K+1$. Niech $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_k]^T$ oznacza hipotetyczny wektor współczynników regresji liniowej oraz niech $\boldsymbol{b}$ oznacza $EMNK[\boldsymbol{\beta}]$. Niech $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$ będzie wektorem reszt. Pokazać, że suma reszt jest zero wtedy i tylko wtedy, gdy istnieje taka kombinacja liniowa kolumn macierzy $\boldsymbol{X}$, która równa jest wektorowi jedynek.

20.17. Zakładamy, iż oczekiwany roczny koszt obsługi grupy ubezpieczonych jest liniową funkcją liczebności grupy (wielkości nielosowej), co możemy sformalizować następująco: $EY = ax + b$, gdzie Y jest rocznym kosztem obsługi grupy (w złotówkach), x jest ilością ubezpieczonych w grupie, natomiast a, b są nieznanymi parametrami. Zakłada się ponadto, że wariancja zmiennej losowej Y nie zależy od x. Zanotowano roczny koszt obsługi dla czterech grup o różnych liczebnościach:

x	50	100	200	500
Y	2000	3000	7000	9000

Do estymacji parametrów a, b stosujemy estymator najlepszy wśród wszystkich estymatorów liniowych i równocześnie nieobciążonych. Jaka jest wyestymowana wartość kosztu stałego (parametru b)?

20.18. Zakładamy, że zależność czynnika Y od nielosowego czynnika x opisuje model regresji liniowej $Y_i = \beta x_i + \varepsilon_i$. Obserwujemy n elementową próbę, w której $x_i = i$ dla $i = 1, \ldots, n$. Zmienne losowe $Y_1, \ldots, Y_n$ są niezależne, a błędy losowe mają rozkłady $N\left(0, i\sigma^2\right)$, dla $i = 1, 2, \ldots, n$. Wyznaczono estymator $\hat{\beta}$ parametru β wykorzystując ważoną metodę najmniejszych kwadratów, to znaczy minimalizując sumę $\sum_{i=1}^{5} \frac{(Y_i - \beta x_i)^2}{D^2 \varepsilon_i}$. Wyznaczyć stałą z tak, by $\mathrm{P}(|\hat{\beta} - \beta| < z\sigma) = 0.95$.

20.19. Zakładamy, że zależność czynnika Y od czynnika x (nielosowego) opisuje model regresji liniowej $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. Obserwujemy $2k$ elementową próbę, w której $x_1 = \cdots = x_k = a$ i $x_{k+1} = \cdots = x_{2k} = b$ ($a < b$). Zmienne losowe $Y_1, \ldots, Y_{2k}$ są niezależne. Błędy losowe mają rozkłady $N\left(0, \sigma^2\right)$ dla $i = 1, \ldots, k$ oraz $N\left(0, \gamma^2\sigma^2\right)$ dla $i = k+1, \ldots, 2k$, γ jest znaną liczbą. Wyznaczono estymatory $\hat{\beta}_0$ i $\hat{\beta}_1$ parametrów β_0 i β_1 wykorzystując ważoną metodę najmniejszych kwadratów, to znaczy minimalizując sumę $\sum_{i=1}^{2k} \frac{(Y_i - \beta_0 - \beta_1 x_i)^2}{D^2 \varepsilon_i}$. Wyznaczyć stałe z_0 i z_1 tak, aby $P(|\hat{\beta}_0 - \beta_0| < z_0\sigma) = 0.95$ i $P(|\hat{\beta}_1 - \beta_1| < z_1\sigma) = 0.95$.

21. Estymacja przedziałowa

Niech $(\mathcal{X}, \mathcal{B}, \{P_\theta,\ \theta \in \Theta\})$ będzie modelem statystycznym wektora losowego X. Niech x będzie realizacją wektora X.

Definicja 21.1 Zbiór $S(x) \subset \Theta$ taki, że

$$P_\theta\{\theta \in S(X)\} \geq 1 - \alpha,\ \forall \theta \in \Theta$$

nazywamy **obszarem ufności** dla parametru θ na **poziomie ufności** $1 - \alpha$.

Definicja 21.2 Jeżeli obszar ufności $S(x)$ ma postać przedziału $(\underline{\theta}(x), \overline{\theta}(x))$ to nazywamy go **przedziałem ufności**.

Konstrukcja obszaru ufności

1. Wyznaczyć taki zbiór $Z(\theta) \subset \mathcal{X}$, że

$$P_\theta\{X \in Z(\theta)\} \geq 1 - \alpha,\ \forall \theta \in \Theta.$$

Wówczas dla danego $x \in \mathcal{X}$ zbiór

$$S(x) = \{\theta \in \Theta :\ x \in Z(\theta)\}$$

jest obszarem ufności na poziomie ufności $1 - \alpha$.

2. W przypadku, gdy zmienna losowa X ma rozkład o dystrybuancie F_θ, można przyjąć:

$$Z(\theta) = \{x \in \mathcal{X} :\ x \in (t_1(\theta), t_2(\theta)]\},$$

gdzie $t_1(\theta),\ t_2(\theta)$ spełniają warunek:

$$F_\theta(t_2(\theta)) - F_\theta(t_1(\theta)) \geq 1 - \alpha,\ \forall \theta \in \Theta.$$

W szczególnym przypadku, gdy dla każdego $x \in \mathcal{X}$ funkcja $h(\theta; x) = F_\theta(x)$ jest malejąca ze względu na θ, zbiór $S(x)$ przyjmuje postać:

$$S(x) = (\underline{\theta}(x), \overline{\theta}(x)),$$

gdzie $\underline{\theta}(x)$ jest rozwiązaniem nierówności $F_\theta(x) \leq a$, a $\overline{\theta}(x)$ nierówności $F_\theta(x) \geq b$, przy czym $a, b \geq 0$ oraz $a+b = 1-\alpha$. Najczęściej przyjmuje się $a = \alpha/2$. W przypadku, gdy $h(\theta; x)$ jest rosnąca konstrukcja jest analogiczna.

3. Konstruujemy taką funkcję $t(x, \theta)$, że (jeżeli $\Theta \subset \mathbb{R}$) jest ona monotoniczną funkcją θ oraz przy wartości parametru θ rozkład zmiennej losowej $t(X, \theta)$ nie zależy od θ.

Taka funkcja nazywa się **funkcją centralną**. Wówczas można wyznaczyć takie wartości t_1 t_2, że

$$P_\theta\{t_1 < t(X,\theta) < t_2\} \geq 1-\alpha.$$

Ze względu na monotoniczność funkcji centralnej, nietrudno jest wyznaczyć takie $\underline{\theta}(x), \overline{\theta}(x)$, że

$$t_1 < t(x,\theta) < t_2 \Leftrightarrow \underline{\theta}(x) < \theta < \overline{\theta}(x).$$

Dodatkowe kryteria

1. Najkrótszy przedział ufności. Jeżeli $\Theta \subset \mathbb{R}$, to wyznaczamy takie funkcje $\underline{\theta}(x)$ oraz $\overline{\theta}(x)$, że

$$\overline{\theta}(x) - \underline{\theta}(x) = \min!$$

2. Przedział ufności o zadanej długości. Jeżeli $\Theta \subset \mathbb{R}$, $d > 0$, to

$$\overline{\theta}(x) - \underline{\theta}(x) \leq d$$

Przykład. *Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami. Skonstruować przedział ufności dla μ.*

Rozwiązanie. *Do konstrukcji przedziału ufności dla μ skorzystamy z trzeciej metody pokazując, że*

$$\frac{\overline{X}-\mu}{\mathrm{var}X}\sqrt{n(n-1)}$$

jest funkcją centralną.

Model dla próby $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ ma postać

$$\left(\mathbb{R}^n, \left\{N_n\left(\mu\mathbf{1}_n, \sigma^2\boldsymbol{I}_n\right),\ \theta = (\mu, \sigma^2) \in \mathbb{R}\times\mathbb{R}_+\right\}\right)$$

Korzystając z własności wielowymiarowego rozkładu normalnego wiemy, że $\overline{X}$ ma rozkład $N\left(\mu, \sigma^2/n\right)$. A zatem,

$$\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \sim N(0,1).$$

Jak nietrudno zauważyć, $\mathrm{var}X$ jest formą kwadratową wektora $\boldsymbol{X}$:

$$\mathrm{var}X = \boldsymbol{X}^T\left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)\boldsymbol{X}$$

i korzystając ze związku rozkładu normalnego z rozkładem chi-kwadrat stwierdzamy, że $\mathrm{var}X/\sigma^2$ ma rozkład $\chi^2(n-1)$. Ponadto, $\overline{X}$ oraz $\mathrm{var}X$ są niezależnymi zmiennymi losowymi. Rozważmy funkcję

$$t(\boldsymbol{X},\mu) = \frac{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}}{\sqrt{\frac{\mathrm{var}X}{\sigma^2}/(n-1)}} = \frac{\overline{X}-\mu}{\sqrt{\mathrm{var}X}}\sqrt{n(n-1)}$$

Dla każdej obserwacji $\boldsymbol{x}$ wektora $\boldsymbol{X}$ funkcja $t(\boldsymbol{x}, \cdot)$ jest funkcją monotoniczną; dla każdego μ zmienna losowa $t(\cdot, \mu)$ ma rozkład $t\,(n-1)$. Czyli $t(\cdot, \cdot)$ jest funkcją centralną.

A zatem, można znaleźć takie dwie liczby t_1 oraz t_2, że (t oznacza zmienną losową o rozkładzie $t\,(n-1)$)

$$P\,(t_1 < t < t_2) = 1 - \alpha.$$

Rozwiązując względem μ nierówność

$$t_1 < \frac{\overline{X} - \mu}{\sqrt{\mathrm{var}X}} \sqrt{n(n-1)} < t_2$$

otrzymujemy przedział ufności dla μ:

$$\left[\overline{X} - t_2\sqrt{\frac{\mathrm{var}X}{n(n-1)}}, \overline{X} + t_1\sqrt{\frac{\mathrm{var}X}{n(n-1)}}\right].$$

Wybór liczb t_1 i t_2 jest arbitralny. Znajdziemy takie liczby, by uzyskany przedział był najkrótszy. Rozwiązujemy więc zadanie

$$\begin{cases} t_2 - t_1 = \min! \\ F(t_2) - F(t_1) = 1 - \alpha \end{cases}$$

Tutaj $F(\cdot)$ oznacza dystrybuantę rozkładu $t\,(n-1)$.

Stosując technikę mnożników Lagrange'a otrzymujemy układ równań

$$\begin{cases} 1 - \lambda f(t_2) = 0, \\ 1 - \lambda f(t_1) = 0, \\ F(t_2) - F(t_1) = 1 - \alpha, \end{cases}$$

gdzie $f(\cdot)$ oznacza funkcję gęstości rozkładu $t\,(n-1)$.

Jak nietrudno sprawdzić, liczby t_1 oraz t_2 muszą spełniać warunek $t_1 + t_2 = 0$. Z własności rozkładu t wynika, że liczba t_2 jest rozwiązaniem równania

$$F(t_2) = 1 - \frac{\alpha}{2},$$

czyli t_2 jest kwantylem rozkładu $t\,(n-1)$. W statystyce, taką wielkość nazywamy **dwustronną wartością krytyczną** *rozkładu $t\,(n-1)$ i oznaczamy $t(\alpha; n-1)$.*

Podsumowując, najkrótszy przedział ufności dla wartości oczekiwanej μ ma postać

$$\left[\overline{X} - t(\alpha; n-1)\sqrt{\frac{\mathrm{var}X}{n(n-1)}}, \overline{X} + t(\alpha; n-1)\sqrt{\frac{\mathrm{var}X}{n(n-1)}}\right].$$

□

Zadania do samodzielnego rozwiązania

21.1. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Bin\,(m, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0, 1)$. Wyznaczyć przedział ufności dla parametru θ. Skorzystać z drugiej metody oraz związku rozkładu dwumianowego z rozkładem Beta.

21.2. Niech $X_1, \dots, X_n$ będzie próbą z rozkładu $Ge(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć przedział ufności dla parametru θ. Skorzystać z drugiej metody oraz związku rozkładu geometrycznego z rozkładem Beta.

21.3. Niech $X_1, \dots, X_n$ będzie próbą z rozkładu $NB(r, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Wyznaczyć przedział ufności dla parametru θ. Skorzystać z drugiej metody oraz związku rozkładu dwumianowego z rozkładem Beta.

21.4. Niech $X_1, \dots, X_n$ będzie próbą z rozkładu $E(\lambda)$ z nieznanym parametrem $\lambda > 0$. Wyznaczyć przedział ufności dla parametru λ. Skorzystać z trzeciej metody pokazując, że $\sum_{i=1}^n X_i/\lambda$ jest funkcją centralną.

21.5. Niech $X_1, \dots, X_n$ będzie próbą z rozkładu $N(\mu, \sigma^2)$ ze znaną wariancją $\sigma^2 > 0$. Wyznaczyć przedział ufności dla parametru μ. Skorzystać z trzeciej metody pokazując, że $(\sum_{i=1}^n X_i - n\mu)/\sqrt{n\sigma^2}$ jest funkcją centralną.

21.6. Niech $X_1, \dots, X_n$ będzie próbą z rozkładu $U(0, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Znaleźć najmniejsze c takie, żeby przedział

$$[\max\{X_1, \dots, X_n\}, c\max\{X_1, \dots, X_n\}]$$

był przedziałem ufności dla θ na poziomie ufności $1 - \alpha$.

21.7. Rozważmy próbę $X_1, \dots, X_n$ z rozkładu $U(0, \theta)$ (z nieznanym $\theta > 0$). Niech $M = \max\{X_1, \dots, X_n\}$. Należy zbudować przedział ufności dla θ na poziomie 90%. Chcemy, żeby ten przedział był postaci $[aM, bM]$, gdzie liczby a i b są tak dobrane, żeby

$$P_\theta(\theta < aM) = P_\theta(\theta > bM) = 0.05.$$

Podać długość tego przedziału.

21.8. Losujemy n ($n \geq 3$) niezależnych realizacji zmiennej losowej o rozkładzie $U(0, \theta)$. Po uporządkowaniu zaobserwowanych wartości w ciąg rosnący $\{x_1, \dots, x_n\}$ tworzymy przedział $(2x_1, 2x_{n-1})$. Dobrać najmniejsze n, przy którym prawdopodobieństwo tego, że tak utworzony przedział pokrywa wartość parametru θ jest większe niż 0.9.

21.9. Niech $X_1, \dots, X_n$ będzie próbą z rozkładu $E(1/\theta)$. Parametr θ jest nieznany. Wiadomo, że ENW tego parametru jest $\hat{\theta} = n/S_n$, gdzie $S_n = X_1 + \cdots + X_n$. Należy zbudować przedział ufności dla parametru θ postaci

$$[\underline{\theta}, \overline{\theta}] = \left[\frac{\underline{a}}{S_n}, \frac{\overline{a}}{S_n}\right].$$

Żądamy, żeby ten przedział był symetryczny w tym sensie, że $P_\theta(\theta < \underline{\theta}) = P_\theta(\theta > \overline{\theta})$. Wyznaczyć stałe $\underline{a}$ i $\overline{a}$ tak, żeby otrzymać przedział na poziomie ufności 0.95.

21.10. Niech X oraz Y będą zmiennymi losowymi o rozkładach $E\left(1/\lambda\right)$ i $E\left(1/\mu\right)$ odpowiednio. Pokazać, że

$$C_{X,Y} = \{(\lambda,\mu) : \lambda X + \mu Y \leq a\}$$

jest obszarem ufności dla (λ,μ) na poziomie ufności $1-(1+a)e^{-a}$.

21.11. Niech X ma rozkład $N\left(\mu,\sigma^2\right)$ o nieznanych parametrach μ i σ. Pokazać, jak za pomocą funkcji $(X-\mu)/\sigma$ można skonstruować taki przedział ufności dla $\theta=(\mu,\sigma)$, który miałby kształt trójkąta.

21.12. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu,\sigma^2\right)$ o nieznanych μ i σ^2 i niech $S^2 = \frac{1}{n-1}\mathrm{var}X$. Pokazać, że S^2/σ^2 jest funkcją centralną za pomocą której można skonstruować przedział ufności dla wariancji σ^2.

21.13. Jeżeli $X_1, X_2, \ldots, X_n$ jest próbą z rozkładu $N\left(\mu,\sigma^2\right)$, to za pomocą funkcji $(\overline{X}-\mu)/S$ można skonstruować $100(1-\alpha)$-procentowy przedział ufności dla μ, powiedzmy $(\underline{\mu},\overline{\mu})$, oraz za pomocą funkcji S^2/σ^2 można skonstruować $100(1-\alpha)$-procentowy przedział ufności dla σ, powiedzmy $(\underline{\sigma},\overline{\sigma})$. W powyższych wzorach $S^2 = \mathrm{var}X/(n-1)$. Rozpatrzmy prostokątny obszar na płaszczyźnie (μ,σ):

$$\left\{(\mu,\sigma):\ \underline{\mu}\leq\mu\leq\overline{\mu},\ \underline{\sigma}\leq\sigma\leq\overline{\sigma}\right\}.$$

Obszar ten można przyjąć za zbiór ufności dla (μ,σ). Co można powiedzieć na temat poziomu ufności?

21.14. Niech X będzie nieobciążonym estymatorem o znanej wariancji σ_1^2 parametru ξ, niech Y będzie nieobciążonym estymatorem o znanej wariancji σ_2^2 parametru η i niech X i Y będą niezależne. Przypuśćmy, że dla każdej liczby λ zmienna losowa $X-\lambda Y$ ma rozkład normalny. Pokazać, jak można zbudować 95-procentowy przedział ufności dla ilorazu ξ/η.

Może się czasami wydarzyć, że przedział ten jest całą prostą liczbową. Sugeruje to, aby takie przypadki wyłączyć z rozważań przy obliczaniu prawdopodobieństwa p, że przedział ufności zawiera prawdziwą wartość. Pokazać, że po takiej modyfikacji prawdopodobieństwo p zależy od ξ/η.

21.15. $X_1,\ldots,X_{10}$ jest próbą z rozkładu $N\left(\mu,\sigma^2\right)$, gdzie μ i σ^2 są nieznanymi parametrami. Niech $[L,U]$ będzie przedziałem ufności dla parametru μ taki, że dla wszystkich μ i σ^2

$$P_{\mu,\sigma^2}(L>\mu) = P_{\mu,\sigma^2}(U<\mu) = 0.025.$$

Niech $(-\infty,W]$ będzie jednostronnym przedziałem ufności dla μ takim, że $P_{\mu,\sigma^2}(W<\mu)=0.01$ dla wszystkich μ i σ^2. Oba przedziały ufności zbudowane są w standardowy sposób w oparciu o $\overline{X}$ i $\mathrm{var}X$. Wiadomo, że $L=-0.262$ i $U=4.262$. Wyznaczyć W.

