Zadanie 1.

Niech $X_1, X_2, \ldots, X_n, n \geq 4$ będą niezależnymi zmiennymi losowymi o rozkładzie wykładniczym o gęstości

$$f(x) = \begin{cases} \theta e^{-\theta x} & \text{dla } x \geq 0, \\ 0 & \text{w p.p.} \end{cases}$$

z nieznanym parametrem $\theta > 0$. Obserwujemy wartości

$$Y_i = \left\lfloor \frac{2}{3} X_i \right\rfloor, \quad i = 1, \ldots, n$$

(gdzie $\lfloor x \rfloor$ oznacza największą liczbę całkowitą m taką, że $m \leq x$). Oznaczmy $S = \sum_{i=1}^{n} Y_i$.
Estymator największej wiarogodności parametru θ oparty na obserwacjach $Y_1, \ldots, Y_n$ podany jest wzorem:

$Y_i = \lfloor \frac{2}{3} X_i \rfloor$ ma rozkład:

$$P(Y_i = 0) = P\left(\tfrac{2}{3} X < 1\right) = \int_0^{\frac{3}{2}} \theta e^{-\theta x}\, dx$$

$$P(Y_i = 1) = P\left(1 < \tfrac{2}{3} X < 2\right) = \int_{\frac{3}{2}}^{3} \theta e^{-\theta x}\, dx$$

$$P(Y_i = k) = P\left(k < \tfrac{2}{3} X < k+1\right) = \int_{\frac{3}{2}k}^{\frac{3}{2}(k+1)} \theta e^{-\theta x}\, dx =$$

$$= -e^{-\theta x}\Big|_{\frac{3}{2}k}^{\frac{3}{2}(k+1)} = e^{-\frac{3}{2}\theta k} - e^{-\frac{3}{2}\theta(k+1)} =$$

$$= \left(e^{-\frac{3}{2}\theta}\right)^k \left(1 - e^{-\frac{3}{2}\theta}\right) \leftarrow \text{rozkład geometryczny}$$

$$L(y;\theta) = \prod_{i=1}^{n} \left(e^{-\frac{3}{2}\theta}\right)^{y_i} \left(1 - e^{-\frac{3}{2}\theta}\right) = \left(e^{-\frac{3}{2}\theta}\right)^{\sum_{i=1}^{n} y_i} \left(1 - e^{-\frac{3}{2}\theta}\right)^n$$

$$\ln(L) = -\frac{3}{2}\theta S + n \ln\left(1 - e^{-\frac{3}{2}\theta}\right)$$

$$\frac{\delta \ln(L)}{\delta \theta} = -\frac{3}{2} S + \frac{\frac{3}{2} n e^{-\frac{3}{2}\theta}}{1 - e^{-\frac{3}{2}\theta}} \quad := 0$$

$$S = \frac{n e^{-\frac{3}{2}\theta}}{1 - e^{-\frac{3}{2}\theta}} \qquad \left| \, p = e^{-\frac{3}{2}\theta} \right|$$

$$S = \frac{np}{1-p}$$

$$S - Sp = np$$

$$S = Sp + np$$

$$S = p(S+n)$$

$$p = \frac{S}{S+n}$$

$$e^{-\frac{3}{2}\theta} = \frac{S}{S+n}$$

$$-\frac{3}{2}\theta = \ln\left(\frac{S}{S+n}\right)$$

$$\theta = -\frac{2}{3} \ln\left(\frac{S}{S+n}\right)$$

Zadanie 2.

Wektor losowy (X,Y) ma łączny rozkład

$$f(x,y) = \begin{cases} 2(x+y) & \text{dla } 0<y<x<1, \\ 0 & \text{w p.p.} \end{cases}$$

Niech $M = Y - X$ oraz $N = 6X - 3Y$. Ile wynosi $E(N|M=-\frac{1}{2})$?

