Zadanie 1.

Wektor losowy (X,Y) ma rozkład ciągły o gęstości

$$f(x,y) = \begin{cases} 2 & \text{dla } \{(x,y) : (x,y) \in (0,1)^2,\ y > x\}, \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Jakie jest prawdopodobieństwo, że $\left[\frac{Y}{X}\right]$ jest liczbą parzystą? Przez $[z]$ oznaczamy liczbę całkowitą najbliższą liczbie z (przyjmujemy $[n+1/2] = n$ dla liczby całkowitej n).

Warunki brzegowe w tym zadaniu to:

$n - \frac{1}{2} < \frac{y}{x} < n + \frac{1}{2}$, gdzie n to liczba parzysta.

Wypiszmy kilka przykładów:

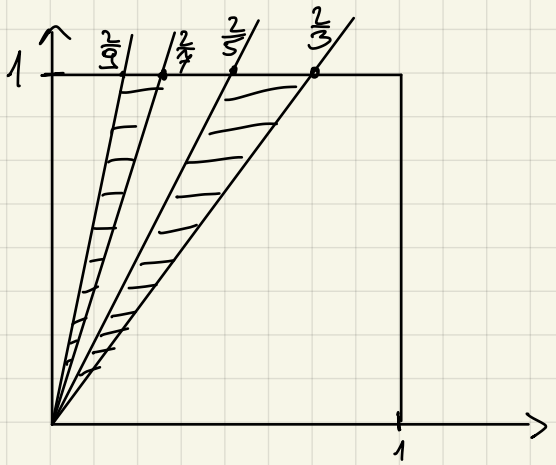
$n=2$: $\frac{3}{2}x < y < \frac{5}{2}x$

$n=4$: $\frac{7}{2}x < y < \frac{9}{2}x$

$n=6$: $\frac{11}{2}x < y < \frac{13}{2}x$

$n=8$: $\frac{15}{2}x < y < \frac{17}{2}x$

Pozostaje policzyć pole dla poszczególnych n: pomnożyć przez gęstość rozkładu, kilka pól jest zaznaczone na powyższym rysunku:



$p_2 = 2 \cdot \frac{1}{2} \cdot 1 \cdot \left(\frac{2}{3} - \frac{2}{5}\right) = \frac{4}{15}$

$p_4 = \frac{2}{7} - \frac{2}{9} = \frac{4}{63}$

$$p_6 = \frac{2}{11} - \frac{2}{13} = \frac{4}{143}$$

$$p_8 = \frac{2}{15} - \frac{2}{17} = \frac{4}{255}$$

Czyli:

$$p = \frac{4}{15} + \frac{4}{63} + \frac{4}{143} + \frac{4}{255} + \dots$$

Trzeba skorzystać z rozwinięcia arctg w szereg potęgowy:

$$\operatorname{arctg}(x) = \frac{\pi}{2} - x + \frac{x^3}{3} - \frac{x^5}{5} + \frac{x^7}{7} - \dots$$

$$\operatorname{arctg}(1) = \frac{\pi}{4} = \frac{\pi}{2} - 1 + \frac{1}{3} - \frac{1}{5} + \frac{1}{7} - \dots = \frac{\pi}{2} - 1 + \frac{2}{15} + \frac{2}{63} + \frac{2}{143} + \frac{2}{255} + \dots$$

$$\frac{\pi}{2} = \pi - 2 + \frac{4}{15} + \frac{4}{63} + \frac{4}{143} + \frac{4}{255} + \dots$$

Wracając do naszego pstwa:

$$p = \frac{4}{15} + \frac{4}{63} + \frac{4}{143} + \frac{4}{255} + \dots = 2 - \frac{\pi}{2}$$

Zadanie 2.

Niech $X_1, \ldots, X_n, n \geq 1$ będą niezależnymi zmiennymi losowymi o rozkładzie jednostajnym na odcinku (0,1). Niech $S_n = \sum_{i=1}^{n} X_i$ oraz $Z_n = \max\{X_1, \ldots, X_n\}$. Ile wynosi $Cov(S_n, Z_n)$?

$Cov(S_n, Z_n) = ?$

Obserwacja

Jeśli $X_{1:n}, \ldots, X_{n:n}$ są statystykami porządkowymi zm. jednostajnych $U(0,1)$ to ciąg $X_{1:n} / X_{n:n}, X_{2:n} / X_{n:n}, \ldots, X_{n-1:n} / X_{n:n}$ ma taki sam rozkład jak ciąg $(n-1)$ statystyk porządkowych zm. $U(0,1)$

Dowód. Wystarczy pokazać, że przed „sortowaniem" zm. $X_i / \max(X_i)$ mają rozkład jednostajny.