21.16. Dwie niezależne próby $X_1,\ldots,X_n$ i $Y_1,\ldots,Y_n$ pochodzą z tego samego rozkładu $N\left(\mu,\sigma^2\right)$. Jeden statystyk ma do dyspozycji pierwszą próbę, drugi zaś drugą. Obaj statystycy znają wariancję σ^2, żaden nie zna wartości oczekiwanej μ. Każdy z nich buduje na podstawie swojej próby przedział ufności dla μ na poziomie ufności $1-\alpha$. Wyznaczyć prawdopodobieństwo, że zbudowane przedziały okażą się rozłączne.

21.17. Na podstawie próby $X_1, \ldots, X_n$ z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami μ i σ^2 budujemy przedział ufności $(\underline{\sigma^2}, \overline{\sigma^2})$ dla wariancji na poziomie ufności 0.95. Metodę wybieramy możliwie prostą, korzystając na przykład z przybliżenia rozkładu $\chi^2(k)$ rozkładem $N(k, 2k)$. Względny błąd estymacji przedziałowej mierzymy za pomocą ilorazu $R = \frac{\overline{\sigma^2} - \underline{\sigma^2}}{2\sigma^2}$. Dla jakiego rozmiaru n próby $E(R) \approx 0.01$.

21.18. Jacek i Placek mają próbę $X_1, \ldots, X_n$ z rozkładu $N\left(\mu, \sigma^2\right)$. Obaj nie znają wartości oczekiwanej μ, ale Jacek zna wariancję σ^2, a Placek jej nie zna. Obaj budują w standardowy sposób przedziały ufności dla μ na poziomie ufności 0.95. Placek się chwali: „mam szansę 10%, że mój przedział ufności będzie przynajmniej x razy krótszy, niż Twój". Znaleźć x.

21.19. Niech $X_1, \ldots, X_{20}$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$, z nieznanymi parametrami μ i σ^2. Niech

$$\bar{X}_{10} = \frac{1}{10}\sum_{i=1}^{10} X_i, \;\; \bar{X}_{20} = \frac{1}{20}\sum_{i=1}^{20} X_i, \;\; S^2 = S_{10}^2 = \frac{1}{9}\sum_{i=1}^{10}(X_i - \bar{X}_{10})^2.$$

Należy skonstruować przedział $[\bar{X}_{10} - aS, \bar{X}_{10} + aS]$ taki, że $P_{\mu,\sigma^2}(\bar{X}_{20} \in [\bar{X}_{10} - aS, \bar{X}_{10} + aS]) = 0.95$. Wyznaczyć odpowiednią liczbę a.

21.20. Załóżmy, że $X_1, \ldots, X_9$ jest próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami μ i σ^2. Pan Ixiński miał podać przedział ufności dla μ na poziomie 0.95, ale nie znalazł tablic rozkładu $t(\cdot)$. Ponieważ miał tablice rozkładu $N(\cdot, \cdot)$ i $\chi^2(\cdot)$, więc poradził sobie tak:

1. obliczył w standardowy sposób jednostronny przedział ufności $[0, \overline{\sigma}^2]$ dla wariancji, na poziomie 0.95;

2. przyjął, że $\left[\overline{X} - \frac{1.96\overline{\sigma}}{\sqrt{9}}, \overline{X} + \frac{1.96\overline{\sigma}}{\sqrt{9}}\right]$ jest potrzebnym przedziałem dla wartości oczekiwanej, gdzie $\overline{\sigma}$ zostało wyznaczone w punkcie 1.

Obliczyć faktyczny poziom ufności takiego przedziału ufności dla μ.

21.21. Załóżmy, że $X_1, \ldots, X_6$ jest próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami. Zadanie polega na zbudowaniu przedziału ufności dla wariancji σ^2. Żądany poziom ufności jest równy $1 - \alpha = 0.99$. Rozpatrzmy dwie metody:

Metoda S jest standardowa: budujemy przedział postaci $[0, G_S]$, gdzie $G_S = \frac{5S^2}{c}$, (S^2 oznacza nieobciążony estymator wariancji, zaś c jest odpowiednim kwantylem rozkładu $\chi^2(\cdot)$);

Metoda N polega na podziale próby na dwie części. Podpróbę X_1, X_2, X_3 wykorzystujemy do zbudowania przedziału ufności $[0, G_{123}]$, zaś podpróbę X_4, X_5, X_6 do zbudowania przedziału $[0, G_{456}]$. Oba te przedziały obliczamy niezależnie w standardowy sposób przyjmując poziom ufności $1 - \sqrt{\alpha} = 0.90$. Ostatecznie naszym przedziałem ufności jest $[0, G_N]$, gdzie $G_N = \max\{G_{123}, G_{456}\}$.

Porównać średnie długości przedziałów otrzymanych obiema metodami.

21.22. Każda ze zmiennych losowych $X_1, \ldots, X_n$ ma rozkład $N(\mu, \sigma^2)$ z nieznaną wartością oczekiwaną μ i znaną wariancją σ^2. Założono, że zmienne są niezależne. W standardowy sposób zbudowano dla μ przedział ufności na poziomie ufności 0.95:

$$\left[\overline{X} - \frac{1.96\sigma}{\sqrt{n}}, \overline{X} + \frac{1.96\sigma}{\sqrt{n}}\right].$$

W rzeczywistości zmienne $X_1, \ldots, X_n$ mają łączny rozkład normalny, ale są skorelowane, $Corr(X_i, X_j) = 0.1$ dla wszystkich $i \neq j$. Obliczyć faktyczny poziom ufności.

21.23. Próbka $X_1, \ldots, X_n$ pochodzi z rozkładu $N(\mu, \sigma^2)$ z nieznaną wartością oczekiwaną μ i nieznaną wariancją σ^2. Na podstawie tej próby zbudowano dla μ w standardowy sposób przedział ufności na poziomie ufności $1 - \alpha = 0.95$

$$\left[\overline{X} - \frac{\sqrt{\mathrm{var} X} \cdot t(0.025; n-1)}{\sqrt{n(n-1)}}, \overline{X} + \frac{\sqrt{\mathrm{var} X} \cdot t(0.025; n-1)}{\sqrt{n(n-1)}}\right].$$

Niech X będzie zmienną losową pochodzącą z tego samego rozkładu, niezależną od $X_1, \ldots, X_n$. Obliczyć prawdopodobieństwo tego, że X należy do uprzednio wyznaczonego przedziału ufności:

$$P_{\mu,\sigma^2}\left(\overline{X} - \frac{\sqrt{\mathrm{var} X} \cdot t(0.025; n-1)}{\sqrt{n(n-1)}} \leq X \leq \overline{X} + \frac{\sqrt{\mathrm{var} X} \cdot t(0.025; n-1)}{\sqrt{n(n-1)}}\right).$$

21.24. Załóżmy, że $X_1, \ldots, X_n$ jest próbą z rozkładu $N(\mu, \sigma^2)$ oraz $Y_1, \ldots, Y_m$ jest próbą z rozkładu $N(\mu, \tau^2)$, μ jest nieznanym parametrem. Znaleźć takie liczby r i d, żeby przedział

$$\left[r\overline{X} + (1-r)\overline{Y} - d, r\overline{X} + (1-r)\overline{Y} + d\right]$$

był najkrótszym przedziałem ufności dla μ na poziomie ufności $1 - \alpha$.

21.25. Załóżmy, że $X_1, \ldots, X_n$ jest próbą z rozkładu normalnego. Zmienna X_i ma rozkład $N\left(\mu, \frac{1}{i}\right)$, dla $i = 1, \ldots, n$. Wartość oczekiwana μ jest nieznana. Należy zbudować przedział ufności dla μ na poziomie $1 - \alpha = 0.95$. Przedział ma być postaci $[\hat{\mu} - d, \hat{\mu} + d]$, gdzie $\hat{\mu}$ jest $ENW(\mu)$. Podać liczbę d taką, że

$$P_\mu(\hat{\mu} - d < \mu < \hat{\mu} + d) = 0.95.$$

21.26. Załóżmy, że $X_1, \ldots, X_n$ jest próbą z rozkładu $N(\mu, \sigma^2)$ o nieznanej wartości oczekiwanej i nieznanej wariancji. Niech X będzie zmienną losową z tego samego rozkładu, niezależną od próby. Zbudować „przedział ufności”

$$[L, U] = [L(X_1, \ldots, X_n), U(X_1, \ldots, X_n)]$$

taki, że

$$P_{\mu,\sigma^2}(L(X_1, \ldots, X_n) \leq X \leq U(X_1, \ldots, X_n)) = 0.95,$$

przy tym żądamy, żeby przedział był symetryczny, tzn. $\frac{1}{2}(L + U) = \overline{X}$.

21.27. Zakładamy, że $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi o rozkładach normalnych, przy czym $EX_i = \mu$ i $D^2X_i = \frac{\sigma^2}{w_i}$. Parametry μ i σ^2 są nieznane, a wagi w_i $(i = 1, \ldots, n)$ są znane. Zbudować przedział ufności $[\hat{\sigma}_1^2, \hat{\sigma}_2^2]$ dla σ^2 na poziomie ufności $1 - \alpha = 0.9$.

21.28. Niech $(X_1, Y_1), \ldots, (X_n, Y_n)$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie normalnym z następującymi parametrami: nieznaną wartością oczekiwaną $EX_i = EY_i = \mu$, wariancją $D^2X_i = \frac{1}{4}$, $D^2Y_i = 1$ i współczynnikiem korelacji $Corr(X_i, Y_i) = 0.5$. Osobno na podstawie prób $X_1, \ldots, X_n$ i $Y_1, \ldots, Y_n$ zbudowano dwa przedziały ufności dla wartości oczekiwanej μ, każdy na poziomie ufności 0.8. Obliczyć prawdopodobieństwo, że tak zbudowane przedziały okażą się rozłączne.

21.29. Zakładamy, że $X_1, \ldots, X_n$ są niezależnymi zmiennymi losowymi o rozkładach normalnych, przy czym $EX_i = \mu$ i $D^2X_i = \frac{\sigma^2}{i}$, gdzie parametry $\mu \in R$ i $\sigma^2 > 0$ są nieznane. Budujemy przedział ufności $[\hat{\sigma}_1^2, \hat{\sigma}_2^2]$ dla parametru σ^2 na poziomie ufności 0.9. Wyznaczyć $\hat{\sigma}_1^2$ i $\hat{\sigma}_2^2$ tak, by $P_{\mu,\sigma^2}(\hat{\sigma}_1^2 > \sigma^2) = P_{\mu,\sigma^2}(\hat{\sigma}_2^2 < \sigma^2) = 0.05$.

21.30. Dysponując pięcioma niezależnymi próbami o tej samej liczebności n, z tego samego rozkładu $N\left(\mu, \sigma^2\right)$ z nieznaną wartością oczekiwaną μ i znaną wariancją σ^2, zbudowano pięć standardowych przedziałów ufności dla parametru μ postaci

$$\left[\overline{X}_i - 1.2816\frac{\sigma}{\sqrt{n}}, \overline{X}_i + 1.2816\frac{\sigma}{\sqrt{n}}\right],$$

gdzie $\overline{X}_i$ jest średnią z obserwacji w i-tej próbce, $i = 1, 2, 3, 4, 5$, natomiast liczba 1.2816 jest kwantylem rzędu 0.9 rozkładu $N(0, 1)$. Zbudowano przedział ufności dla parametru μ postaci

$$\left[m - 1.2816\frac{\sigma}{\sqrt{n}}, m + 1.2816\frac{\sigma}{\sqrt{n}}\right],$$

gdzie $m = med\{X_1, X_2, X_3, X_4, X_5\}$. Wyznaczyć

$$P_\mu\left(m - 1.2816\frac{\sigma}{\sqrt{n}} \le \mu \le m + 1.2816\frac{\sigma}{\sqrt{n}}\right).$$

21.31. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Par(1, a_1)$, a $Y_1, \ldots, Y_m$ będzie próbą z rozkładu $Par(1, a_2)$, gdzie $a_1, a_2 > 0$ są nieznanymi parametrami. Wszystkie zmienne są niezależne. Na poziomie ufności $1-\alpha$ budujemy przedział ufności $[dT, cT]$ dla ilorazu parametrów $\frac{a_1}{a_2}$ na podstawie estymatora największej wiarogodności T tego ilorazu w ten sposób, że

$$P_{a_1,a_2}\left(cT < \frac{a_1}{a_2}\right) = P_{a_1,a_2}\left(dT > \frac{a_1}{a_2}\right) = \frac{\alpha}{2}.$$

Wyznaczyć długość przedziału ufności, gdy $\alpha = 0.1$, $m = 4$, $n = 5$.

21.32. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Wei\,(3, 1/\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. W oparciu o estymator największej wiarogodności $\hat{\theta}_n$ otrzymujemy przedział ufności dla parametru θ rozwiązując nierówność

$$\left| \frac{\hat{\theta}_n - \theta}{\sigma(\theta)} \sqrt{n} \right| \le z,$$

gdzie $\sigma^2(\theta)$ jest wariancją asymptotyczną statystyki $\hat{\theta}_n$ i liczba z spełnia

$$\lim_{n\to\infty} P_\theta \left(\left| \frac{\hat{\theta}_n - \theta}{\sigma(\theta)} \sqrt{n} \right| \le z \right) = 0.95.$$

Wyznaczyć postać przedziału ufności.

21.33. Niech $X_1, \ldots, X_n$ będzie próbą n o rozkładzie ciągłym i niech $X_{1:n} \le \cdots \le X_{n:n}$ oznacza uporządkowaną próbę. Budujemy przedział ufności dla mediany zmiennej X o postaci

$$U_n = (X_{2:n}, X_{n-1:n}).$$

Oznaczmy przez n^* najmniejszą z tych wartości n, dla których prawdopodobieństwo pokrycia mediany przez przedział U_n przekracza 0.95. Wyznaczyć n^*.

21.34. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Pow\,(\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Skonstruować przedział ufności $[\underline{\theta}, \overline{\theta}]$ dla parametru θ (na poziomie 0.90) tak, żeby $P_\theta(\underline{\theta} < \theta) = 0.05 = P_\theta(\overline{\theta} > \theta)$.

22. Weryfikacja hipotez statystycznych

Testy istotności.

Niech $(\mathcal{X}, \mathcal{B}, \{P_\theta,\ \theta \in \Theta\})$ będzie modelem statystycznym.

Definicja 22.1 **Hipotezą statystyczną** nazywamy podzbiór Θ_0 zbioru Θ.

Podzbiór Θ_0 nazywamy **hipotezą zerową**, zaś podzbiór $\Theta_1 = \Theta \setminus \Theta_0$ nazywamy **hipotezą alternatywną**.

Jeżeli zbiór Θ_0 jest jednoelementowy, to mówimy o hipotezie **prostej**, w przeciwnym przypadku mówimy o hipotezie **złożonej**.

Zapis klasyczny

$$H_0 : \theta \in \Theta_0 \qquad H_1 : \theta \in \Theta_1$$

Definicja 22.2 **Testem statystycznym** nazywamy procedurę statystyczną, w wyniku której podejmujemy jedną z dwóch decyzji: *odrzucić hipotezę zerową* H_0 lub *nie ma podstaw do odrzucenia hipotezy zerowej* H_0.

Test hipotezy H_0 utożsamiamy z funkcją $\phi : \mathcal{X} \to \{0, 1\}$ taką, że

$$\phi(x) = \begin{cases} 1 & \text{odrzucić } H_0 \\ 0 & \text{nie odrzucać } H_0 \end{cases}$$

Testem zrandomizowanym nazywamy funkcję $\phi : \mathcal{X} \to [0, 1]$ taką, że

$$\phi(x) \begin{cases} = 1 & \text{odrzucić } H_0 \\ \in (0,1) & \text{na podstawie niezależnego od } X \text{ mechanizmu losowego odrzucić } \\ & H_0 \text{ z prawdopodobieństwem } \phi(x) \\ = 0 & \text{nie odrzucać } H_0 \end{cases}$$

Definicja 22.3 **Obszarem krytycznym** testu nazywamy zbiór $\{x \in \mathcal{X} : \phi(x) = 1\}$.

Obszar krytyczny jest zbiorem tych obserwacji dla których odrzucana jest hipoteza zerowa.

Definicja 22.4 **Błędem I rodzaju** nazywamy błąd wnioskowania polegający na odrzuceniu hipotezy zerowej H_0, gdy w rzeczywistości jest ona prawdziwa.

Niech $\alpha \in (0, 1)$ będzie daną liczbą.

Definicja 22.5 Test ϕ jest na **poziomie istotności** α, jeżeli $E_\theta\phi(X) = P_\theta\{\phi(X) = 1\} \le \alpha,\ \forall\theta \in \Theta_0$.

Inaczej mówiąc, poziom istotności jest prawdopodobieństwem odrzucenia prawdziwej hipotezy zerowej.

Definicja 22.6 **Rozmiarem testu** nazywamy $\sup\limits_{\theta\in\Theta_0} E_\theta\phi(X) = \sup\limits_{\theta\in\Theta_0} P_\theta\{\phi(X) = 1\}$.

Zauważmy, że rozmiar testu jest największym prawdopodobieństwem odrzucenia hipotezy zerowej, gdy jest ona prawdziwa.

Definicja 22.7 **Błędem II rodzaju** nazywamy błąd polegający na nieodrzuceniu hipotezy zerowej H_0, gdy w rzeczywistości jest ona fałszywa.

Definicja 22.8 **Mocą testu** nazywamy $\Theta_1 \ni \theta \to E_\theta\phi(X) = P_\theta\{\phi(X) = 1\}$.

Moc testu jest prawdopodobieństwem odrzucenia hipotezy zerowej, gdy nie jest ona prawdziwa.