$$\begin{cases} M = Y - X \\ N = 6X - 3Y \end{cases} \qquad \begin{cases} Y = M + X \\ 6X = N + 3(M+X) \end{cases} \qquad \begin{cases} Y = M + X \\ 6X = N + 3M + 3X \end{cases}$$

$$3X = N + 3M$$

$$\begin{cases} X = \frac{N+3M}{3} \\ Y = \frac{6M+N}{3} \end{cases}$$

$$0 < \frac{6M+N}{3} < \frac{N+3M}{3} < 1$$

$$0 < 6M + N < N + 3M < 3$$

Trzeba ustalić $M = -\frac{1}{2}$:

$$0 < -3 + N < N - 1.5 < 3$$

$$\boxed{3 < N < 4.5}$$

Jakobianu nie trzeba liczyć bo w tym zadaniu nie zależy od żadnej zmiennej więc ostatecznie się skróci

$$f(m,n) = 2\left(\frac{n+3m}{3} + \frac{6m+n}{3}\right)|J|$$

$$f(-\tfrac{1}{2}, n) = 2\left(\frac{n-1.5}{3} + \frac{-3+n}{3}\right)|J| = \frac{2}{3}(2n-4.5)|J|$$

$$E(N|M=-\tfrac{1}{2}) = \frac{\int_3^{4.5} n(2n-4.5)\,dn}{\int_3^{4.5} 2n-4.5\,dn} = \frac{279}{16} \cdot \frac{2}{9} = \frac{31}{8}$$

Zadanie 3.

Wybieramy z koła jednostkowego dwa punkty $P_1 = (X_1, Y_1)$ oraz $P_2 = (X_2, X_2)$ w następujący sposób:

$$X_1 = R_1 \cos(\theta_1), \quad X_2 = R_2 \cos(\theta_2),$$

$$Y_1 = R_1 \sin(\theta_1), \quad Y_2 = R_2 \sin(\theta_2),$$

gdzie R_1, R_2 są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym $U(0,1)$, a θ_1, θ_2 są niezależnymi (również od R_1 i R_2) zmiennymi losowymi o rozkładzie jednostajnym $U(0,2\pi)$.
Niech $D(P_1,P_2)$ oznacza kwadrat odległości między tymi punktami, tzn.
$D(P_1,P_2) = (X_1 - X_2)^2 + (Y_1 - Y_2)^2$. Ile wynosi wartość oczekiwana $E[D(P_1,P_2)]$?

$$E[D(P_1,P_2)] = E[(X_1-X_2)^2 + (Y_1-Y_2)^2] = E[(X_1-X_2)^2] + E[(Y_1-Y_2)^2]$$

$$E[(X_1-X_2)^2] = E[X_1^2 - 2X_1X_2 + X_2^2] = EX_1^2 - 2EX_1X_2 + EX_2^2$$

$$E[X_1^2] = E[R_1^2]E[\cos^2(\theta_1)] = \int_0^1 r_1^2 \, dr_1 \cdot \int_0^{2\pi} \cos^2(\theta_1)\frac{1}{2\pi} \, d\theta_1 = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$

$$E[X_1X_2] = E[R_1]E[\cos(\theta_1)]\, E[R_2]\, E[\cos(\theta_2)] = 1 \cdot 1 \cdot \int_0^{2\pi} \cos(\theta_1)\frac{1}{2\pi} \, d\theta_1 \cdot \int_0^{2\pi} \cos(\theta_2)\frac{1}{2\pi} \, d\theta_2 = 0$$

$$E[X_2^2] = E[R_2^2]\, E[\cos(\theta_1)] = \frac{1}{6}$$

Dla $E[(Y_1-Y_2)^2]$ obliczenia analogiczne, trzeba tylko wstawić sin zamiast cos.

$$E[(X_1-X_2)^2] = \frac{1}{3}$$

$$E[(Y_1-Y_2)^2] = \frac{1}{3}$$

$$E[D(P_1,P_2)] = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

Zadanie 4.

Mamy dwie urny, A i B. Początkowo w urnie A znajdują się 3 czare kule, a w urnie B znajdują się 4 białe kule. Losujemy po jednej kuli z obu urn - po czym je zamieniamy (kulę wylosowaną z urny A wkładamy do urny B, a tą wylosowaną z urny B wkładamy do urny A). Ile wynosi granica (gdy $n \to \infty$) prawdopodobieństwa zdarzenia, że w n-tym kroku obie wylosowane kule są tego samego koloru?