Uwaga – małe p oznacza gęstość p-stwa, duże P, p-stwo.

$$p\left(\frac{X_k}{\max(X_i)} = x \mid X_k < \max(X_i)\right) = \left| \text{warunkowanie po wartości } \max(X_i) \right| =$$

$$= \int_0^1 p\left(\frac{X_k}{\max(X_i)} = x \mid X_k < \max(X_i), \max(X_i) = t\right) p(\max(X_i) = t)\, dt =$$

$$= \left| \text{pod warunkiem } X_k < \max(X_i) \text{ zm. } X_k \text{ oraz } \max(X_i) \text{ są niezależne} \right| =$$

$$= \int_0^1 p\left(\frac{X_k}{t} = x \mid X_k < t\right) p(\max(X_i) = t)\, dt =$$

$$= \int_0^1 \mathbb{1}(0 < x < 1)\, p(\max(X_i) = t)\, dt = \left| \text{gęstość zm. } \max(X_i) \text{ całkuje się do } 1 \right| =$$

$= \mathbb{1}(0 < x < 1)$ – czyli dostajemy gęstość rozkładu $U(0,1)$.

Analogiczne rozumowanie należałoby przeprowadzić dla wektorów zm. $(X_1, \ldots, X_m)$ aby dostać rozkład łączny (a więc w tym przypadku – niezależność zm. przed „sortowaniem".

Wracamy do zadania.

zauważmy, że

$$S_n = X_{1:n} + X_{2:n} + \dots + X_{n:n}$$

więc

$$\operatorname{cov}(S_n, Z_n) = \sum_{i=1}^{n} \operatorname{cov}(X_{i:n}, X_{n:n}) = \sum_{i=1}^{n} \left(E[X_{i:n} X_{n:n}] - EX_{i:n} \cdot EX_{n:n} \right) =$$

$$\left| \; U_{i:n} = \frac{X_{i:n}}{X_{n:n}} \Rightarrow \begin{cases} U_{i:n} \sim U(0,1) & \text{dla } i \in \{1, \dots, n-1\} \\ U_{i:n} = 1 & \text{dla } i = n \end{cases} \right|$$

$$= \sum_{i=1}^{n} \left(E[U_{i:n} (X_{n:n})^2] - EX_{i:n} \cdot EX_{n:n} \right) =$$

$$= E(X_{n:n})^2 \sum_{i=1}^{n} (EU_{i:n}) - (EX_{n:n})^2 \sum_{i=1}^{n} EU_{i:n} =$$

$$\left| \begin{array}{l} F_X(x) = x \\ F_{max}(x) = x^n \\ EX_{n:n} = \int_0^1 1 - x^n \, dx = x - \frac{x^{n+1}}{n+1} \Big|_0^1 = 1 - \frac{1}{n+1} = \frac{n}{n+1} \\ E(X_{n:n})^2 = \int_0^1 2x - 2x^{n+1} \, dx = \frac{2x^2}{2} - \frac{2x^{n+2}}{n+2} \Big|_0^1 = 1 - \frac{2}{n+2} = \\ \quad = \frac{n}{n+2} \end{array} \right|$$

$$= \frac{n}{n+2} \left(\frac{n-1}{2} + 1 \right) - \left(\frac{n}{n+1} \right)^2 \left(\frac{n-1}{2} + 1 \right) = \frac{n}{n+2} \cdot \frac{n+1}{2} - \frac{n^2}{(n+1)^2} \cdot \frac{n+1}{2} =$$

$$= \frac{n(n+1)}{2(n+2)} - \frac{n^2}{2(n+1)} = \frac{n(n^2+2n+1) - n^2(n+2)}{2(n+1)(n+2)} = \frac{n^3 + 2n^2 + n - n^3 - 2n^2}{2(n+1)(n+2)} =$$

$$= \frac{n}{2(n+1)(n+2)}$$

Zadanie 3.

Załóżmy, że zmienna losowa X ma rozkład Poissona z parametrem $\lambda_1 > 0$, a zmienna losowa Y ma rozkład Poissona z parametrem $\lambda_2 > 0$ oraz, że X i Y są niezależne.
Ile wynosi $Var(X|X+Y=n)$ dla $n \geq 1$?

$$X|X+Y=n \sim Bin\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$

$$Var(X|X+Y=n) = n \cdot \frac{\lambda_1}{\lambda_1+\lambda_2}\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right) = \frac{n\lambda_1}{\lambda_1+\lambda_2}\left(\frac{\cancel{\lambda_1}+\lambda_2-\cancel{\lambda_1}}{\lambda_1+\lambda_2}\right) =$$

$$= \frac{n\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}$$

Zadanie 4.