Twierdzenie 22.1. *(lemat Neymana–Pearsona)*
Niech P_{θ_0} oraz P_{θ_1} będą rozkładami prawdopodobieństwa o gęstościach f_0 i f_1. Niech $\alpha \in (0,1)$ będzie ustaloną liczbą

a. (Istnienie testu) *Istnieją takie stałe t i γ, że*

$$\phi(x) = \begin{cases} 1, & \text{gdy } f_1(x) > tf_0(x) \\ \gamma, & \text{gdy } f_1(x) = tf_0(x) \\ 0, & \text{gdy } f_1(x) < tf_0(x) \end{cases}$$

jest testem hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$ na poziomie istotności α, tzn.

$$(*) \qquad E_{\theta_0}\phi(X) = \alpha$$

b. (Dostateczność) *Jeżeli test ϕ spełnia warunek $(*)$ i dla pewnego t warunek*

$$(**) \qquad \phi(x) = \begin{cases} 1, & \text{gdy } f_1(x) > tf_0(x) \\ 0, & \text{gdy } f_1(x) < tf_0(x) \end{cases}$$

to ϕ jest testem najmocniejszym na poziomie istotności α

c. (Konieczność) *Jeżeli ϕ jest testem najmocniejszym na poziomie istotności α, to spełnia on warunek $(**)$*

Lemat Neymana–Pearsona podaje sposób konstrukcji najmocniejszego testu w problemie weryfikacji dwóch hipotez prostych.

Przykład. *Niech $a, b > 1$ będą dwiema ustalonymi liczbami. Rozważmy model statystyczny*

$$\{(0,1), \{U(0,1), Bet(a,b)\}\}$$

Korzystając z lematu Neymana–Pearsona skonstruować test dla problemu weryfikacji hipotezy $H_0 : U(0,1)$ przeciwko $H_1 : Bet(a,b)$.

Rozwiązanie.

Niech f_0 oznacza gęstość rozkładu jednostajnego $U(0,1)$, zaś f_1 gęstość rozkładu beta $B(a,b)$. Mamy

$$f_0(x) = 1 \cdot \mathbf{1}_{(0,1)}(x)$$
$$f_1(x) \propto x^{a-1}(1-x)^{b-1}\mathbf{1}_{(0,1)}(x)$$

Zgodnie z lematem Neymana–Pearsona, test $\phi(x)$ ma postać

$$\phi(x) = 1 \Leftrightarrow \frac{f_1(x)}{f_0(x)} > t \Leftrightarrow x^{a-1}(1-x)^{b-1} > t$$

dla pewnej liczby t.

Zauważmy, że $x^{a-1}(1-x)^{b-1} > t$ dla $\{x : x_0 < x < x_1\}$ dla pewnych x_0 oraz x_1 takich, że $x_0^{a-1}(1-x_0)^{b-1} = x_1^{a-1}(1-x_1)^{b-1}$. Zatem, obszar krytyczny testu ma postać przedziału (x_0, x_1), przy czym liczby x_0, x_1 dobrane są tak, że poziom istotności testu wynosi α, czyli

$$E_{H_0}\phi(X) = P_{U(0,1)}\{x_0 < X < x_1\} \le \alpha.$$

Ponieważ $P_{U(0,1)}\{x_0 < X < x_1\} = x_1 - x_0$, więc obszar krytyczny testu ma postać $(x_0, x_0 + \alpha)$. Liczba x_0 jest rozwiązaniem równania:

$$\left[\frac{x_0 + \alpha}{x_0}\right]^{a-1} \left[\frac{1 - x_0 - \alpha}{1 - x_0}\right]^{b-1} = 1$$

To równanie, dla danych a i b można rozwiązać numerycznie. Przykładowe rozwiązanie pokazane są w poniższej tabeli.

a	b	x_0	x_1
2	2	0.47500	0.52500
2	3	0.30865	0.35865
2	4	0.22556	0.27556
3	2	0.64135	0.69135
3	3	0.47500	0.52500
3	4	0.37517	0.42517
4	2	0.72444	0.77444
4	3	0.57483	0.62483
4	4	0.47500	0.52500

Na przykład, test najmocniejszy dla problemu weryfikacji hipotezy $H_0 : U(0,1)$ przeciwko $H_1 : B(2,3)$ wygląda następująco: odrzucić hipotezę zerową, jeżeli zaobserwowano wartość z przedziału $(0.30865, 0.35865)$. W przeciwnym przypadku nie ma podstaw do odrzucenia hipotezy zerowej. Moc tego testu jest równa

$$P_{B(2,3)}\{X \in (0.30865, 0.35865)\} = 0.08876.$$

□

Niech co najmniej jedna z hipotez H_0 i/lub H_1 będzie hipotezą złożoną. Do konstrukcji odpowiedniego testu wykorzystywany jest iloraz wiarogodności

$$\lambda(x) = \frac{\sup\limits_{\theta\in\Theta_1} f_\theta(x)}{\sup\limits_{\theta\in\Theta_0} f_\theta(x)} \quad \text{lub} \quad \lambda(x) = \frac{\sup\limits_{\theta\in\Theta} f_\theta(x)}{\sup\limits_{\theta\in\Theta_0} f_\theta(x)}$$

Test hipotezy H_0 przeciwko H_1 ma postać:

$$\phi(x) = \begin{cases} 1, & \text{gdy } \lambda(x) > t \\ \gamma, & \text{gdy } \lambda(x) = t \\ 0, & \text{gdy } \lambda(x) < t \end{cases}$$

Stała t dobierana jest tak, by test miał z góry zadany poziom istotności:

$$\sup_{\theta\in\Theta_0} E_\theta \phi(X) \le \alpha$$

Przykład. *Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami. Niech $\mu_0 \in \mathbb{R}$. Skonstruować test dla hipotezy $H_0 : \mu = \mu_0$ przeciwko $H_1 : \mu \neq \mu_0$*

Rozwiązanie.

Model dla próby $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ ma postać

$$\left(\mathbb{R}^n, \left\{N_n\left(\mu \mathbf{1}_n, \sigma^2 \boldsymbol{I}_n\right),\ \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+\right\}\right)$$

Przestrzenią parametrów oraz zbiorami Θ_0 i Θ_1 są odpowiednio:

$$\Theta = \mathbb{R} \times \mathbb{R}_+ \quad \Theta_0 = \{\mu_0\} \times \mathbb{R}_+ \quad \Theta_1 = (\mathbb{R} \setminus \{\mu_0\}) \times \mathbb{R}_+$$

Zauważmy, że obie weryfikowane hipotezy są złożone. Do konstrukcji testu wykorzystamy iloraz wiarogodności

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{(\mu,\sigma)\in\Theta} f_{\mu,\sigma}(\boldsymbol{x})}{\sup\limits_{(\mu,\sigma)\in\Theta_0} f_{\mu,\sigma}(\boldsymbol{x})}$$

Próba ma rozkład o gęstości

$$f_{\mu,\sigma}(\boldsymbol{x}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right\}$$

Z teorii metody największej wiarogodności wiadomo, że $\sup_{(\mu,\sigma)\in\Theta} f_{\mu,\sigma}(\boldsymbol{x})$ *jest realizowane w punkcie* $(\bar{X}, S^2)$*, gdzie*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

są estymatorami największej wiarogodności odpowiednio wartości oczekiwanej i wariancji. A zatem

$$\sup_{(\mu,\sigma)\in\Theta} f_{\mu,\sigma}(\boldsymbol{x}) = f_{\bar{X},S^2}(\boldsymbol{x}) = \left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\right)^n \exp\left\{-\frac{n}{2}\right\}$$

Niech

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2$$

Jak łatwo sprawdzić, jest to estymator największej wiarogodności wariancji przy założeniu, że $\mu = \mu_0$*. Zatem*

$$\sup_{(\mu,\sigma)\in\Theta_0} f_{\mu,\sigma}(\boldsymbol{x}) = f_{\mu_0,\tilde{\sigma}^2}(\boldsymbol{x}) = \left(\frac{1}{\tilde{\sigma}\sqrt{2\pi}}\right)^n \exp\left\{-\frac{n}{2}\right\}$$

Iloraz wiarogodności przyjmuje więc postać

$$\lambda(\boldsymbol{x}) = \frac{\sup\limits_{\theta\in\Theta} f_{\mu,\sigma}(x)}{\sup\limits_{\theta\in\Theta_0} f_{\mu,\sigma}(x)} = \left(\frac{\tilde{\sigma}}{\hat{\sigma}}\right)^n$$

Ponieważ

$$\sum_{i=1}^{n}(X_i - \mu_0)^2 = \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu_0)^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2$$

więc

$$\lambda(\boldsymbol{x}) = \left(1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)^{\frac{n}{2}}$$

Poszukiwany test ma postać

$$\phi(x) = 1 \Leftrightarrow \lambda(\boldsymbol{x}) > t \Leftrightarrow \frac{n(\bar{X} - \mu_0)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} > t'$$

Stała t' musi być tak dobrana, by test miał poziom istotności α, tzn.

$$E_{H_0}\phi(X) = \alpha$$

Jak wiadomo

a. *jeżeli H_0 jest prawdziwa, to*

$$\sqrt{n}(\bar{X} - \mu_0) \sim N(0, \sigma^2) \quad \text{czyli} \quad \frac{n(\bar{X} - \mu_0)^2}{\sigma^2} \sim \chi_1^2$$

b. *dla wszystkich $\theta \in \Theta$*

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sim \chi_{n-1}^2$$

c. $\bar{X}$ *oraz* $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ *są niezależne*

A zatem, jeżeli hipoteza H_0 jest prawdziwa, to

$$\frac{n(\bar{X} - \mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \sim F_{1,n-1}$$

Hipoteza H_0 jest odrzucana na poziomie istotności α, jeżeli

$$(*) \qquad \frac{n(\bar{X} - \mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} > F\,(\alpha; 1, n-1)$$

Ponieważ $t_v = \sqrt{F_{1,v}}$, więc $(*)$ jest równoważne

$$(**) \qquad \frac{|\bar{X} - \mu_0|}{S}\sqrt{n} > t(\alpha; n-1)$$

Test $(**)$ nazywa się **testem Studenta**

□

Testy zgodności.

Rozważmy problem weryfikacji hipotezy H_0 : Cecha X ma rozkład F.

Test chi–kwadrat zgodności

Niech F będzie dowolnym rozkładem prawdopodobieństwa. Test chi–kwadrat zgodności przebiega w następujący sposób.

Próba zapisywana jest w postaci szeregu rozdzielczego:

Klasa	Liczebność
1	n_1
2	n_2
$\vdots$	$\vdots$
k	n_k

Statystyką testową jest

$$\chi^2_{\text{emp}} = \sum_{i=1}^{k} \frac{(n_i - n_i^t)^2}{n_i^t}$$

gdzie

$$n_i^t = Np_i^t, \; N = \sum_{i=1}^{k} n_i,$$

$$p_i^t = P_F\{X \text{ przyjęła wartość z klasy } i\}$$

Hipotezę H_0 odrzucamy na poziomie istotności α, jeżeli $\chi^2_{\text{emp}} > \chi^2(\alpha; k-u-1)$. Tutaj $\chi^2(\alpha; k-u-1)$ jest wartością krytyczną (u jest liczbą nieznanych parametrów hipotetycznego rozkładu F).

Test Kołmogorowa

Niech F będzie znanym rozkładem ciągłym. Próbę $X_1, \ldots, X_n$ porządkujemy niemalejąco: $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Tak uporządkowane obserwacje nazywamy statystykami pozycyjnymi

Statystyką testową jest

$$D_n = \max_{1 \leq i \leq n} \left\{ \max\left\{ \left| F(X_{(i)}) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - F(X_{(i)}) \right| \right\} \right\}$$

Wartością krytyczną testu Kołmogorowa jest $D(\alpha; n)$. Jeżeli $D_n > D(\alpha; n)$, to hipotezę H_0 odrzucamy.

Test Shapiro–Wilka

Niech F będzie rozkładem normalnym. Próbę $X_1, \ldots, X_n$ porządkujemy niemalejąco: $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$.

Statystyką testową jest

$$W = \frac{\left(\sum_{i=1}^{[n/2]} a_{i:n} (X_{(n-i+1)} - X_{(i)}) \right)^2}{\text{var} X}$$

gdzie $a_{i:n}$ są stablicowanymi współczynnikami oraz

$$[n/2] = \begin{cases} n/2, & \text{dla } n \text{ parzystych} \\ (n-1)/2, & \text{dla } n \text{ nieparzystych} \end{cases}$$

Wartością krytyczną testu Shapiro–Wilka jest $W_n(\alpha)$. Jeżeli $W \leq W_n(\alpha)$, to hipotezę H_0 odrzucamy.

Zadania do samodzielnego rozwiązania

22.1. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $D(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0, 1)$. Skonstruować

a. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$ dla znanych wartości $\theta_0 < \theta_1$.

b. test hipotezy $H_0 : \theta \leq \theta_0$ przeciwko $H_1 : \theta \geq \theta_0$ dla ustalonej wartości θ_0.

c. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \neq \theta_0$ dla ustalonej wartości θ_0.

22.2. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Bin\,(m, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Skonstruować

a. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$ dla znanych wartości $\theta_0 < \theta_1$.

b. test hipotezy $H_0 : \theta \le \theta_0$ przeciwko $H_1 : \theta \ge \theta_0$ dla ustalonej wartości θ_0.

c. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \ne \theta_0$ dla ustalonej wartości θ_0.

22.3. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Ge\,(\theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Skonstruować

a. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$ dla znanych wartości $\theta_0 < \theta_1$.

b. test hipotezy $H_0 : \theta \le \theta_0$ przeciwko $H_1 : \theta \ge \theta_0$ dla ustalonej wartości θ_0.

c. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \ne \theta_0$ dla ustalonej wartości θ_0.

22.4. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $NB\,(r, \theta)$ z nieznanym prawdopodobieństwem $\theta \in (0,1)$. Skonstruować

a. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$ dla znanych wartości $\theta_0 < \theta_1$.

b. test hipotezy $H_0 : \theta \le \theta_0$ przeciwko $H_1 : \theta \ge \theta_0$ dla ustalonej wartości θ_0.

c. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \ne \theta_0$ dla ustalonej wartości θ_0.

22.5. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Po\,(\lambda)$ z nieznanym parametrem $\lambda > 0$. Skonstruować

a. test hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda = \lambda_1$ dla znanych wartości $\lambda_0 < \lambda_1$.

b. test hipotezy $H_0 : \lambda \le \lambda_0$ przeciwko $H_1 : \lambda \ge \lambda_0$ dla ustalonej wartości λ_0.

c. test hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda \ne \lambda_0$ dla ustalonej wartości λ_0.

22.6. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U\,(0, \theta)$ z nieznanym parametrem $\theta > 0$. Skonstruować

a. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta = \theta_1$ dla znanych wartości $\theta_0 < \theta_1$.

b. test hipotezy $H_0 : \theta \le \theta_0$ przeciwko $H_1 : \theta \ge \theta_0$ dla ustalonej wartości θ_0.

c. test hipotezy $H_0 : \theta = \theta_0$ przeciwko $H_1 : \theta \ne \theta_0$ dla ustalonej wartości θ_0.

22.7. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $E\,(\lambda)$ z nieznanym parametrem $\lambda > 0$. Skonstruować

a. test hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda = \lambda_1$ dla znanych wartości $\lambda_0 < \lambda_1$.

b. test hipotezy $H_0 : \lambda \le \lambda_0$ przeciwko $H_1 : \lambda \ge \lambda_0$ dla ustalonej wartości λ_0.

c. test hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda \ne \lambda_0$ dla ustalonej wartości λ_0.

22.8. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $G\,(\alpha, \lambda)$ z nieznanym parametrem $\lambda > 0$ i znanym parametrem α. Skonstruować

a. test hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda = \lambda_1$ dla znanych wartości $\lambda_0 < \lambda_1$.

b. test hipotezy $H_0 : \lambda \le \lambda_0$ przeciwko $H_1 : \lambda \ge \lambda_0$ dla ustalonej wartości λ_0.

c. test hipotezy $H_0 : \lambda = \lambda_0$ przeciwko $H_1 : \lambda \ne \lambda_0$ dla ustalonej wartości λ_0.

22.9. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ ze znaną wariancją $\sigma^2 > 0$. Skonstruować

a. test hipotezy $H_0 : \mu = \mu_0$ przeciwko $H_1 : \mu = \mu_1$ dla znanych wartości $\mu_0 < \mu_1$.

b. test hipotezy $H_0 : \mu \leq \mu_0$ przeciwko $H_1 : \mu \geq \mu_0$ dla ustalonej wartości μ_0.

c. test hipotezy $H_0 : \mu = \mu_0$ przeciwko $H_1 : \mu \neq \mu_0$ dla ustalonej wartości μ_0.

22.10. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami. Skonstruować

a. test hipotezy $H_0 : \mu = \mu_0$ przeciwko $H_1 : \mu = \mu_1$ dla znanych wartości $\mu_0 < \mu_1$.

b. test hipotezy $H_0 : \mu \leq \mu_0$ przeciwko $H_1 : \mu \geq \mu_0$ dla ustalonej wartości μ_0.

c. test hipotezy $H_0 : \mu = \mu_0$ przeciwko $H_1 : \mu \neq \mu_0$ dla ustalonej wartości μ_0.

d. test hipotezy $H_0 : \sigma^2 = \sigma_0^2$ przeciwko $H_1 : \sigma^2 = \sigma_1^2$ dla znanych wartości $\sigma_0^2 < \sigma_1^2$.

e. test hipotezy $H_0 : \sigma^2 \leq \sigma_0^2$ przeciwko $H_1 : \sigma^2 \geq \sigma_0^2$ dla ustalonej wartości σ_0^2.

f. test hipotezy $H_0 : \sigma^2 = \sigma_0^2$ przeciwko $H_1 : \sigma^2 \neq \sigma_0^2$ dla ustalonej wartości σ_0^2.

22.11. Niech X będzie pojedynczą obserwacją z rozkładu $Pow\,(a+1)$. Testujemy $H_0 : a = 1$ przeciwko $H_1 : a = 2$. Dysponujemy pojedynczą obserwacją X. Wyznaczyć obszar krytyczny najmocniejszego testu o rozmiarze 0.1.

22.12. Zmienna losowa X ma gęstość prawdopodobieństwa $f(x)$. Na podstawie pojedynczej obserwacji X przeprowadzamy test hipotezy

$$H_0: \; f(x) = \begin{cases} 1, & \text{dla } 0 < x < 1, \\ 0, & \text{poza tym,} \end{cases} \quad \text{przeciwko } H_1: \; f(x) = \begin{cases} 5x^4, & \text{dla } 0 < x < 1, \\ 0, & \text{poza tym.} \end{cases}$$

Wyznaczyć moc najmocniejszego testu na poziomie istotności α.