Jak będziemy losować w nieskończoność i zmieniać miejscami kule to w żaden sposób nie będziemy mogli nic powiedzieć o urnach A i B, więc losowanie z dwóch urn będzie dla nas równoważne z losowaniem z jednej urny, w której są 3 kule czarne i 4 kule białe. Pozostaje więc obliczyć prawdopodobieństwo:

$$P(\text{czarna} : \text{czarna}) + P(\text{biała} : \text{biała}) = \frac{3}{7} \cdot \frac{2}{6} + \frac{4}{7} \cdot \frac{3}{6} = \frac{3}{7}$$

Zadanie 5.

Niech $X_1, \ldots, X_n, n \geq 4$ będą niezależnymi zmiennymi losowymi o rozkładzie o gęstości

$$f(x) = \begin{cases} 2xe^{-x^2}, & x > 0 \\ 0 & \text{w przeciwnym przypadku,} \end{cases}$$

Niech F oznacza dystrybuantę zmiennej losowej o gęstości f. Zdefiniujmy

$$M_n = \max(X_1, \ldots, X_n).$$

Ile wynosi $Pr\left(M_n \leq F^{-1}\left(1 - \frac{t}{n}\right)\right)$ dla ustalonego $t \in (0, n)$?

Dystrybuanta maksimum to $(F(x))^n$ stąd:

$$P\left(M_n \leq F^{-1}\left(1 - \frac{t}{n}\right)\right) = \left[F\left(F^{-1}\left(1 - \frac{t}{n}\right)\right)\right]^n = \left(1 - \frac{t}{n}\right)^n$$

Zadanie 7.

Wektor (X,Y) ma dwuwymiarowy rozkład normalny $N(\mu,\Sigma)$ ze średnią $\mu=(0,0.5)$ i macierzą kowariancji

$$\Sigma=\begin{pmatrix}1 & 1\\ 1 & 4\end{pmatrix}.$$

Zdefiniujmy zmienne losowe: $U=e^X$ oraz $V=e^Y$. Ile wynosi $Cov(U,V)$?

$$Cov(U,V)=Cov(e^X,e^Y)=E[e^Xe^Y]-E[e^X]E[e^Y]$$

$$E[e^Xe^Y]=E[e^{X+Y}]=$$

$$W=X+Y\sim N(0.5;\,1+4+2)=N(0.5;\,7)$$

$$=E[e^W]=\exp\{0.5+\tfrac{7}{2}\}=\exp\{4\}$$

$$E[e^X]=\exp\{\tfrac{1}{2}\}$$

$$E[e^Y]=\exp\{0.5+\tfrac{4}{2}\}=\exp\{2.5\}$$

$$Cov(U,V)=e^4-e^{\frac{1}{2}}e^{\frac{5}{2}}=e^4-e^3$$

Zadanie 8.

Zmienna losowa X przyjmuje wartości 1,2,3. Wiadomo, że $Pr(X=1)=1/4$. Niech $X_1,\dots,X_n, n \geq 3$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie co X. Zdefiniujmy

$$Y_i = |\{k : X_k = i\}|, \quad i = 1,2,3$$

(tj. Y_i mówi o tym, ile razy pojawiła się liczba i wśród n eksperymentów).
Wiadomo, że $Cov(Y_1+Y_2, Y_2+Y_3) = -\frac{1}{16}n$. Ile wynosi EX?

X	1	2	3
$P(X)$	$1/4$	$3/4-p$	p

$Cov(Y_1+Y_2, Y_2+Y_3) = Cov(Y_1,Y_2) + Cov(Y_2,Y_2) + Cov(Y_1,Y_3) +$
$+ Cov(Y_2,Y_3) = Cov(Y_1,Y_2) + Var(Y_2) + Cov(Y_1,Y_3) + Cov(Y_2,Y_3)$