Niech $X_1, \ldots, X_{2n}$ będą niezależnymi zmiennymi losowymi o rozkładzie normalnym $N(\mu, \sigma^2)$, gdzie średnia μ i wariancja σ^2 nie są znane. Oznaczmy $\hat{X}_{2n} = \frac{1}{2n}\sum_{i=1}^{2n} X_i$.
Wiemy, ile wynosi $\hat{X}_{2n}$, ale mamy dostęp tylko do obserwacji $X_1, \ldots, X_n$. Statystyka

$$S^2 = \frac{1}{n-\rho}\sum_{i=1}^{n}(X_i - \hat{X}_{2n})^2$$

jest nieobciążonym estymatorem parametru σ^2. Ile wynosi ρ?

$$S^2 = \frac{1}{n-\rho}\sum_{i=1}^{n}(X_i - \hat{X}_{2n})^2 = \frac{1}{n-\rho}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\hat{X}_{2n} + \hat{X}_{2n}^2\right) =$$

$$= \frac{1}{n-\rho}\left[\sum_{i=1}^{n} X_i^2 - 2\hat{X}_{2n}\sum_{i=1}^{n} X_i + n\hat{X}_{2n}^2\right]$$

$$E\hat{X}_{2n}^2 = \frac{1}{4n^2}E\left[\sum_{i=1}^{2n} X_i\right]^2 = \frac{1}{4n^2}E(Y^2), \text{ gdzie } Y \sim N(2n\mu, 2n\sigma^2)$$

$$E\hat{X}_{2n}^2 = \frac{1}{4n^2}\left(4n^2\mu^2 + 2n\sigma^2\right) = \mu^2 + \frac{\sigma^2}{2n}$$

$$E\left[\sum_{i=1}^{n} X_i^2\right] = \sum_{i=1}^{n} EX^2 = n\cdot(\mu^2+\sigma^2) = n\mu^2 + n\sigma^2$$

$$E\left[\hat{X}_{2n}\sum_{i=1}^{n} X_i\right] = E\left[\frac{1}{2n}\sum_{i=1}^{2n} X_i \sum_{i=1}^{n} X_i\right] = \frac{1}{2n}E\left[\left(\sum_{i=1}^{n} X_i + \sum_{i=n+1}^{2n} X_i\right)\sum_{i=1}^{n} X_i\right] =$$

$$= \frac{1}{2n} E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] + \frac{1}{2n} E\left[\sum_{i=n+1}^{2n} X_i \sum_{i=1}^{n} X_i\right] = \left| \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \right| =$$

$$= \frac{1}{2n}\left(n^2\mu^2 + n\sigma^2\right) + \frac{1}{2n} n\mu\, n\mu = \frac{1}{2n}\left(n^2\mu^2 + n\sigma^2 + n^2\mu^2\right) =$$

$$= \frac{1}{2}\left(2n\mu^2 + \sigma^2\right)$$

$$ES^2 = \frac{1}{n-\rho}\left[n\mu^2 + n\sigma^2 - 2n\mu^2 - \sigma^2 + n\left(\mu^2 + \frac{\sigma^2}{2n}\right)\right] =$$

$$= \frac{1}{n-\rho}\left[\cancel{n\mu^2} + n\sigma^2 - \cancel{2n\mu^2} - \sigma^2 + \cancel{n\mu^2} + \frac{1}{2}\sigma^2\right]$$

$$n\sigma^2 - \frac{1}{2}\sigma^2 = (n-\rho)\sigma^2$$

$$\cancel{n} - \frac{1}{2} = \cancel{n} - \rho$$

$$\rho = \frac{1}{2}$$

Zadanie 5.

Niech (X,Y) będzie wektorem losowym o dystrybuancie

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y})(1+\alpha e^{-x}e^{-y}), & x>0, y>0, \\ 0, & \text{w przeciwnym przypadku,} \end{cases}$$

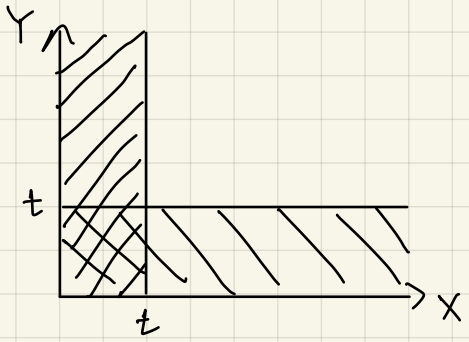
gdzie $\alpha \geq 0$. Ile wynosi $E(\min(X,Y))$?