22.13. Rozważamy zadanie testowania, na podstawie pojedynczej obserwacji X, hipotezy prostej H_0 : X*pochodzi z rozkladu* f_0 przeciwko prostej alternatywie H_1 : X*pochodzi z rozkladu* f_1. Wiadomo, że dla każdego $\alpha \in (0,1)$ najmocniejszy test na poziomie istotności α odrzucający H_0, jeśli $X > k = k(\alpha)$, ma moc $1-\beta$ taką, że $1 - \beta = 1 - \beta(\alpha) = \alpha^2$. Gęstość f_0 dla $x > 0$ określona jest wzorem $\frac{1}{(1+x)^2}$. Jakim wzorem (dla dodatnich x) określona jest gęstość f_1?

22.14. Załóżmy, że dysponujemy pojedynczą obserwacją X z rozkładu $P \in \{P_0, P_1\}$, gdzie P_0 jest rozkładem $N\,(0,1)$ i P_1 jest rozkładem $Lap\,(0,1)$. Rozważmy zadanie testowania hipotezy $H_0 : P = P_0$ przeciw alternatywie $H_1 : P = P_1$. Podać rozmiar testu najmocniejszego, jeśli wiadomo, że obszar krytyczny testu jest sumą przedziałów rozłącznych, z których jeden jest równy $(-\infty, -1.9)$.

22.15. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\,(\mu, 4)$. Rozważmy najmocniejszy test hipotezy $H_0 : \mu = 0$ przeciw alternatywie $H_1 : \mu = 1$, na poziomie istotności 0.01. Ile potrzeba obserwacji, żeby moc testu była większa niż 0.9?

22.16. Załóżmy, że $X_1, \ldots, X_n$ jest próbą z rozkładu $N(\mu, 1)$ z nieznanym μ. Znaleźć najmniejsze n, dla którego istnieje test hipotezy $H_0 : \mu = 10.0$ przeciwko alternatywie $H_1 : \mu = 10.1$ na poziomie istotności 0.05 o mocy przynajmniej 0.50.

22.17. Niech $X_1, \ldots, X_9$ będzie próbą z rozkładu normalnego $N(\mu, 1)$ o nieznanej wartości oczekiwanej μ. Rozpatrzmy zadanie testowania hipotezy $H_0 : \mu = 0$ przeciwko alternatywie $H_1 : \mu = 0.5$. Należy zbudować taki test, dla którego suma prawdopodobieństw błędów I i II rodzaju, oznaczanych odpowiednio przez α i β jest najmniejsza. Obliczyć tę najmniejszą wartość.

22.18. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N(\mu, 1)$. Hipotezę $\mu = 0$ weryfikuje się za pomocą testu z obszarem krytycznym $\{(x_1, x_2, \ldots, x_n) : \sqrt{n}|\overline{x}| > 2\}$. Jaki jest rozmiar tego testu? Naszkicować wykres mocy tego testu.

22.19. Zmienne losowe $X_1, \ldots, X_9$ są niezależne, X_i ma rozkład $N\left(\mu\sqrt{i}, 1\right)$ dla $i = 1, \ldots, 9$. Rozważamy hipotezy statystyczne $H_0 : \mu = 0$ i $H_1 : \mu > 0$. Znaleźć obszar krytyczny testu jednostajnie najmocniejszego hipotezy zerowej H_0 przeciw alternatywie H_1 na poziomie istotności 0.025.

22.20. Próba $X_1, \ldots, X_n$ pochodzi z rozkładu $N(\mu, 1)$ z nieznaną wartością oczekiwaną μ i wariancją 1. Na podstawie tej próby zbudowano w standardowy sposób przedział ufności na poziomie 0.95 dla μ:

$$\left[\overline{X} - \frac{1.96}{\sqrt{n}}, \overline{X} + \frac{1.96}{\sqrt{n}}\right] = [\hat{\mu}_-, \hat{\mu}_+]$$

Pokazać, że jednostajnie najmocniejszy test hipotezy $H_0 : \mu = 3$ przeciw alternatywie $H_1 : \mu > 3$ na poziomie istotności 0.025 odrzuca H_0 wtedy i tylko wtedy, gdy $\hat{\mu}_- > 3$.

22.21. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$ z nieznanymi parametrami μ i σ^2. Rozważamy problem testowania hipotezy $H_0 : \mu = 0$ przy alternatywie $H_1 : \mu \neq 0$ za pomocą testu, który odrzuca H_0, jeśli $\frac{|\overline{X}|}{Z} > t$, gdzie $Z = \sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}$. Dobrać stałą t tak, aby prawdopodobieństwo błędu pierwszego rodzaju testu było równe 0.05, jeśli wiadomo, że $n = 9$.

22.22. Za pomocą próby $X_1, \ldots, X_n$ z rozkładu $N(\mu, 1)$ weryfikuje się hipotezę, że $\mu = 0$. Rozważa się dwa testy dla weryfikacji tej hipotezy, mianowicie testy z obszarami krytycznymi

$$R_1 = \{(x_1, \ldots, x_n) : |\overline{x}| > k_1\}, \quad R_2 = \{(x_1, \ldots, x_n) : \sum x_i^2 > k_2\},$$

gdzie k_1 i k_2 są dobrane tak, aby testy miały rozmiar α. Czy któryś z tych testów jest jednostajnie mocniejszy od drugiego?

22.23. Każda ze zmiennych losowych $X_1, \ldots, X_n$ ma rozkład $N\left(\mu, \sigma^2\right)$ z nieznaną wartością oczekiwaną μ i znaną wariancją σ^2. Założono, że zmienne są niezależne i wyznaczono test jednostajnie najmocniejszy dla testowania hipotezy $H_0 : \mu = \mu_0$ przy alternatywie $H_1 : \mu > \mu_0$ na poziomie istotności α. W rzeczywistości zmienne losowe $X_1, \ldots, X_n$ mają łączny rozkład normalny, ale są skorelowane i współczynnik korelacji $Corr(X_i, X_j) = \varrho$ dla wszystkich $i \neq j$. Wyznaczyć rzeczywisty rozmiar testu.

22.24. Rozważmy następujące zagadnienie testowania hipotez statystycznych. Dysponujemy próbą $X_1, \ldots, X_n$ z rozkładu $N(\mu, 1)$. Przeprowadzamy najmocniejszy test hipotezy $H_0 : \mu = 0$ przeciwko alternatywie $H_1 : \mu = 1$ na poziomie istotności 0.5. Oczywiście, moc tego testu zależy od rozmiaru próby. Niech β_n oznacza prawdopodobieństwo błędu drugiego rodzaju, dla rozmiaru próby n. Udowodnić, że

$$\lim_{n\to\infty} \frac{\beta_n}{e^{-n/2}/\sqrt{2\pi n}} = 1.$$

22.25. Zmienne losowe $X_1, \ldots, X_n$ są niezależne, przy czym X_i ma rozkład $N(\theta_i, 1)$. Weryfikuje się hipotezę, że wszystkie θ_i są równe zeru przy hipotezie alternatywnej, że $\theta_i = 0.5$ dla $i = 1, 2, \ldots, r$ i $\theta_i = -0.5$ dla $i = r+1, \ldots, n$. Pokazać, że najmocniejszy test o rozmiarze 0.05 ma obszar krytyczny

$$\left\{(x_1, x_2, \ldots, x_n) : \sum_{i=1}^{r} x_i - \sum_{i=r+1}^{n} x_i > 1.645\sqrt{n}\right\}.$$

Jak duże musi być n, aby moc tego testu była równa co najmniej 0.9?

22.26. $X_1, \ldots, X_n$ jest próbą z rozkładu $N(0, \sigma^2)$. Rozważmy najmocniejszy test hipotezy $H_0 : \sigma^2 = 1$ przeciwko alternatywie $H_1 : \sigma^2 = 3$ na poziomie istotności 0.1. Wyznaczyć moc tego testu.

22.27. Niech X_1, X_2, X_3, X_4 będzie czteroelementową próbą z rozkładu $N(\mu, \sigma^2)$ o nieznanych parametrach. Testujemy hipotezę $H_0 : \sigma^2 = \sigma_0^2$ przeciw alternatywie $H_0 : \sigma^2 > \sigma_0^2$ za pomocą statystyki $\frac{\mathrm{var}X}{\sigma_0^2}$, przyjmując poziom istotności 0.05. Zaobserwowaliśmy $(1, -1.2, 3, 0.7)$. Jaka jest minimalna możliwa wartość σ_0^2, skoro wiadomo, że test wykazał, że nie mamy podstaw do odrzucenia hipotezy H_0?

22.28. Załóżmy, że $X_1, \ldots, X_n$ jest próbą z rozkładu $N(\mu, \sigma^2)$ ze znaną średnią μ i nieznaną wariancją σ^2. Rozważmy test jednostajnie najmocniejszy hipotezy $H_0 : \sigma^2 \leq 1$ przeciw alternatywie $H_1 : \sigma^2 > 1$ na poziomie istotności 0.05. Dla jakich wariancji σ^2 moc testu przekracza 0.9?

22.29. $X_1, \ldots, X_n$ jest próbą z rozkładu $N(\mu, \sigma^2)$ o znanej wartości oczekiwanej μ i nieznanej wariancji σ^2. Rozważmy test hipotezy $H_0 : \sigma^2 \leq 4$ przeciwko alternatywie $H_1 : \sigma^2 > 4$, który jest najmocniejszy na poziomie istotności 0.05. Dla jakich wartości wariancji moc tego testu jest nie mniejsza niż 0.95?

22.30. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi z rozkładu normalnego, przy czym $EX_i = 0$ i $D^2X_i = \frac{\sigma^2}{i}$, gdzie σ^2 jest nieznanym parametrem. Rozważamy jednostajnie najmocniejszy test hipotezy $H_0 : \sigma^2 \leq 4$ przy alternatywie $H_1 : \sigma^2 > 4$ na poziomie istotności 0.05. Dla jakich wartości wariancji moc tego testu jest nie mniejsza niż 0.95?

22.31. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $N\left(\mu, \sigma^2\right)$, przy czym parametry μ i σ^2 są nieznane. Weryfikuje się hipotezę zerową, że $\sigma = \sigma_0$ wobec hipotezy alternatywnej $\sigma \neq \sigma_0$. Pokazać, że obszar krytyczny testu opartego na ilorazie wiarogodności ma postać

$$\left\{(x_1, x_2, \ldots, x_n) : k_1 \leq \frac{s^2}{\sigma_0^2} \leq k_2\right\},$$

gdzie $S^2 = \mathrm{var}X/n$. Jak należy ustalić k_1 i k_2, aby test miał rozmiar α.

22.32. Załóżmy, że dysponujemy pojedynczą obserwacją X z rozkładu $N\left(\mu, \sigma^2\right)$. Rozważmy zadanie testowania hipotezy $H_0 : \mu = 0, \sigma^2 = 1$ przeciw alternatywie $H_1 : \mu = 1, \sigma^2 = 4$. Najmocniejszy test na poziomie istotności α jest postaci

$$\text{odrzucić } H_0, \text{ gdy } X \notin (-2, b).$$

Podać b i poziom istotności α.

22.33. Wiadomo, że $X_1, \ldots, X_n$ jest próbą z rozkładu $N(\mu - \theta, 1)$, zaś $Y_1, \ldots, Y_n$ jest niezależną próbą z rozkładu $N(\mu + \theta, 1)$. Liczby μ i θ są nieznanymi parametrami. Rozpatrujemy zadanie testowania hipotezy $H_0 : \theta = 0$ przeciw alternatywie $H_1 : \theta = 1$. Dla jakich n można skonstruować test na poziomie istotności 0.05 o mocy przynajmniej 0.95?

22.34. Niech $X_1, \ldots, X_m$ będzie próbą z rozkładu $N\left(\mu_1, \sigma^2\right)$ i niech $Y_1, \ldots, Y_n$ będzie próbą z rozkładu $N\left(\mu_2, \sigma^2\right)$ każda. Wszystkie zmienne są niezależne. Hipotezę $H_0 : \mu_1 = \mu_2$ przy alternatywie $H_1 : \mu_1 > \mu_2$ weryfikujemy w następujący sposób. Zliczamy liczbę S elementów w próbie $X_1, \ldots, X_m$ większych od wszystkich elementów próby $Y_1, \ldots, Y_n$. Hipotezę H_0 odrzucamy, gdy $S \geq s$, gdzie s jest wartością krytyczną. Przypuśćmy, że $m = 7$ i $n = 8$. Wyznaczyć rozmiar testu, gdy $s = 2$.

22.35. Niech $X_1, \ldots, X_9$ będzie próbą z rozkładu $N(\mu_1, 1)$, a $Y_1, \ldots, Y_9$ próbą z rozkładu $N(\mu_2, 4)$. Zakładając, że wszystkie zmienne losowe są niezależne wyznaczono test oparty na ilorazie wiarogodności dla testowania hipotezy $H_0 : \mu_1 = \mu_2$ przy alternatywie $H_1 : \mu_1 \neq \mu_2$ na poziomie istotności 0.05. W rzeczywistości założenie to nie jest spełnione: co prawda pary zmiennych $(X_1, Y_1), \ldots, (X_9, Y_9)$ są niezależne, ale X_i, Y_i są zależne i współczynnik korelacji $Corr(X_i, Y_i) = 0.5$ dla $i = 1, \ldots, 9$. Obliczyć faktyczne prawdopodobieństwo błędu pierwszego rodzaju α i moc testu β przy alternatywie $\mu_1 = \mu_2 + 2$.

22.36. Niech $X_1, \ldots, X_{15}$ będzie próbą z rozkładu $N\left(\mu_1, \sigma^2\right)$, a $Y_1, \ldots, Y_{15}$ próbą z rozkładu $N\left(\mu_2, \sigma^2\right)$. Wszystkie zmienne są niezależne, a parametry μ_1, μ_2, σ^2 są nieznane. Testujemy hipotezę $H_0 : \mu_1 = \mu_2$ przy alternatywie $H_1 : \mu_1 \neq \mu_2$. Hipotezę H_0 odrzucamy, gdy spełniona jest nierówność

$$\frac{|\overline{X} - \overline{Y}|}{S} > c,$$

gdzie $S^2 = \sum_{i=1}^{15}(X_i - Y_i)^2$. Wyznaczyć c tak, aby rozmiar testu był równy 0.05.

22.37. Obserwujemy parę (X, Y) zmiennych losowych. Zakładamy, że są to zmienne niezależne, X ma rozkład $N(\mu_X, 1)$ i Y ma rozkład normalny $N(\mu_Y, 1/9)$. Rozważmy najmocniejszy test hipotezy $H_0 : (\mu_X, \mu_Y) = (0, 0)$ przeciwko alternatywie $H_1 : (\mu_X, \mu_Y) = (1, 1)$ na poziomie istotności 0.1. Wyznaczyć moc tego testu.

22.38. Obserwujemy pary $(X_1, Y_1), \ldots, (X_{20}, Y_{20})$. Zmienne $X_1, \ldots, X_{20}$ są zmiennymi losowymi o rozkładzie $N(\mu_X, 1)$, a zmienne $Y_1, \ldots, Y_{20}$ o rozkładzie $N(\mu_Y, 4)$. Wszystkie zmienne są niezależne. Dla najmocniejszego testu hipotezy $H_0 : (\mu_X, \mu_Y) = (0, 0)$ przeciw alternatywie $H_1 : (\mu_X, \mu_Y) = (1, 1)$ na poziomie istotności 0.01, wyznaczyć prawdopodobieństwo błędu drugiego rodzaju tego testu.

22.39. Mamy próbę $(X_1, Y_1), \ldots, (X_n, Y_n)$ z dwuwymiarowego rozkładu normalnego o nieznanych parametrach:

$$EX_i = EY_i = \mu,\ D^2X_i = D^2Y_i = \sigma^2,\ Cov(X_i, Y_i) = \varrho\sigma^2.$$

Niech $Z_i = X_i + Y_i$, $R_i = X_i - Y_i$, $S_Z^2 = \frac{1}{n-1}\mathrm{var}Z$, $S_R^2 = \frac{1}{n-1}\mathrm{var}R$. Niech $\varrho_0 \in (-1, 1)$ będzie ustaloną liczbą, $\varrho_0 \neq 0$. Do testowania hipotezy $H_0 : \varrho = \varrho_0$ przeciwko alternatywie $H_1 : \varrho \neq \varrho_0$ możemy użyć testu opartego na statystyce S_Z^2/S_R^2. Wyznaczyć rozkład tej statystyki przy założeniu prawdziwości hipotezy zerowej.

22.40. Rozważmy prosty model regresji liniowej bez wyrazu wolnego $Y_i = \theta x_i + \varepsilon_i$ $(i = 1, \ldots, n)$, gdzie (Y_i, x_i) to obserwacje par (losowa zmienna zależna, nielosowa zmienna niezależna), θ jest nieznanym parametrem, a $(\varepsilon_1, \ldots, \varepsilon_n)$ są nawzajem niezależnymi zmiennymi losowymi o identycznym rozkładzie $N(0, \sigma^2)$. Rozważmy jednostajnie najmocniejszy test hipotezy $H_0 : \theta = 0$ przeciw alternatywie $H_1 : \theta > 0$ na poziomie istotności 0.05. Wyznaczyć obszar krytyczny tego testu.

22.41. Zależność czynnika Y od czynnika x (nielosowego) opisuje model regresji liniowej $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, gdzie błędy ε_i są niezależne i mają rozkłady $N(0, 1)$. Obserwujemy zmienne losowe $Y_1, \ldots, Y_n$ przy danych wartościach $X_1, \ldots, X_n$. Wyznaczyć obszar krytyczny testu najmocniejszego dla weryfikacji hipotezy $H_0 : \beta_0 = 0$ i $\beta_1 = 1$ przy alternatywie $H_1 : \beta_0 = 1$ i $\beta_1 = 2$ na poziomie istotności 0.05.

22.42. Niech $X_1, \ldots, X_n$ oraz $Y_1, \ldots, Y_n$ będą niezależnymi próbami z dwóch rozkładów $E(\lambda)$ i $E(\mu)$ o nieznanych parametrach. Weryfikuje się hipotezę zerową $\lambda = \mu$ wobec hipotezy alternatywnej $\lambda \neq \mu$. Pokazać, że obszar krytyczny testu opartego na ilorazie wiarogodności zależy tylko od ilorazu $\overline{Y}/\overline{X}$. Podać sposób konstrukcji testu o rozmiarze α.