$Var(Y_1+Y_2+Y_3) = 0$

$Var(Y_1)+Var(Y_2)+Var(Y_3)+2Cov(Y_1,Y_2)+2Cov(Y_1,Y_3)+2Cov(Y_2,Y_3) = 0$

$-\frac{Var(Y_1)+Var(Y_2)+Var(Y_3)}{2} = Cov(Y_1,Y_2)+Cov(Y_1,Y_3)+Cov(Y_2,Y_3)$

$Y_1 \sim Bin(n, \frac{1}{4})$, $E[Y_1] = \frac{n}{4}$, $Var(Y_1) = \frac{n}{4}\cdot\frac{3}{4} = \frac{3n}{16}$

$Y_2 \sim Bin(n, \frac{3}{4}-p)$, $E[Y_2] = n(\frac{3}{4}-p)$, $Var(Y_2) = n(\frac{3}{4}-p)(\frac{1}{4}+p)$

$Y_3 \sim Bin(n,p)$, $E[Y_3] = np$, $Var(Y_3) = np(1-p)$

$-\frac{1}{16}n = \frac{1}{2}(-Var(Y_1) - Var(Y_2) - Var(Y_3))$

$-\frac{1}{16}n = \frac{1}{2}n(\frac{3}{4}-p)(\frac{1}{4}+p) - \frac{1}{2}\cdot\frac{3n}{16} - \frac{1}{2}np(1-p) \quad |\cdot\frac{2}{n}$

$-\frac{1}{8} = \cancel{\frac{3}{16}} + \frac{3}{4}p - \frac{1}{4}p \cancel{-p^2} \cancel{-\frac{3}{16}} - p \cancel{+p^2} \quad |\cdot 8$

$-1 = (\frac{2}{4}p - p)8 = 4p - 8p$

$-1 = -4p$

$p = \frac{1}{4}$

$EX = 1\cdot\frac{1}{4} + 2\cdot\frac{1}{2} + 3\cdot\frac{1}{4} = \frac{1}{4} + 1 + \frac{3}{4} = 2$

Zadanie 9.

Zmienna losowa o rozkładzie Pareto(m, α) (gdzie $m > 0, \alpha > 1$) ma gęstość

$$g(x) = \begin{cases} \dfrac{\alpha m^\alpha}{x^{\alpha+1}} & x > m, \\ 0, & \text{w przeciwnym przypadku,} \end{cases}$$

Zmienne losowe $X_1, \ldots, X_n, n \geq 3$ są niezależnymi zmiennymi losowymi o rozkładzie jednostajnym $U(0, \theta), \theta > 0$, tj. o gęstości $f(x) = \dfrac{1}{\theta}$ dla $x \in (0, \theta)$. Załóżmy, że rozkład *a priori* parametru θ ma rozkład Pareto(m, α). Ile wynosi $E(\theta)$, gdzie θ jest rozkładem *a posteriori* – pod warunkiem, iż zaobserwowano $X_1 = x_1 > 0, \ldots, X_n = x_n > 0$?

$$g_{\Omega}(\theta) = \frac{\alpha m^\alpha}{\theta^{\alpha+1}} \quad \theta > m$$

$$P(\theta \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{P(X_1 = x_1, \ldots, X_n = x_n \mid \theta)\, P_{\Omega}(\theta)}{P(X)} =$$

$$= \frac{P(X_1 = x_1 \mid \theta) \cdot \ldots \cdot P(X_n = x_n \mid \theta)\, P_{\Omega}(\theta)}{P(X)} = \frac{\left(\frac{1}{\theta}\right)^n \mathbb{1}(x_1 \leq \theta) \cdot \ldots \cdot \mathbb{1}(x_n \leq \theta)\, P_{\Omega}(\theta)}{P(X)} =$$

$$= \frac{\left(\frac{1}{\theta}\right)^n \mathbb{1}(\max(x_i) \leq \theta) \cdot \frac{\alpha m^\alpha}{\theta^{\alpha+1}}\, \mathbb{1}(\theta > m)}{P(X)} = \left| \mathbb{1}_E \mathbb{1}_F = \mathbb{1}_{E \cap F} \right| =$$

$$= C \cdot \frac{1}{\theta^{\alpha+n+1}} \cdot \mathbb{1}(\theta \geq \max(x_1, x_2, \ldots, x_n, m))$$

Stała $C = (\alpha+n)\, m^{\alpha+n}$, gdzie $m = \max(x_1, x_2, \ldots x_n, m)$

Wartość oczekiwana rozkładu Pareto to:

$$E[\theta] = \frac{\alpha+n}{\alpha+n-1} \max\left(m, \max_{i \in \{1, \ldots, n\}}(x_i)\right)$$

Zadanie 10.