22.43. Niech X będzie pojedynczą obserwacją z przesuniętego rozkładu wykładniczego $E(1; \theta)$, gdzie $\theta \geq 0$ jest nieznanym parametrem. Rozważmy jednostajnie najmocniejszy test hipotezy $H_0 : \theta = 0$ przeciwko alternatywie $H_1 : \theta > 0$, na poziomie istotności 0.05. Wyznaczyć zbiór tych wartości θ, dla których moc testu wynosi co najmniej 0.90.

22.44. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu o dystrybuancie

$$F_\theta(x) = \begin{cases} 1 - 2^{-(x-\theta)}, & \text{dla } x > \theta, \\ 0, & \text{poza tym,} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Rozważmy jednostajnie najmocniejszy test hipotezy $H_0 : \theta = 0$ przeciw alternatywie $H_1 : \theta > 0$ na poziomie istotności 0.01. Przy jakiej liczebności próby n, w danym punkcie $\theta_1 > 0$ funkcja mocy tego testu przybiera wartość większą lub równą 0.64?

22.45. Zmienne losowe $X_1, \ldots, X_n$ są niezależne i mają jednakowy rozkład $E(1;\theta)$. Hipotezę zerową $\{\theta : \theta \le 1\}$ przy hipotezie alternatywnej $\{\theta : \theta > 1\}$ weryfikuje się za pomocą testu z obszarem krytycznym postaci

$$\{(x_1, x_2, \ldots, x_n) : \min(x_1, x_2, \ldots, x_n) > c\}.$$

Wyznaczyć c tak, aby test miał rozmiar α. Naszkicować wykres funkcji mocy.

22.46. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $G(q, 1/\theta)$. Zakładamy, że q jest znane. Skonstruować najmocniejszy test na poziomie istotności α dla weryfikacji hipotezy $\{\theta_0\}$ wobec alternatywy $\{\theta_1\}$, gdzie $\theta_1 > \theta_0$, i pokazać, że istnieje test JNM dla weryfikacji $\{\theta_0\}$ wobec $\{\theta : \theta > \theta_0\}$. Pokazać, że gdy $q = 1/n$, moc tego testu jest równa $1 - (1-\alpha)^{\theta/\theta_0}$.

22.47. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $G(p, \lambda)$, gdzie $\lambda > 0$ jest znane, zaś $p > 0$ jest nieznane. Wyznaczyć obszar krytyczny jednostajnie najmocniejszego testu hipotezy $H_0 : p = 2$ przeciw hipotezie $H_1 : p > 2$ na poziomie istotności α.

22.48. Niech X będzie pojedynczą obserwacją z rozkładu $Par(0, \lambda)$. Rozważmy najmocniejszy test hipotezy $H_0 : \lambda = 1$ przeciwko alternatywie $H_1 : \lambda = 101$ na poziomie istotności 0.01. Wyznaczyć moc tego testu.

22.49. Niech X_1, X_2, X_3, X_4 będzie próbą z rozkładu $Par(2, \lambda_1)$, a Y_1, Y_2, Y_3, Y_4, Y_5 będzie próbą z rozkładu $Par(2, \lambda_2)$, gdzie θ_1 i θ_2 są nieznanymi parametrami dodatnimi. Wszystkie zmienne losowe są niezależne. Testujemy hipotezę $H_0 : \frac{\theta_1}{\theta_2} = 1$ przy alternatywie $H_1 : \frac{\theta_1}{\theta_2} > 1$ za pomocą testu o obszarze krytycznym

$$K = \left\{ \frac{\hat{\theta}_1}{\hat{\theta}_2} > t \right\},$$

gdzie $\hat{\theta}_1$ i $\hat{\theta}_2$ są estymatorami największej wiarogodności odpowiednio parametrów θ_1 i θ_2 wyznaczonymi na podstawie prób X_1, X_2, X_3, X_4 i Y_1, Y_2, Y_3, Y_4, Y_5. Dobrać stałą t tak, aby otrzymać test o rozmiarze 0.05.

22.50. Niech $X_1, \ldots, X_5$ będzie próbą z rozkładu $Par(1, \theta)$. Skonstruować jednostajnie najmocniejszy (na poziomie istotności 0.01) test hipotezy $H_0 : \theta = 0.5$ przeciwko $H_1 : \theta > 0.5$.

22.51. Dysponujemy pojedynczą obserwacją X z rozkładu $Lap\,(\mu, \lambda)$, gdzie $\lambda > 0$ i $\mu \in R$ są parametrami. Rozważmy zadanie testowania hipotezy

$$H_0 : \mu = 0 \text{ i } \lambda = 1 \text{ przeciw } H_1 : \mu = -1 \text{ i } \lambda = 0.5$$

Obszar krytyczny najmocniejszego testu na poziomie istotności α jest postaci

$$K = \{x : x \notin (a, 3)\}.$$

Wyznaczyć a i poziom istotności α.

22.52. Niech X będzie pojedynczą obserwacją z rozkładu $Lap\,(\theta, 1)$, gdzie θ jest nieznanym parametrem. Rozważmy jednostajnie najmocniejszy test hipotezy $H_0 : \theta = 0$ przeciw hipotezie alternatywnej $H_1 : \theta > 0$ na poziomie istotności α, gdzie $\alpha < 0.5$, oparty na pojedynczej obserwacji X. Dla jakiego θ funkcja mocy tego testu osiąga wartość 0.75?

22.53. Urna zawiera r kul ponumerowanych liczbami $1, 2, \ldots, r$. Liczba kul r jest nieznanym parametrem, o którym wiemy, że jest większy od 5. Wybieramy z urny pięć kul, losując je bez zwracania. Na podstawie numerów wylosowanych kul testujemy hipotezę zerową $H_0 : r = 25$ przeciwko alternatywie $H_1 : r = 48$. Obliczyć moc najmocniejszego testu na poziomie istotności 0.2.

22.54. W celu zweryfikowania hipotezy, że nieznane prawdopodobieństwo sukcesu jest mniejsze od 0.5 wykonuje się 20 niezależnych prób i hipotezę uważa się za prawdziwą, gdy liczba sukcesów jest mniejsza od 12. Wykreślić funkcje prawdopodobieństw błędów tego testu.

22.55. Właściciel jeziora odsprzedający prawo połowów w tym jeziorze twierdzi, że w jeziorze jest co najmniej N ryb, przy czym N jest dużą liczbą. Dla sprawdzenia tego twierdzenia wyławia się m ryb, znakuje się je i wpuszcza z powrotem do jeziora. Po pewnym czasie, gdy oznakowane ryby rozproszą się po całym jeziorze, wyławia się n ryb; okazuje się, że wśród nich jest r ryb oznakowanych. Twierdzenie właściciela jeziora odrzuca się, gdy iloraz r/n jest większy od pewnej liczby k. Jak należy wybrać k, aby prawdopodobieństwo odrzucenia twierdzenia, gdy jest ono prawdziwe, było nie większe niż 0.1. (Można założyć, że liczby m i n są małe w porównaniu z N.)

22.56. Prawdopodobieństwo, że w pewnym doświadczeniu zaobserwuje się r cząstek jest równe $e^{-\lambda}\lambda^r/r!$ $(r = 0, 1, 2, \ldots)$. Pokazać, że prawdopodobieństwo zdarzenia polegającego na tym, że w n niezależnych powtórzeniach tego doświadczenia zaobserwuje się łącznie N cząstek, jest równe $e^{-n\lambda}(n\lambda)^N/N!$. Przypuśćmy, że wiadomo, iż λ jest równe 0.5 lub 1. Na podstawie pięciu niezależnych powtórzeń doświadczenia należy zdecydować, jaką wartość ma λ. Porównać dwie następujące reguły.

REGUŁA 1. Przyjąć, że $\lambda = 0.5$ wtedy i tylko wtedy, gdy łączna liczba zaobserwowanych cząstek jest mniejsza od czterech.

REGUŁA 2. Przyjąć, że $\lambda = 0.5$ wtedy i tylko wtedy, gdy w więcej niż dwóch powtórzeniach doświadczenia nie zaobserwowano ani jednej cząstki.

22.57. Zakładamy, że liczba roszczeń w ciągu roku dla pewnego portfela ryzyk jest zmienną losową X o rozkładzie $Po(\lambda)$. Zaobserwowano $X = 2600$. W oparciu o przybliżenie rozkładu Poissona rozkładem normalnym zbudowano test hipotezy H_0 : $EX = 2500$ przeciwko alternatywie H_1 : $EX > 2500$ o obszarze krytycznym postaci $X > c$. Czy test odrzuca H_0 na poziomie istotności 0.05?

22.58. Wykonując pewne doświadczenie można otrzymać jeden z $N+1$ możliwych wyników $z_0, z_1, \ldots, z_N$. Hipoteza zerowa H_0 przypisuje tym wynikom następujące prawdopodobieństwa:

$$P(z_0) = \frac{1}{2}, \quad P(z_i) = \frac{1}{2N}, \quad i = 1, 2, \ldots, N.$$

Hipoteza alternatywna jest hipotezą złożoną: przypisuje prawdopodobieństwa $(N-1)/N$ wynikowi z_0 i nie specyfikuje prawdopodobieństw wyników $z_1, z_2, \ldots, z_N$. Pokazać, że test o rozmiarze 0.5, oparty na ilorazie wiarogodności i wykorzystujący tylko pojedynczą obserwację, przyjmuje hipotezę H_0 wtedy i tylko wtedy, gdy wynikiem tej obserwacji jest z_0. Jaka jest moc tego testu? Czy jest to „dobry” test?

22.59. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(0, \theta)$. Pokazać, że istnieje test jednostajnie najmocniejszy na poziomie istotności α dla weryfikacji hipotezy $\{\theta_1\}$ przy hipotezie alternatywnej $\{\theta : \theta < \theta_1\}$. Czy istnieje test jednostajnie najmocniejszy na poziomie istotności α dla hipotezy $\{\theta_1\}$ wobec alternatywy $\{\theta : \theta \neq \theta_1\}$?

22.60. $U_1, \ldots, U_n, \ldots$ jest próbą z rozkładu $U(0, 1)$. Obserwujemy zmienną losową X. Niech H_0 : X *ma rozkład taki jak* $\min\{U_1, U_2, U_3\}$ oraz H_1 : X *ma rozkład taki jak* $\min\{U_1, U_2\}$. Wyznaczyć moc najmocniejszego testu na poziomie istotności 1/8.

22.61. Niech $X_1, \ldots, X_9$ będzie próbą z rozkładu $U(-\theta, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Hipotezę $H_0 : \theta = 2$ przy alternatywie $H_1 : \theta = 4$ weryfikujemy testem najmocniejszym na poziomie istotności 0.1. Wyznaczyć prawdopodobieństwo błędu drugiego rodzaju.

22.62. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(a, b)$, gdzie $0 \leq a < b$. Rozważmy zagadnienie weryfikacji hipotezy $H_0 : a = 0$ przeciw $H_1 : a > 0$. Wyznaczyć obszar krytyczny testu ilorazu wiarogodności.

22.63. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(0, \theta)$. W celu zweryfikowania hipotezy $H_0 : \theta = 1$ przeciw $H_1 : \theta > 1$ wykorzystujemy test o obszarze krytycznym $\max\{X_1, \ldots, X_n\} > c$. Niech n_0 będzie najmniejszą liczebnością próby taką, przy której test na poziomie istotności 0.20 ma w punkcie $\theta = \sqrt[3]{2}$ moc co najmniej 0.95. Obliczyć n_0.

22.64. Niech $X_1, \ldots, X_6$ będzie próbą z rozkładu $U(0, \theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Zbudowano test jednostajnie najmocniejszy dla weryfikacji hipotezy H_0 : $\theta = 1$ przy alternatywie $H_1 : \theta \neq 1$ na poziomie istotności 0.125. Wyznaczyć obszar krytyczny tego testu.

22.65. Niech $X_1, \ldots, X_{10}$ będzie próbą z rozkładu *Pow* (θ). Rozważmy test jednostajnie najmocniejszy hipotezy $H_0 : \theta = 1$ przeciwko alternatywie $H_1 : \theta > 1$ na poziomie istotności 0.01. Dla jakich wartości parametru θ ten test ma moc nie mniejszą niż 0.99?

22.66. Niech $X_1, \ldots, X_6$ będzie próbą z rozkładu *Pow* (θ), gdzie $\theta > 0$ jest nieznanym parametrem. Weryfikujemy hipotezę $H_0 : \theta = 1$ przy alternatywie $H_1 : \theta > 1$ testem jednostajnie najmocniejszym na poziomie istotności 0.05. Wyznaczyć moc tego testu przy $\theta = 3$.

Test chi-kwadrat

22.67. Wykonuje się k serii niezależnych prób, przy czym każdą serię kontynuuje się dopóty, dopóki określone zdarzenie E nie pojawi się dokładnie r razy. Niech prawdopodobieństwo zdarzenia E w każdej próbie i–tej serii będzie równe θ_i i niech n_i będzie liczbą prób w tej serii ($i = 1, 2, \ldots, k$). Poszczególne serie prób są niezależne. Weryfikuje się hipotezę, że $\theta_1 = \cdots = \theta_k$ wobec alternatywy, że nie wszystkie θ_i są sobie równe. Pokazać, że za podstawę testu można przyjąć statystykę

$$\sum_{i=1}^{k} \left\{ n_i \log \left[\frac{\bar{n}}{n_i} \right] + (n_i - r) \log \left[\frac{n_i - r}{\bar{n} - r} \right] \right\},$$

gdzie $\bar{n} = \sum n_i / k$. Pokazać, że dla dużych r test oparty na tej statystyce jest w przybliżeniu podobny.

22.68. Przypuśćmy, że za podstawę konstrukcji testu hipotezy $\theta_1 = \cdots = \theta_k$ w poprzednim zadaniu przyjęto koncepcje leżące u podstaw testu chi–kwadrat. Pokazać, że otrzymuje się test oparty na statystyce

$$\frac{r}{\bar{n}(\bar{n} - r)} \sum_{i=1}^{k} (n_i - \bar{n})^2.$$

22.69. Wykonuje się k serii niezależnych prób, po n prób w każdej serii. Niech $r_1, r_2, \ldots, r_k$ będzie liczbą tych prób w każdej serii, w których zaobserwowano zdarzenie E. Weryfikuje się hipotezę, że prawdopodobieństwo zdarzenia E jest takie samo w każdej próbie wobec alternatywnej, że prawdopodobieństwo tego zdarzenia jest stałe w każdej serii, ale w różnych seriach może być różne. Wyznaczyć statystykę chi–kwadrat.

22.70. Wyznaczyć test chi–kwadrat dla weryfikacji hipotezy o niezależności w tablicy kontyngencji.

22.71. Testujemy niezależność dwóch cech z tablicy kontygencji. Jedna z cech przyjmuje n, a druga m możliwych wartości. Ilość obserwacji w każdej z $(n \cdot m)$ komórek jest wystarczająca, aby zastosować test niezależności chi-kwadrat. Odpowiednia statystyka będzie miała asymptotycznie rozkład chi-kwadrat. Wyznaczyć ilość stopni swobody tego rozkładu.

22.72. Za pomocą testu chi-kwadrat zgodności testowano hipotezę, iż n elementowa próba pochodzi z rozkładu Poissona o wartości oczekiwanej równej jeden. Mamy niepełną informację o próbie, na podstawie której przeprowadzono test:

k	0	1	2	3 lub więcej
Ilość obserwacji	$n-70-4-25$	70	40	25

Wiadomo, iż na poziomie istotności 0.05 nie znaleziono podstaw do odrzucenia hipotezy o zgodności. Podać najmniejszą możliwą liczebność próby n.

22.73. Wykonano 120 rzutów dwiema kośćmi do gry: czarną i białą. 45 razy na białej kości wypadło więcej oczek niż na czarnej, 50 razy na białej kości wypadło mniej oczek niż na czarnej oraz 25 razy na białej kości wypadła ta sama liczba oczek co na czarnej. Rozważmy hipotezę H_0 : *obie kości są rzetelne i wynik rzutu kością białą jest niezależny od wyniku rzutu kością czarną*. W oparciu o podane wyniki przeprowadzono test chi-kwadrat. Pokazać, że na poziomie istotności 0.1 test nie prowadzi do odrzucenia H_0.

22.74. Wykonujemy n rzutów kością do gry i weryfikujemy hipotezę H_0 mówiącą, że kość jest rzetelna. Standardowy test chi-kwadrat na poziomie istotności 0.001 odrzuca hipotezę zerową, jeśli obliczona wartość statystyki testowej przekracza 20.515. Przypuśćmy, że wykonaliśmy tylko $n = 6$ rzutów. Jest to zbyt mało, żeby asymptotyczne przybliżenie rozkładu chi-kwadrat było zadowalające. Wyznaczyć faktyczny rozmiar testu.

22.75. Wykonujemy n rzutów kością do gry i weryfikujemy hipotezę H_0 mówiącą, że kość jest rzetelna, tzn. że każda liczba oczek pojawia się z jednakowym prawdopodobieństwem $\frac{1}{6}$. Standardowy test chi-kwadrat na poziomie istotności 0.01 odrzuca hipotezę zerową, jeżeli wartość statystyki chi-kwadrat przekracza 15.0863 (kwantyl rzędu 0.99 rozkładu chi-kwadrat z pięcioma stopniami swobody). Przypuśćmy, że wykonaliśmy tylko $n = 6$ rzutów. Jest to zbyt mało, żeby asymptotyczne przybliżenie rozkładu chi-kwadrat było zadowalające. Wyznaczyć faktyczny rozmiar testu odrzucamy H_0, jeśli wartość statystyki chi-kwadrat przekroczy 15.0863.

22.76. Zmienna losowa X przyjmuje wartości 1 lub 2 z jednakowym prawdopodobieństwem 0.5. Zmienna losowa Y przyjmuje wartości $1, 2, \ldots, k$. Dysponujemy próbą z łącznego rozkładu prawdopodobieństwa zmiennych losowych X i Y, złożoną z n par obserwacji. Niech n_{ij} oznacza liczbę takich par, dla których zmienna X przyjęła wartość i, zaś Y wartość j ($i = 1, 2;\ j = 1, 2, \ldots, k$). W celu weryfikacji hipotezy o niezależności zmiennych X i Y, czyli hipotezy

$$H_0 : P(X = i, Y = j) = \frac{1}{2} P(Y = j) \text{ dla } i = 1, 2;\ j = 1, 2, \ldots, k,$$

używamy statystyki

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(n_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}/2}, \quad \text{gdzie} \quad \hat{p}_{ij} = \frac{n_{1j} + n_{2j}}{2}.$$

Przy $n \to \infty$, rozkład tej statystyki zmierza do rozkładu chi-kwadrat. Wyznaczyć liczbę stopni swobody rozkładu granicznego.

22.77. Dysponujemy danymi o liczbie szkód zgłoszonych przez klientów $1, 2, \ldots, k$ w ciągu n lat. Niech $S_i(n)$ oznacza sumaryczną liczbę szkód dla klienta numer i w ciągu n lat. Wiemy, że $S_1(n), \ldots, S_k(n)$ są niezależnymi zmiennymi o rozkładzie Poissona. Mamy też pewne przypuszczenia dotyczące intensywności pojawiania się szkód, czyli wartości oczekiwanych tych zmiennych, ale nie jesteśmy ich pewni. Weryfikujemy hipotezę statystyczną

$$H_0 : \forall\ i = 1, \ldots, k \text{ zmienna losowa } S_i(n) \text{ ma rozkład Poissona z parametrem } n\lambda_i.$$

Hipotetyczne intensywności $\lambda_1, \ldots, \lambda_k$ są danymi, ustalonymi liczbami dodatnimi. Używamy pewnej odmiany testu chi-kwadrat: obliczamy statystykę

$$\chi^2 = \sum_{i=1}^{k} \frac{(S_i(n) - n\lambda_i)^2}{n\lambda_i}.$$

Jaki jest rozkład graniczny tej statystyki, jeśli H_0 jest prawdziwa i $n \to \infty$?

22.78. W pewnym portfelu ubezpieczeń samochodowych liczącym 450 ubezpieczonych odnotowano, że w ciągu roku ubezpieczenia 40 ubezpieczonych kobiet uczestniczyło w wypadku, 30 ubezpieczonych mężczyzn uczestniczyło w wypadku, 160 ubezpieczonych kobiet nie uczestniczyło w wypadku oraz 220 ubezpieczonych mężczyzn nie uczestniczyło w wypadku. Ile wynosi wartość statystyki testu chi-kwadrat niezależności i czy (na poziomie istotności 0.05) istnieją podstawy do odrzucenia hipotezy o niezależności wystąpienia wypadku od płci ubezpieczonego?

Inne zagadnienia testowania

22.79. Obserwujemy n niezależnych zmiennych losowych $X_1, \ldots, X_n$ o tym samym rozkładzie o gęstości

$$f_\theta(x) = \begin{cases} \frac{2x}{\theta}, & \text{dla } x \in (0, \theta), \\ 0, & \text{poza tym,} \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Rozważmy test jednostajnie najmocniejszy dla weryfikacji hipotezy $H_0 : \theta = 1$ przeciw $H_1 : \theta > 1$ na poziomie istotności 0.1. Jak najmniej liczną próbą należy dysponować, aby moc otrzymanego testu przy alternatywie $\theta = 2/3$ była nie mniejsza niż 0.9?

22.80. Niezależne zmienne losowe $X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3, Y_4, Y_5$ są takie, że X_1, X_2, X_3, X_4 mają ten sam rozkład o dystrybuancie F_{μ_1}, a Y_1, Y_2, Y_3, Y_4, Y_5 mają ten sam rozkład o dystrybuancie F_{μ_2}. Dystrybuanta F_μ spełnia warunek

$$F_\mu(x) = F(x - \mu)$$

dla pewnej ustalonej, nieznanej, ciągłej, ściśle rosnącej dystrybuanty F. Weryfikujemy hipotezę $H_0 : \mu_1 = \mu_2$ przy alternatywie $H_1 : \mu_1 \neq \mu_2$ stosując test o obszarze krytycznym

$$K = \{S : S \leq 13 \vee S \geq 27\},$$

gdzie S jest sumą rang zmiennych losowych X_1, X_2, X_3, X_4 w próbie złożonej ze wszystkich obserwacji ustawionych w ciąg rosnący. Wyznaczyć rozmiar testu.

22.81. Obserwujemy niezależne zmienne losowe $X_1, X_2, X_3, Y_1, Y_2, Y_3, Y_4$. Zmienne losowe X_1, X_2, X_3 mają ten sam rozkład o dystrybuancie F_{μ_1}, a zmienne losowe Y_1, Y_2, Y_3, Y_4 mają ten sam rozkład o dystrybuancie F_{μ_2}. Dystrybuanta F_μ spełnia warunek $F_\mu = F(x-\mu)$ dla pewnej ustalonej, nieznanej, ciągłej, ściśle rosnącej dystrybuanty F. Weryfikujemy hipotezę $H_0 : \mu_1 = \mu_2$ przy alternatywie $H_1 : \mu_1 > \mu_2$ stosując test o obszarze krytycznym $K = \{S : S > 13\}$, gdzie S jest sumą rang zmiennych losowych X_1, X_2, X_3 w próbie złożonej ze wszystkich obserwacji ustawionych w ciąg rosnący. Wyznaczyć rozmiar testu.

22.82. Na podstawie próby $X_1, \ldots, X_{20}$ testowano hipotezę $H_0 : \sigma^2 = 1$ przy alternatywie $H_1 : \sigma^2 > 1$, gdzie σ^2 jest parametrem odpowiadającym za wariancję zmiennej losowej X_i, za pomocą testu o obszarze krytycznym $K = \left\{\sum_{i=1}^{20} X_i^2 > t\right\}$. Wiadomo, że zmienne losowe mają rozkład zadany gęstością

$$f_\theta(x) = \theta|x|e^{-\theta x^2}, \quad \text{dla} \quad x \in R,$$

gdzie $\theta > 0$ jest nieznanym parametrem. Wyznaczyć wartość krytyczną t tak, by test miał poziom istotności 0.05.

22.83. Niech $X_1, \ldots, X_{10}$ będzie próbą z rozkładu ciągłego o ściśle rosnącej dystrybuancie F. Hipotezę H_0 : *F jest dystrybuantą rozkładu symetrycznego*, tzn. takiego że dla każdego x zachodzi: $F(-x) = 1 - F(x)$, odrzucamy, gdy spełniona jest nierówność $K > 7$ lub $K < 3$, gdzie K jest liczbą elementów w próbie losowej $X_1, \ldots, X_{10}$ o wartościach większych od zera. Wyznaczyć rozmiar testu.

22.84. Mamy trzy niezależne, dziesięcioelementowe próby proste pobrane z trzech populacji normalnych: $X_{i,1}, \ldots, X_{i,10} \sim N(\mu_i, \sigma^2)$, $i = 1, 2, 3$ o tej samej (nieznanej) wariancji σ^2. W każdym z trzech przypadków policzono średnią $\bar{X}_i = \frac{1}{10}\sum_{i=1}^{10} X_{i,j}$ oraz wariancję z próby $S_i^2 = \frac{1}{9}\sum_{i=1}^{10}(X_{i,j} - \bar{X}_i)^2$. Uzyskano następujące wyniki:

i	1	2	3
$\bar{X}_i$	30	31	32
S_i^2	15/9	25/9	20/9

Przeprowadzono testy F analizy wariancji na poziomie istotności 0.05 dla weryfikacji każdej z następujących hipotez: $H_{12} : \mu_1 = \mu_2$, $H_{23} : \mu_2 = \mu_3$, $H_{13} : \mu_1 = \mu_3$, $H_{123} : \mu_1 = \mu_2 = \mu_3$. Które z hipotez zostaną odrzucone, a które nie?

22.85. Rozważamy model K obiektów obserwowanych przez T okresów czasu, gdzie zarówno K jak i T są dużymi liczbami. Przyjmujemy następujące założenia:

• dla każdego $k = 1, \ldots, K$ oraz $t = 1, \ldots, T$ warunkowy rozkład zmiennej losowej $X_{t,k}$ przy danej wartości zmiennej μ_k, jest rozkładem normalnym o wartości oczekiwanej i wariancji (μ_k, σ^2);

• dla każdego $k = 1, \ldots, K$ rozkład zmiennej losowej μ_k jest rozkładem normalnym o wartości oczekiwanej i wariancji (μ, a^2).

Przyjmijmy typowe oznaczenia dla średnich obiektowych i średniej ogólnej:

$$\bar{X}_k = \frac{1}{T}\sum_{t=1}^{T} X_{t,k},\ k = 1,\dots,K, \quad \text{oraz} \quad \bar{X} = \frac{1}{K}\sum_{k=1}^{K}\bar{X}_k.$$

Międzyobiektową i wewnątrzobiektową sumę kwadratów odchyleń oznaczmy:

$$SSB = \sum_{k=1}^{K}(\bar{X}_k - \bar{X})^2, \quad SSW = \sum_{t=1}^{T}\sum_{k=1}^{K}(X_{t,k} - \bar{X}_k)^2.$$

Wiadomo, że zmienne losowe SSB i SSW są niezależne,

$$E(SSB) = (K-1)\left(a^2 + \frac{\sigma^2}{T}\right), \quad E(SSW) = K(T-1)\sigma^2.$$

Dobrać stałą c tak, by wartość oczekiwana wyrażenia $c \cdot \frac{SSW}{SSB}$ wyniosła $\frac{\sigma^2}{a^2T+\sigma^2}$.

22.86. Rozpatrzmy standardowy model jednokierunkowej analizy wariancji. Niech X_{ij} będą niezależnymi zmiennymi losowymi o rozkładach normalnych ($i = 1,\dots,k$, $j = 1,\dots,n_i$), przy czym $EX_{ij} = \mu_i$ i $D^2X_{ij} = \sigma^2$. Przyjmijmy typowe oznaczenia:

$$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2, \quad SST = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2,$$

gdzie

$$\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij} \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}, \quad n = \sum_{i=1}^{k} n_i.$$

Przy założeniu, że hipoteza o jednorodności jest prawdziwa, czyli że $\mu_1 = \cdots = \mu_k$, obliczyć $E\left(\frac{SSW}{SST}\right)$.

23. Elementy statystyki bayesowskiej

Model bayesowski. Niech $(\mathcal{X}, \mathcal{B}, \{P_\theta,\ \theta \in \Theta\})$ będzie model statystycznyem. Parametr θ traktujemy jako zmienną losową o rozkładzie prawdopodobieństwa Π na zbiorze Θ.

Rozkład Π nazywamy **rozkładem** *a priori* i niech $\pi(\cdot)$ oznacza gęstość rozkładu *a priori*.

Rozkład $P_\theta(\cdot)$ oznaczać będziemy $P(\cdot|\theta)$, a jego gęstość przez $p(\cdot|\theta)$.

Rozkładem *a posteriori* nazywamy rozkład prawdopodobieństwa na zbiorze Θ określony funkcją gęstości.

$$\pi(\theta|x) = \pi(\theta)\frac{p(x|\theta)}{p(x)}, \qquad p(x) = \int_\Theta \pi(\theta)p(x|\theta)d\theta.$$

Estymacja bayesowska.

Niech $L(\cdot,\cdot)$ będzie funkcją straty.

Ryzykiem bayesowskim reguły $\hat{\theta} : \mathcal{X} \to \Theta$ nazywamy

$$\int_\Theta \left[\int_\mathcal{X} L(\hat{\theta}(x), \theta)p(x|\theta)dx\right]\pi(\theta)d\theta = \int_\mathcal{X} \left[\int_\Theta L(\hat{\theta}(x), \theta)\pi(\theta|x)d\theta\right]p(x)dx.$$

Regułę $\hat{\theta}$ nazywamy **estymatorem bayesowskim**, jeżeli minimalizuje ona stratę *a posteriori*, czyli

$$\int_\Theta L(\hat{\theta}(x), \theta)\pi(\theta|x)d\theta.$$

Jeżeli $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$, to estymatorem bayesowskim jest wartość oczekiwana rozkładu *a posteriori*.

Bayesowskie przedziały ufności.

Bayesowskim przedziałem ufności na poziomie ufności $1-\alpha$ nazywamy taki podzbiór $K_\alpha \subset \Theta$, którego prawdopodobieństwo *a posteriori* wynosi $1-\alpha$, tzn.

$$\int_{\theta \in K_\alpha} \pi(\theta|x)d\theta = 1 - \alpha.$$

Bayesowskie testowanie hipotez.

Rozważmy problem weryfikacji hipotezy $H_0 : \theta = \theta_1$ przeciwko $H_1 : \theta = \theta_2$. Mamy następujący zbiór decyzji

$$\mathcal{D} = \{d_1 = \{\theta = \theta_1\}, d_2 = \{\theta = \theta_2\}\}.$$

Przyjmujemy następującą funkcję straty $L(d, \theta)$

	$\theta = \theta_1$	$\theta = \theta_2$
d_1	L_{11}	L_{12}
d_2	L_{21}	L_{22}

Liczby L_{ij} są dodatnie. Zazwyczaj przyjmuje się, że $L_{11} = L_{22} = 0$, czyli że nie ponosimy starty przy podjęciu poprawnej decyzji.

Niech rozkład *a priori* będzie

$$P\{\theta = \theta_1\} = \pi = 1 - P\{\theta = \theta_2\}.$$

Oczekiwana strata *a posteriori* reguły δ wynosi

$$\begin{aligned} \delta(x) = d_1 : &\quad \pi p(x|\theta_1) L_{11} + (1-\pi) p(x|\theta_2) L_{12}, \\ \delta(x) = d_2 : &\quad \pi p(x|\theta_1) L_{21} + (1-\pi) p(x|\theta_2) L_{22}. \end{aligned}$$

Test bayesowski jest regułą bayesowską o następującym obszarze przyjęć decyzji d_2

$$\left\{ x \in \mathcal{X} : \frac{p(x|\theta_2)}{p(x|\theta_1)} > \frac{\pi[L_{21} - L_{11}]}{(1-\pi)[L_{12} - L_{22}]} \right\}.$$

Przykład. *Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $D(\theta)$. Zakładamy, że prawdopodobieństwo θ jest zmienną losową o rozkładzie $U(0,1)$. Znaleźć*

1. *bayesowski estymator parametru θ;*

2. *przedział ufności dla parametru θ;*

3. *test hipotezy $H_0 : \theta = \theta_1$ przeciwko $H_1 : \theta = \theta_2$ $(\theta_1 < \theta_2)$.*

Rozwiązanie. *Jak wiadomo statystyką dostateczną dla wnioskowania o prawdopodobieństwie θ jest $T = \sum_{i=1}^n X_i$, której modelem jest*

$$\left(\{0, 1, \ldots, n\}, \{Bin(n, \theta),\ \theta \in [0,1]\}\right).$$

Rozkładem a priori jest rozkład $U(0,1)$. Powiedzmy, że zaobserwowano $T = t$ sukcesów. Znajdziemy rozkład a posteriori $\pi(\theta|t)$:

$$\pi(\theta|t) = \pi(\theta) \frac{p(t|\theta)}{p(t)},\ p(t) = \int_\Theta p(t|\theta)\pi(\theta) d\theta, \quad \text{dla } t = 0, 1, \ldots, n,$$

gdzie $p(t|\theta) = \binom{n}{t} \theta^t (1-\theta)^{n-t}$. Mamy

$$p(t) = \binom{n}{t} \int_0^1 \theta^t (1-\theta)^{n-t} d\theta = \binom{n}{t} B(t+1, n-t+1) = \frac{1}{n+1}.$$

A zatem

$$\pi(\theta|t) = \frac{\binom{n}{t}\theta^t(1-\theta)^{n-t}\pi(\theta)}{p(t)} = (n+1)\binom{n}{k}\theta^t(1-\theta)^{n-t}.$$

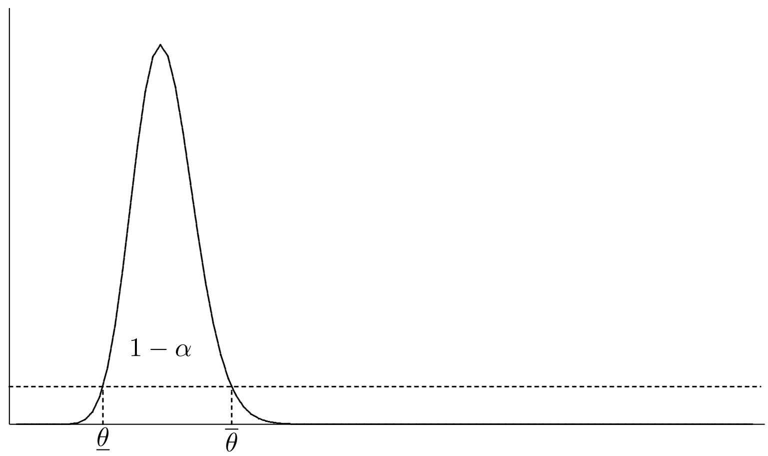
W celu znalezienia bayesowskiego estymatora parametru θ przyjmijmy kwadratową funkcję straty $L(\hat{\theta},\theta) = (\hat{\theta}-\theta)^2$. Wówczas estymatorem bayesowskim jest wartość oczekiwana rozkładu a posteriori:

$$\begin{aligned}\hat{\theta}(t) = \int_0^1 \theta\pi(\theta|t)d\theta &= (n+1)\binom{n}{t}\int_0^1 \theta^{t+1}(1-\theta)^{n-t}d\theta \\ &= (n+1)\binom{n}{t} B\left(t+2, n-t+1\right) \\ &= \frac{t+1}{n+2}\end{aligned}$$

Bayesowski przedział ufności (na poziomie ufności $1-\alpha$) jest zbiorem tych wartości parametru θ, których prawdopodobieństwo a posteriori wynosi $1-\alpha$. Szukamy zatem takich dwóch wartości $\underline{\theta}$ i $\overline{\theta}$, że

$$\begin{cases}\displaystyle\int_{\underline{\theta}}^{\overline{\theta}} \pi(\theta|t)d\theta = 1-\alpha, \\ \pi(\underline{\theta}|t) = \pi(\overline{\theta}|t).\end{cases}$$

Dla danego n oraz zaobserwowanej liczby sukcesów t powyższy układ można rozwiązać numerycznie. Na poniższym wykresie przedstawiona jest gęstość rozkładu a posteriori (dla $n = 10$ i $t = 2$) wraz z zaznaczonym przedziałem ufności dla parametru θ.