Rozpatrzmy następujący algorytm:
[$U(0,1)$ oznacza rozkład jednostajny na odcinku $(0,1)$, wszystkie symulacje U_1, U_2 są niezależne, $\lfloor x \rfloor$ oznacza największą liczbę całkowitą m taką, że $m \le x$].

1. Wysymuluj $U_1 \sim U(0,1)$ i podstaw $Y = \left\lfloor -\frac{\ln(U_1)}{\ln(2)} \right\rfloor$.
2. Wysymuluj $U_2 \sim U(0,1)$. Jeśli $U_2 \le \frac{2^{Y+1}}{4Y!}$ to zwróć $X := Y$, **Koniec**,

   w przeciwnym przypadku **Idź do Linii** 1.

Jaki rozkład ma wynik działania algorytmu, tj. zmienna X?

Oznaczenia (dla $p \in (0,1)$ oraz $\lambda > 0$):

- $X \sim Geo0(p)$ oznacza rozkład $Pr(X = k) = (1-p)^k p, \quad k = 0, 1, \ldots$
- $X \sim Geo1(p)$ oznacza rozkład $Pr(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, \ldots$
- $X \sim Pois(\lambda)$ oznacza rozkład $Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1 \ldots$

Metoda eliminacji:

1. Generowanie zm. los. Y o gęstości
2. Generowanie zm. los. U niezależnej
3. Jeśli $U \le \frac{f(X)}{c\, g(Y)}$ to $X := Y$, w przeciwnym przypadku punkt 1.

Trzeba wiedzieć, że $P\left(U \le \frac{f(X)}{c\,g(Y)}\right) = \frac{1}{c}$

$$P\left(U \le \frac{2^{Y+1}}{4Y!}\right) = \int_0^\infty P\left(U \le \frac{2^{Y+1}}{4Y!} \,\middle|\, Y = y\right) f_Y(y)\,dy = E\left[P\left(U \le \frac{2^{y+1}}{4y!}\right)\right] =$$

$$= E\left[\frac{2^{Y+1}}{4Y!}\right] = \sum_{i=0}^{\infty} \frac{2^{i+1}}{4i!} \cdot \left(\frac{1}{2}\right)^{i+1} = \sum_{i=0}^{\infty} \frac{1}{4i!} = \frac{e}{4} \Rightarrow c = \frac{4}{e}$$

Rozkład $g(Y)$:

$$-\frac{\ln(U_1)}{\ln(2)} = W \sim Exp(\ln(2))$$

$Y = 0: \quad P(0 < W < 1)$

$Y = 1: \quad P(1 < W < 2)$

$$Y=k: \; P(k < W < k+1) = F_W(k+1) - F_W(k) = 1 - e^{-\ln(2)(k+1)} - 1 + e^{-\ln(2)k} =$$

$$= e^{-k\ln(2)} - e^{-(k+1)\ln(2)} = \left(e^{-\ln(2)}\right)^k \left(1 - e^{-\ln(2)}\right)$$

rozkład geometryczny z parametrem $e^{-\ln(2)} = \frac{1}{2}$

$$g(Y) \sim Geo\left(\tfrac{1}{2}\right)$$

$$\frac{f(X)}{c\,g(Y)} = \frac{f(X)}{\frac{4}{e} \cdot \left(\frac{1}{2}\right)^{k+1}} = \frac{2^{k+1}}{4k!}$$

$$f(X) \cdot \frac{e}{4} = \frac{2^{k+1} \cdot \left(\frac{1}{2}\right)^{k+1}}{4k!}$$

$$f(X) = \frac{1}{4k!} \cdot \frac{4}{e} = \frac{e^{-1}}{k!} \sim Pois(1)$$