Skonstruujemy bayesowski test hipotezy $H_0 : \theta = \theta_1$ przeciwko $H_1 : \theta = \theta_2$ $(\theta_1 < \theta_2)$. Rozważany model statystyczny ma postać

$$\left\{\{0,1,\ldots,n\}, \{Bin\left(n,\theta\right), \theta \in \Theta = \{\theta_1,\theta_2\}\}\right\},$$

a rozkładem a priori jest

$$P(\theta = \theta_1) = \pi = 1 - P(\theta = \theta_2).$$

Mamy następujący zbiór decyzji

$$\mathcal{D} = \{d_1 = \{\theta = \theta_1\}, d_2 = \{\theta = \theta_2\}\}.$$

Przyjmując funkcję straty

	$\theta = \theta_1$	$\theta = \theta_2$
d_1	L_{11}	L_{12}
d_2	L_{21}	L_{22}

otrzymujemy bayesowski obszar przyjęć decyzji d_2

$$\left\{ t \in \{0, 1, \ldots, n\} : \frac{p(t|\theta_2)}{p(t|\theta_1)} > \frac{\pi[L_{21} - L_{11}]}{(1-\pi)[L_{12} - L_{22}]} \right\}.$$

Po dokonaniu odpowiednich przekształceń, otrzymujemy obszar przyjęć decyzji d_2:

$$\left\{ t \in \{0, 1, \ldots, n\} : \left[\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)}\right]^t > \gamma \left[\frac{1-\theta_1}{1-\theta_2}\right]^n \right\}.$$

Tutaj $\gamma = \frac{\pi[L_{21}-L_{11}]}{(1-\pi)[L_{12}-L_{22}]}$.

Mamy więc następującą regułę bayesowską podjęcia decyzji d_2:

$$t > n \frac{\log\left[\frac{1-\theta_1}{1-\theta_2}\right]}{\log\left[\frac{\theta_2}{\theta_1}\right] + \log\left[\frac{1-\theta_1}{1-\theta_2}\right]} + \log\gamma.$$

Niech na przykład $\pi = 0.5$ oraz niech funkcja straty ma postać

	$\theta = \theta_1$	$\theta = \theta_2$
d_1	0	1
d_2	1	0

Reguła bayesowska podjęcia decyzji d_2 ma postać:

$$k > n \frac{\log\left[\frac{1-\theta_1}{1-\theta_2}\right]}{\log\left[\frac{\theta_2}{\theta_1}\right] + \log\left[\frac{1-\theta_1}{1-\theta_2}\right]}.$$

Zauważmy, że przy rozważanej funkcji straty ryzykiem reguły bayesowskiej jest prawdopodobieństwo podjęcia błędnej decyzji.

□

Zadania do samodzielnego rozwiązania

23.1. Niech prawdopodobieństwo θ sukcesu w każdej z niezależnych prób ma rozkład *a priori* $U(0,1)$. Pokazać, że prawdopodobieństwo otrzymania r sukcesów w n próbach jest równe $\frac{1}{n+1}$, $r = 0, 1, \ldots, n$. Pokazać, że jeżeli w n pierwszych próbach otrzymano r sukcesów, to prawdopodobieństwo sukcesu w następnej próbie wynosi $\frac{r+1}{n+2}$.

23.2. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $U(0, \theta)$. Zakładamy, że nieznany parametr θ jest zmienną losową o rozkładzie *a priori* $U(0, 1)$. Wyznaczyć gęstość rozkładu *a posteriori* parametru θ.

23.3. Niech $X_1, X_2, \ldots, X_n$ będzie próbą z rozkładu $N(\theta, \sigma^2)$, w którym wariancja σ^2 jest znana. Niech rozkład *a priori* parametru θ będzie rozkładem $N(\mu, \tau^2)$. Pokazać, że rozkład *a posteriori* jest rozkładem $N(\mu_n, \tau_n^2)$, gdzie

$$\mu_n = \frac{n\bar{X}/\sigma^2 + \mu/\tau^2}{n/\sigma^2 + 1/\tau^2} \qquad \text{oraz} \qquad \tau_n^{-2} = n\sigma^{-2} + \tau^{-2}.$$

23.4. Produkuje się duże serie lamp elektrycznych. Czas życia (trwałość) lampy jest zmienną losową o rozkładzie $E(1/\theta)$. Serię uważa się za udaną, gdy średnia trwałość lamp w serii jest większa od t. Parametr θ w poszczególnych partiach lamp zmienia się zależnie od jakości wolframu, z którego produkuje się włókna żarzenia; zmienność tego parametru opisuje się za pomocą rozkładu *a priori* $G(k, 1)$. Z pewnej serii produkcyjnej wybrano losowo n lamp i w wyniku zmierzenia ich trwałości otrzymano $X_1, X_2, \ldots, X_n$. Obliczyć na tej podstawie prawdopodobieństwo *a posteriori*, że seria ta jest udana.

23.5. Zakłada się, że liczba cząstek wypromieniowanych w czasie T przez pewne radioaktywne źródło ma rozkład $Po(\lambda T)$, gdzie λ jest intensywnością źródła. Zarejestrowano, że w czasie T dwa niezależne źródła o nieznanych intensywnościach λ_1 i λ_2 wypromieniowały odpowiednio X i Y cząstek. Przed dokonaniem tych obserwacji zakłada się, że λ_1 i λ_2 są niezależnymi zmiennymi losowymi o rozkładach $E(1)$. Przy założeniu, że δ jest małe i X oraz Y duże, obliczyć prawdopodobieństwa *a priori* i *a posteriori*, że λ_1/λ_2 ma wartość z przedziału $(1 - \delta, 1 + \delta)$.

23.6. Zmienna losowa N ma rozkład $Po(\lambda)$ z nieznanym $\lambda > 0$. O parametrze λ zakładamy, że podlega rozkładowi *a priori* gamma $G(2, 8)$. Zmienna losowa θ ma rozkład $Bet(1, 2)$. Zmienne N i θ są niezależne oraz zmienne λ i θ są niezależne. Obserwujemy zmienną losową X, która przy znanych wartościach N i θ ma rozkład $Bin(N, \theta)$. Wyznaczyć wartości a i b najlepszego liniowego predyktora zmiennej losowej N, to znaczy liczby a i b minimalizujące wielkość $E(N - aX - b)^2$.

23.7. Niech X będzie zmienną losową o rozkładzie $Ge(\theta)$. Załóżmy, że nieznany parametr θ jest realizacją zmiennej losowej Θ o rozkładzie *a priori* $Pow(3)$. Wyznaczyć wartość bayesowskiego estymatora parametru θ, przy kwadratowej funkcji straty, obliczoną na podstawie zaobserwowanej wartości $X = 0$, czyli $E(\Theta | X = 0)$.

23.8. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Par(0, \theta)$. Zakładamy, że nieznany parametr θ jest zmienną losową o rozkładzie *a priori* $E(1/\lambda)$. Wyznaczyć bayesowski estymator parametru θ przy kwadratowej funkcji straty.

23.9. Niech $X_1, \ldots, X_n$ będzie próbą z rozkładu $Par(\theta, 3)$, gdzie $\theta > 0$ jest nieznanym parametrem. Dla parametru θ zakładamy rozkład *a priori* o gęstości

$$\pi(\theta) = \begin{cases} \frac{\theta}{2}, & \text{dla } \theta \in (0, 2), \\ 0, & \text{poza tym.} \end{cases}$$

Wyznaczyć wartość estymatora bayesowskiego parametru θ przy kwadratowej funkcji straty, jeżeli zaobserwowano próbę spełniającą warunek $\min\{X_1,\ldots,X_n\} = 1$.

23.10. Niech $X_1,\ldots,X_n$ będzie próbą z rozkładu $E\,(1/\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Dla parametru θ zakładamy rozkład *a priori* $G\,(2,1/3)$. Estymujemy parametr θ przy funkcji straty postaci

$$L(\theta,a) = e^{(\theta-a)} - (\theta-a) - 1.$$

Wyznaczyć estymator bayesowski a parametru θ.

23.11. Niech $X_1,\ldots,X_n$, gdzie $n > 1$, będzie próbą z rozkładu $Wei\,(2,1/\theta)$, gdzie $\theta > 0$ jest nieznanym parametrem. Zakładamy, że parametr θ ma rozkład *a priori* $G\,(\alpha,1/\beta)$. Wyznaczyć estymator bayesowski $\hat{\theta}$ parametru θ przy funkcji straty Esschera $L(\theta,\hat{\theta}) = e^{c\theta}(\theta-\hat{\theta})^2$, gdzie $c \neq 0$ jest ustaloną liczbą.

23.12. Niech $X_1,X_2,\ldots,X_n$ będzie próbą z rozkładu $N\,(\theta_1,\theta_2)$ o nieznanych parametrach θ_1 i θ_2 i niech rozkład *a priori* parametru $\boldsymbol{\theta} = (\theta_1,\theta_2)$ ma „gęstość”

$$\pi(\boldsymbol{\theta}) \propto \frac{1}{\theta_2} \qquad (-\infty < \theta_1 < \infty,\ \theta_2 > 0).$$

Pokazać, że rozkład *a posteriori* parametru $\boldsymbol{\theta}$ jest taki, że wielkość $\sqrt{n(n-1)}(\theta_1 - \overline{X})/\sqrt{\mathrm{var}X}$ ma rozkład $t\,(n-1)$. Wyznaczyć najkrótszy bayesowski przedział ufności na poziomie ufności 0.95.

23.13. W każdym z niezależnych doświadczeń z prawdopodobieństwem θ może pojawić się pewne zdarzenie E; prawdopodobieństwo θ nie jest znane, ale przypuszcza się, że jest małe. W celu zorientowania się co do wielkości θ przeprowadza się serię doświadczeń, którą kontynuuje się dopóty, dopóki nie pojawi się E. Przypuśćmy, że zdarzenie po raz pierwszy pojawia się w n-tym doświadczeniu. Niech θ ma rozkład *a priori* o gęstości $m(1-\theta)^{m-1}$ $(0 < \theta < 1)$. Pokazać, że najkrótszy bayesowski przedział ufności na poziomie ufności $1-\alpha$ jest dla dużych $m+n$ w dobrym przybliżeniu równy

$$\left[\frac{c_1}{(m+n)}, \frac{c_2}{(m+n)}\right],$$

gdzie $c_1e^{-c_1} = c_2e^{-c_2}$, $e^{-c_1} - e^{-c_2} = 1-\alpha$.

23.14. Niech $X_1,\ldots,X_8$ będzie próbą z rozkładu $N\,(\theta,1)$. Nieznany parametr θ jest z kolei zmienną losową o rozkładzie $N\,(0,1)$. Wyznaczyć bayesowski przedział ufności dla parametru θ, to znaczy przedział $[a,b]$, gdzie

$$a = a(X_1,\ldots,X_8),\ b = b(X_1,\ldots,X_8),$$

taki, że $P(\theta < a|X_1,\ldots,X_8) = 0.05 = P(\theta > b|X_1,\ldots,X_8)$.

23.15. Niech X_1,X_2,X_3,X_4 będzie próbą z rozkładu $U\,(0,\theta)$. Zakładamy, że nieznany parametr θ jest zmienną losową o rozkładzie *a priori* $G\,(5,1/2)$. Hipotezę $H_0 : \theta \leq 3$ przy alternatywie $H_1 : \theta > 3$ odrzucamy dla tych wartości $\{x_1,x_2,x_3,x_4\}$, dla których prawdopodobieństwo *a posteriori* zbioru $\{\theta : \theta > 3\}$ jest większe niż 0.5. Wyznaczyć obszar krytyczny testu.

24. Podejmowanie decyzji statystycznych

Model statystyczny

$$\{\mathcal{X}, \{P_\theta : \theta \in \Theta\}\}$$

1. Zbiór obserwacji $\mathcal{X}$

2. Zbiór stanów natury Θ

3. Zbiór decyzji $\mathcal{D}$

4. Funkcja straty $L(d, \theta) : \mathcal{D} \times \Theta \to \mathbb{R}$

5. Reguła decyzyjna $\delta : \mathcal{X} \to \mathcal{D}$

6. Ryzyko reguły δ: $R_\delta(\theta) = E_\theta\{L(\delta(X), \theta)\}$

Zadanie: znaleźć regułę δ „optymalizującą” ryzyko

Optymalizacja:

1. jednostajna minimalizacja ryzyka

2. zasada minimaksu

3. reguła Bayesa

Estymacja

1. Zbiór decyzji $\mathcal{D} = \Theta = \mathbb{R}$

2. Funkcja straty $L(d, \theta) = (d - \theta)^2$

3. Reguła decyzyjna δ: estymator parametru θ

4. Ryzyko reguły δ: błąd średniokwadratowy

Jeżeli ograniczymy się do takich reguł δ, że

$$E_\theta \delta(X) = \theta, \qquad \forall \theta \in \Theta$$

to reguła **jednostajnie minimalizująca ryzyko** jest ENMW.

Weryfikacja hipotez

1. Zbiór decyzji

$$\mathcal{D} = \{d_1 = \{\theta \in \Theta_0\}, d_2 = \{\theta \notin \Theta_0\}\}$$

2. Funkcja straty $L(d, \theta)$

	$\theta \in \Theta_0$	$\theta \notin \Theta_0$
d_1	0	1
d_2	1	0

3. Reguła decyzyjna δ: test ϕ

4. Ryzyko reguły δ: prawdopodobieństwo błędnego wnioskowania

Jeżeli ograniczymy się do takich reguł δ, że

$$E_\theta \delta(X) \leq \alpha, \qquad \forall \theta \in \Theta_0$$

to reguła **jednostajnie minimalizująca ryzyko** jest testem jednostajnie najmocniejszym.

Optymalizacja

1. Jednostajna minimalizacja ryzyka

Znaleźć taką regułę δ, że jeżeli δ' jest jakąkolowiek inną regułą, to

$$R_\delta(\theta) \leq R_{\delta'}(\theta) \qquad \forall \theta \in \Theta$$

2. Zasada minimaksu

Znaleźć taką regułę δ, że jeżeli δ' jest jakąkolowiek inną regułą, to

$$\max_\theta R_\delta(\theta) \leq \max_\theta R_{\delta'}(\theta)$$

3. Zasada Bayesa

Znaleźć taką regułę δ, że jeżeli δ' jest jakąkolowiek inną regułą, to

$$\int_\Theta R_\delta(\theta)\Pi(d\theta) \leq \int_\Theta R_{\delta'}(\theta)\Pi(d\theta)$$

gdzie Π jest taką miarą na zbiorze Θ, że $\int_\Theta \Pi(d\theta) = 1$

Zadania do samodzielnego rozwiązania

24.1. Sprzedawca jest zainteresowany maksymalizacją zysku ze sprzedaży pewnego, łatwo psującego się towaru. Zakup jednego opakowania kosztuje go \$20, zaś sprzedaje je za \$50. Po jednym dniu niesprzedany towar należy wyrzucić jako niezdatny do spożycia. W ciągu stu dni sprzedawca zaobserwował, że klienci kupowali następujące ilości towaru:

Ilość towaru	10	11	12	13
Ilość dni	15	20	40	25

Ile towaru powinien sprzedawca zamówić, by mógł oczekiwać największego zysku?

24.2. Przedsiębiorca zainteresowany jest zatrudnieniem w swojej firmie samochodowej kilku mechaników. Na kolejny rok przewidywane są następujące ilości godzin pracy dla mechaników:

Ilość godzin	10000	12000	14000	16000
Prawdopodobieństwo	.2	.3	.4	.1

Planowana zapłata za jedną godzinę pracy mechanika wynosi \$9 zaś spodziewany zysk z jednej godziny pracy — \$16. Mechanik może pracować 40 godzin w tygodniu oraz ma prawo do dwutygodniowego urlopu. Na podstawie podanych informacji podać optymalną liczbę mechaników, która powinna być zatrudniona.

24.3. Przedsiębiorstwo lotnicze otrzymuje propozycję zakupienia dziesięciu używanych samolotów, przy czym wszystkie są mniej więcej w takim samym stanie technicznym. Pewna nieznana liczba θ tych samolotów może latać 1000 godzin bez potrzeby większych napraw i każdy z takich samolotów przyniesie zysk w wysokości $1000p$; każdy z pozostałych samolotów dozna jakiegoś poważniejszego uszkodzenia w ciągu pierwszych 1000 godzin eksploatacji, co przyniesie przedsiębiorstwu stratę w wysokości $1000q$. Należy zdecydować, czy przyjąć czy odrzucić ofertę. Przed podjęciem decyzji przedsiębiorstwo może otrzymać pewne dalsze informacje, a mianowicie za cenę $1000r$ może otrzymać jeden samolot do prób, eksploatować go przez 1000 godzin i uzależnić swoją decyzję od tego, czy ten samolot latał 1000 godzin bez awarii, czy nie. Podać wszystkie reguły decyzyjne i ich ryzyka. Wybrać optymalną regułę decyzyjną.

24.4. Pewna osoba zamierza sprzedawać napój *Migotka* w trakcie meczu piłki nożnej i musi zawczasu zdecydować o wielkości zamówienia. Przypuśćmy, że na każdym sprzedawanym w trakcie gry litrze zyskuje m, zaś traci c na każdym zamówionym, lecz nie sprzedanym. Załóżmy, że popyt na *Migotkę* w trakcie gry, mierzony w litrach, jest ciągłą zmienną losową X o funkcji gęstości f i dystrybuancie F. Przy jakiej wysokości zamówienia oczekiwany zysk będzie maksymalny?

24.5. Rozważmy zagadnienie decyzyjne, w którym $\Theta = \{\theta_1, \theta_2\}$ oraz $\mathcal{D} = \{d_1, d_2, d_3\}$, funkcja straty określona jest w następujący sposób:

	d_1	d_2	d_3
θ_1	0	10	3
θ_2	10	0	3

Niech obserwowana zmienna losowa X ma rozkład

$$P(X = 1|\theta = \theta_1) = 0.75;\ P(X = 0|\theta = \theta_1) = 0.25;$$
$$P(X = 1|\theta = \theta_2) = 0.25;\ P(X = 0|\theta = \theta_2) = 0.75$$

Zmienna losowa może być obserwowana n krotnie. Niech $P\{\theta = \theta_1\} = p = 1 - P\{\theta = \theta_2\}$ $(0 \le p \le 1)$. Określić bayesowską regułę decyzyjną na podstawie obserwacji $X_1, \ldots, X_n$ i naszkicować jej ryzyko jako funkcję prawdopodobieństwa p. Zakładając, że koszt każdej obserwacji wynosi c wyznaczyć optymalną wielkość n. Wyznaczyć optymalną wielkość próby, gdy koszt obserwacji o wartości 1 wynosi c_1, zaś koszt obserwacji o wartości 0 wynosi 0.

24.6. Niech $X_1, X_2, \ldots, X_n$ będzie próbką losową z rozkładu $N(\mu, \sigma^2)$ o nieznanych μ i σ^2 i niech rozkład *a priori* tych parametrów będzie taki, że μ i $\log\sigma$ są niezależne, a brzegowy rozkład każdej z tych wielkości jest jednostajny. Pokazać, że estymatorem bayesowskim parametru μ przy kwadratowej funkcji strat $L(\hat{\mu}, \mu) = (\hat{\mu} - \mu)^2$ jest $\bar{X}$. Ile wynosi ryzyko bayesowskie tego estymatora?

24.7. Niech X będzie zmienną losową o rozkładzie $N(\theta, 1)$ i niech strata spowodowana szacowaniem θ za pomocą $\hat{\theta}(X)$ ma postać

$$a\{\hat{\theta}(X) - \theta\}, \quad \text{gdy} \quad \hat{\theta}(X) \ge \theta,$$
$$b\{\theta - \hat{\theta}(X)\}, \quad \text{gdy} \quad \hat{\theta}(X) < \theta,$$

gdzie $a > 0$ oraz $b > 0$. Pokazać, że ryzyko estymatora $\hat{\theta}_k$ postaci

$$\hat{\theta}_k(X) = X - k$$

wyraża się wzorem

$$(a+b)\{\varphi(k) + k\Phi(k)\} - ka,$$

gdzie φ i Φ są odpowiednio gęstością i dystrybuantą rozkładu $N(0,1)$. Następnie pokazać, że w klasie $\{\hat{\theta}_k : k \text{ rzeczywiste}\}$ istnieje estymator o jednostajnie minimalnym ryzyku i że to minimalne ryzyko jest równe

$$(a+b)\varphi\left[\Phi^{-1}\left(\frac{a}{a+b}\right)\right].$$

24.8. Bochenek chleba musi ważyć co najmniej w gramów. W pewnej piekarni ciężar chleba w dużym wypieku jest zmienną losową o rozkładzie $N(\mu, 1/\tau)$. Parametr μ jest wielkością regulowaną, natomiast τ zmienia się od wypieku do wypieku według rozkładu chi–kwadrat o ν stopniach swobody. Koszt produkcji bochenka chleba, gdy średni ciężar jest równy μ wynosi $k + l\mu$, a cena zbytu bochenka o prawidłowym ciężarze jest równa m. Bochenki o zbyt małym ciężarze nie przynoszą zysku. Pokazać, że średni zysk na bochenku, gdy wypiek ustawiony jest na wielkość μ, wynosi

$$m\Psi\{(\mu - w)\sqrt{n}\} - (k + l\mu),$$

gdzie Ψ jest rozkładem t o ν stopniach swobody. Wyznaczyć na tej podstawie najlepszą wartość μ.

24.9. W procesie mierzenia zawartości RNA w pewnych komórkach pojawia się trudność związana z tym, że dwie komórki mogą znajdować się tak blisko siebie, iż stają się nierozróżnialne. Wykonuje się dwa niezależne pomiary X_1 i X_2, przy czym każdy z nich może odnosić się do jednej lub dwóch (niezależnych) komórek. Należy zdecydować, czy dana para pomiarów dotyczy dwóch pojedynczych komórek, jednej pojedynczej komórki i jednej pary, czy też dwóch par komórek.

Przypuśćmy, iż wiadomo, że zawartość pojedynczej komórki, mierzona w odpowiednich jednostkach, ma rozkład o gęstości xe^{-x} $(x \geq 0)$ i że prawdopodobieństwo *a priori* tego, że pomiar dotyczy dwóch komórek zamiast jednej wynosi π. Przypuśćmy również, że strata jest równa zeru, gdy decyzja jest prawidłowa i jest równa jedności, gdy decyzja jest błędna. Naszkicować na płaszczyźnie (x_1, x_2) rozwiązanie bayesowskie tego zagadnienia.

24.10. Niech $X_1, X_2, \ldots, X_n$ będą kolejnymi pomiarami intensywności sygnału radiowego rejestrowanymi za pomocą pewnego odbiornika. Jeżeli był nadawany pewien sygnał, to wynik pomiaru ma postać $X_j = a_j + \varepsilon_j$, gdzie $a_1, a_2, \ldots, a_n$ są znane, a losowe zakłócenia $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ są realizacją wielowymiarowej zmiennej losowej o rozkładzie normalnym ze średnią równą zeru i z macierzą kowariancji V. Jeżeli sygnał nie był nadawany, to $X_j = \varepsilon_j$ $(j = 1, 2, \ldots, n)$. Należy zdecydować, czy sygnał został rzeczywiście nadany. Prawdopodobieństwo *a priori* tego zdarzenia wynosi p. Straty związane z nieprawidłowym orzeczeniem, że sygnał został lub nie został nadany, są równe odpowiednio $L-1$ i L_2. Wyznaczyć regułę decyzyjną realizującą minimum oczekiwanych strat.

Przypuśćmy, że decyzję podejmuje się zgodnie z tą optymalną regułą i że $pL_2 = (1-p)L_1$. Jak wygląda optymalny ciąg $\{a_i\}$ sygnałów, jeżeli analiza mocy transmisji prowadzi do ograniczenia $\sum a_i^2 = 1$?

24.11. W celu podjęcia decyzji, którą z dwóch odmian pszenicy wprowadzić do masowej produkcji, wykonuje się eksperyment polegający na tym, że każdą z odmian bada się na n poletkach doświadczalnych. Obserwowane plony X_{ij} $(i = 1, 2, j = 1, 2, \ldots, n)$ są niezależnymi zmiennymi losowymi o rozkładzie normalnym ze średnimi μ_i i wariancją σ^2, przy czym rozkład *a priori* parametru (μ_1, μ_2) jest taki, że μ_1 i μ_2 są niezależne i mają rozkłady brzegowe normalne o średnich równych odpowiednio α_1 i α_2 i wariancji σ_0^2. Wyznaczyć rozkład *a posteriori* (μ_1, μ_2) przy założeniu, że wariancja σ^2 jest znana. Pokazać, że gdy funkcje strat mają postać $L_i = -k\mu_i$ $(i = 1, 2)$, wtedy oczekiwane ryzyko osiąga minimum dla decyzji: wybrać odmianę 1, gdy $X_{1\cdot} - X_{2\cdot} > c$ lub odmianę 2, gdy $X_{1\cdot} - X_{2\cdot} < c$, gdzie $nX_{i\cdot} = \sum_j X_{ij}$ oraz $c = (\alpha_2 - \alpha_1)\sigma^2/(n\sigma_0^2)$.

24.12. Niech (X_n) będzie ciągiem niezależnych zmiennych losowych o rozkładzie Poissona z jednakową średnią θ. W celu oszacowania θ obserwuje się ten ciąg sekwencyjnie. Zakładamy, że: początkowa wiedza o θ opisana jest za pomocą niewłaściwego rozkładu *a priori* o stałej gęstości na przedziale $(0, \infty)$; strata spowodowana oszacowaniem θ przez $\hat{\theta}$ wynosi $(\hat{\theta}-\theta)^2$; koszt każdej obserwacji jest równy c. Udowodnić, że optymalna bayesowska sekwencyjna reguła decyzyjna ma postać: przerwać obserwacje, gdy tylko $\sum_{i=1}^n X_i + 1 < cn^2(n+1)$ i za oszacowanie θ przyjąć $\sum_{i=1}^n (X_i+1)/n$.

24.13. Niech $X_1, X_2, \ldots, X_{10}$ będzie próbką losową z rozkładu normalnego $N(0, \theta)$. Wyznaczyć rozkład *a posteriori* parametru θ, gdy rozkład *a priori* tego parametru jest taki sam jak rozkład zmiennej losowej $\theta_0\nu_0/\chi^2(\nu_0)$, gdzie θ_0 i ν_0 są dodatnimi stałymi.

Wyznaczyć najlepszy bayesowski estymator punktowy $\hat{\theta}$, gdy funkcja straty ma postać $(\hat{\theta}-\theta)^2$. Wykazać, że gdy $\nu_0 \to 0$ przy stałym θ_0, wtedy

$$\hat{\theta} \to \frac{1}{8}\sum X_i^2.$$

Porównując estymator $\frac{1}{8}\sum X_i^2$ z estymatorem różniącym się od niego pewnym mnożnikiem pokazać, że jest to estymator niedopuszczalny.

25. Statystyki próbkowe

Niech $X_1, \ldots, X_n$ będzie próbą i niech $X_{1:n} \le \cdots \le X_{n:n}$ będzie próbą uporządkowaną.

Kwantyl próbkowy. Kwantylem próbkowym rzędu q nazywamy $X_{\lfloor qn \rfloor:n}$.

Mediana próbkowa. Medianą próbkową nazywamy kwantyl próbkowy rzędu 0.5, czyli $X_{\lfloor 0.5n \rfloor:n}$. Medianę próbkową oznaczamy przez $med\{X_1, \ldots, X_n\}$.

Średnia arytmetyczna.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Suma kwadratów odchyleń.

$$\mathrm{var} X = \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Niech $(X_1, Y_1), \ldots, (X_n, Y_n)$ będzie próbą dwuwymiarową.

Suma iloczynów odchyleń.

$$\mathrm{cov}(X, Y) = \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}).$$

Bibliografia

ABT S. 1972: Matematyczno-statystyczne podstawy analizy rynku, PWE, Warszawa.

BOX G. E. P., HUNTER W. G., HUNTER J. S. 1978: Statistics for experimenters, Wiley.

BOX G. E. P., JENKINS G. M. 1983: Analiza szeregów czasowych, PWN, Warszawa.

BOŻYK Z., RUDZKI W. 1977: Metody statystyczne w badaniu jakości produktów żywnościowych i chemicznych, WNT, Warszawa.

BRANDT S. 1976: Metody statystyczne i obliczeniowe analizy danych, PWN, Warszawa.

BROSS J. B. 1965: Jak podejmować decyzje, PWN, Warszawa.

CRAMÉR H. 1957: Metody matematyczne w statystyce, PWN, Warszawa.

DĄBKOWSKI J. 1992:, Statgraphics, Komputerowa Oficyna Wydawnicza „HELP", Warszawa.

DĄBROWSKI A., GNOT S., MICHALSKI A., SRZEDNICKA J. 1994: Statystyka, 15 godzin z pakietem Statgraphics, Wydawnictwo Akademii Rolniczej, Wrocław.

DITTMANN P. 1998: Metody prognozowania sprzedaży w przedsiębiorstwie, wydanie 3, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.

DOMAŃSKI C. 1979: Statystyczne testy nieparametryczne, PWE, Warszawa.

DRAPER N. R., SMITH H. 1966: Applied Regression Analysis, Wiley.

ELANDT R. 1964: Statystyka matematyczna w zastosowaniu do doświadczalnictwa rolniczego, PWN, Warszawa.

FISZ M. 1958: Rachunek prawdopodobieństwa i statystyka matematyczna, PWN, Warszawa.

GAWĘCKI J., WAGNER W. 1984: Podstawy metodologii badań doświadczalnych w nauce o żywieniu i żywności, PWN, Warszawa.

GNANADESIKAN R. 1982: Statistical data analysis, Proceedings of Symposia in Applied Mathematics, Vol 28, American Mathematical Society, Providence, Rhode Island.

GOLDBERGER A. S. 1975: Teoria ekonometrii, PWE, Warszawa.

GÓRCZYŃSKI J. 1993: Podstawy statystyki z przykładami w arkuszach kalkulacyjnych, Fundacja „Rozwój SGGW", Warszawa.

GREŃ J. 1984: Statystyka matematyczna. Modele i zadania, PWN, Warszawa

GUPTA R. P. 1975: Applied Statistics, North-Holland.

HALD A. 1957: Statistical Theory with Engineering Applications, Wiley.

HELLWIG Z. 1975: Elementy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa.

JAJUGA K. 1993: Statystyczna analiza wielowymiarowa, Biblioteka ekonometryczna, PWN, Warszawa.

JAWORSKI S., KOBUS P., KOZIOŁ D., PIETRZYKOWSKI R., ZIELIŃSKI W. 2001: Zbiór zadań z podstaw statystyki i ekonometrii, WSEI Warszawa

JAWORSKI S., KOBUS P., KOZIOŁ D., PIETRZYKOWSKI R., ZIELIŃSKI W. 2005: Zbiór zadań z podstaw ekonometrii, WSEI Warszawa

JÓŹWIAK J., PODGÓRSKI J. 1993: Statystyka od podstaw, PWE, Warszawa.

KALA R. 1996: Elementy wnioskowania parametrycznego dla przyrodników, Wydawnictwo Akademii Rolniczej w Poznaniu, Poznań.

KASSYK-ROKICKA H. 1997: Statystyka. Zbiór zadań, PWE, Warszawa.

KASSYK-ROKICKA H. 1998: Mierniki statystyczne, PWE, Warszawa.

KLEIN L. R. 1965: Wstęp do ekonometrii, PWE, Warszawa.

KOBUS P., PIETRZYKOWSKI R., ZIELIŃSKI W. 2000: Statystyka z pakietem STATISTICA, wydanie II, Fundacja „Rozwój SGGW", Warszawa.

KRISHNAIAH P. R. 1980: Handbook of Statistics, vol. 1 – Analysis of Variance, North-Holland.

KRYSICKI W., BARTOS J., DYCZKA W., KRÓLIKOWSKA K., WASILEWSKI M. 1994: Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach, Część II. Statystyka matematyczna, wyd. 2, PWN, Warszawa.

KRZYŚKO M. 1994: Statystyka matematyczna, UAM Poznań

KRZYŚKO M. 1997: Statystyka matematyczna, Część II, UAM Poznań

MICHALSKI T. 2000: Statystyka, WSiP, Warszawa.

NIEMIRO W. 1999: Rachunek prawdopodobieństwa i statystyka matematyczna, Szkoła Nauk Ścisłych, Warszawa

NOWAK E. 1994: Zarys metod ekonometrii, Zbiór zadań, PWN, Warszawa.

LANGE O. 1958: Wstęp do ekonometrii, PWN, Warszawa.

LEVIN R. I. 1987: Statistics for managament, Prentice-Hall.

LUSZNIEWICZ A., SŁABY T. 1997: Statystyka stosowana, PWE, Warszawa.

OKTABA W. 1982: Elementy statystyki matematycznej i metodyka doświadczalnictwa, PWN, Warszawa.

OKTABA W. 1982: Metody statystyki matematycznej w doświadczalnictwie, PWN, Warszawa.

PAWŁOWSKI Z. 1978: Ekonometria, PWN, Warszawa.

PAWŁOWSKI Z. 1981: Statystyka matematyczna, PWN, Warszawa.

PLATT C. 1981: Problemy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa.

PODGÓRSKI J. 1993: Statystyka z komputerem, Pakiet STATGRAPHICS, „DIAGRAM", Warszawa.

RAO C. R. 1982: Modele liniowe statystyki matematycznej, PWN, Warszawa.

RAO C. R. 1994: Statystyka i prawda, PWN, Warszawa.

REICHMANN W. J. 1968: Drogi i bezdroża statystyki, PWN, Warszawa.

RÓSZKIEWICZ M. 1993: Statystyka – kurs podstawowy, SGH, Warszawa.

SCHEFFÉ H. 1959: The Analysis of Variance, Wiley.

SEARLE E. S. R. 1971: Linear models, Wiley.

SEARLE E. S. R. 1987: Linear Models for Unbalanced Data, Wiley.

SEBER G. A. F. 1977: Linear Regression Analysis, Wiley.

SRIVASTAVA J. N. 1975: A Survey of Statistical Design and Linear Models, North-Holland.

SILVEY S. D. 1978: Wnioskowanie statystyczne, PWN, Warszawa.

SOBCZAK M. 1997: Statystyka, PWE, Warszawa.

STANISZ A. 2001: Przystępny kurs statystyki w oparciu o program STATISTICA PL na przykładach z medycyny, Tom I, wydanie II, StatSoft Polska, Kraków.

STANISZ A. 2000: Przystępny kurs statystyki w oparciu o program STATISTICA PL na przykładach z medycyny, Tom II, StatSoft Polska, Kraków.

STANISZ T. 1993: Funkcje jednej zmiennej w badaniach ekonomicznych, Biblioteka ekonometryczna, PWN, Warszawa.

TUKEY J. W. 1977: Exploratory data analysis, Addison-Wesley Publishing Company.

WELFE W. 1977: Ekonometryczne modele rynku, tom 1, metody ekonometryczne, PWE, Warszawa.

WELFE W. 1990: Gospodarki Polski w okresie transformacji. Zasady modelowania ekonometrycznego, PWE, Warszawa

WÓJCIK A. R. 1987: Statystyka matematyczna z elementami rachunku prawdopodobieństwa i statystyki opisowej, SGGW, Warszawa.

WÓJCIK A. R., LAUDAŃSKI Z. 1989: Planowanie i wnioskowanie statystyczne w doświadczalnictwie, PWN, Warszawa.

ZIELIŃSKI R. 1976: Rachunek prawdopodobieństwa z elementami statystyki matematycznej, wyd. II, WSiP, Warszawa.

ZIELIŃSKI R. 1990: Siedem wykładów wprowadzających do statystyki matematycznej, http://www.impan.pl/~rziel/Books.html

ZIELIŃSKI R., ZIELIŃSKI W. 1987: Podręczne tablice statystyczne, WNT, Warszawa.

ZIELIŃSKI R., ZIELIŃSKI W. 1990: Tablice statystyczne, PWN, Warszawa.

ZIELIŃSKI T. 1999: Jak pokochać statystykę, czyli STATISTICA do poduszki, StatSoft Polska, Kraków

ZIELIŃSKI W. 1998: Analiza regresji, Fundacja „Rozwój SGGW", Warszawa

ZIELIŃSKI W. 1999: Wybrane testy statystyczne, wydanie II poprawione, Fundacja „Rozwój SGGW", Warszawa

ZIELIŃSKI W. 2000: Tablice statystyczne, wydanie IV poprawione, Fundacja „Rozwój SGGW", Warszawa

ZIELIŃSKI W. 2007: Teoretyczne podstawy ekonometrycznych jednorównaniowych modeli liniowych, Wydawnictwo SGGW

ZIELIŃSKI W. 2010: Estymacja wskaźnika struktury, Wydawnictwo SGGW