W treści zadania mamy podaną dystrybuantę, wygląda na to, że liczenie pochodnej będzie skomplikowane, więc rozsądniej skorzystać ze wzoru na wartość oczekiwaną z dystrybuanty:

$$EX^n = \int_0^\infty n t^{n-1} P(X>t)\, dt$$

$$E[\min(X,Y)] = \int_0^\infty P(\min(X,Y)>t)\, dt = \int_0^\infty P(X>t, Y>t)\, dt$$

Pstwo $P(X>t, Y>t)$ chcemy przedstawić w takiej postaci, aby móc skorzystać z dystrybuanty $F_{X,Y}$. Dla ułatwienia spójrzmy na poniższy rysunek. Zdarzenie $\{X>t, Y>t\}$ dla dodatnich X,Y to nieograniczony obszar. Przeciwieństwo tego zdarzenia to zaznaczony obszar.



$$P(X>t, Y>t) = 1 - P(\neg\{X>t, Y>t\}) = 1 - P(X<\infty, Y<t) - P(X<t, Y<\infty) +$$
$$+ P(X<t, Y<t) = 1 - F_{X,Y}(\infty, t) - F_{X,Y}(t,\infty) + F_{X,Y}(t,t)$$

Wracamy do obliczeń:

$$E[\min(X,Y)] = \int_0^\infty 1 - F_{X,Y}(\infty,t) - F_{X,Y}(t,\infty) + F_{X,Y}(t,t)\, dt =$$
$$= \int_0^\infty 1 - (1-e^{-t}) - (1-e^{-t}) + (1-e^{-t})(1-e^{-t})(1+\alpha e^{-t}e^{-t})\, dt =$$

$$= \int_0^\infty 1-1+e^{-t}-1+e^{-t}+(1-e^{-t})^2(1+\lambda e^{-2t})\,dt =$$

$$= \int_0^\infty 2e^{-t}-1+(1-2e^{-t}+e^{-2t})(1+\lambda e^{-2t})\,dt =$$

$$= \int_0^\infty \cancel{2e^{-t}}-\cancel{1}+\cancel{1}+\lambda e^{-2t}-\cancel{2e^{-t}}-2\lambda e^{-3t}+e^{-2t}+\lambda e^{-4t}\,dt =$$

$$= \int_0^\infty \lambda e^{-2t}-2\lambda e^{-3t}+e^{-2t}+\lambda e^{-4t}\,dt =$$

$$= \frac{\lambda e^{-2t}}{-2}+\frac{2\lambda e^{-3t}}{3}-\frac{e^{-2t}}{2}-\frac{\lambda e^{-4t}}{4}\Bigg|_0^\infty =$$

$$= \frac{\lambda}{2}-\frac{2\lambda}{3}+\frac{1}{2}+\frac{\lambda}{4} = \frac{1}{12}\lambda+\frac{1}{2}$$

Zadanie 6.

Wektor (X,Y) ma dwuwymiarowy rozkład normalny $N(\mu,\Sigma)$ ze średnią $\mu=(0,1)$ i z macierzą kowariancji

$$\Sigma = \begin{pmatrix} 2 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}.$$

Ile wynosi $Cov(X^2,Y)$?

(Mnóżga trik) Szukamy zm. los. $\tilde{Y}$ takiej, że $\tilde{Y}=Y+aX$ oraz $cov(\tilde{Y},X)=0$, (dzięki temu będziemy mogli podstawić $Y=\tilde{Y}+\frac{1}{4}$ do $cov(X^2,Y)$ i w rezultacie uprościmy rachunki – do liczenia zostaną tylko momenty zm. X)

Wyznaczamy parametr a:

$Cov(\tilde{Y},X)=Cov(Y+aX,X)=Cov(Y,X)+a\,Var(X)=\frac{1}{2}+2a=0 \iff a=-\frac{1}{4}$.

Otrzymaliśmy, że $\tilde{Y}=Y-\frac{1}{4}X$. Podstawiamy więc $Y=\tilde{Y}+\frac{1}{4}X$:

$$Cov(X^2,Y)=Cov(X^2,\tilde{Y}+\tfrac{1}{4}X)=\underbrace{Cov(X^2,\tilde{Y})}_{=0}+\tfrac{1}{4}Cov(X^2,X)=\tfrac{1}{4}Cov(X^2,X)=$$

$$=\tfrac{1}{4}(E(X^3)-E(X^2)\underbrace{E(X)}_{=0}=\tfrac{1}{4}E(X^3)=0$$

Ostatnia równość wynika stąd, że $X\sim N(0,1)$ więc nieparzyste momenty są równe 0.

DRUGA METODA

$Cov(X^2, Y) = E(X^2 Y) - EX^2 EY$

$E(X^2 Y) = E[E(X^2 Y | X)] = E[X^2 E(Y|X)]$

$E(Y|X) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X) = \mu_Y + \frac{Cov(X,Y)}{\sigma_X^2}(X - \mu_X) = 1 + \frac{1}{2 \cdot 2}(X - 0) = 1 + \frac{1}{4}X$

$E[X^2(1 + \frac{1}{4}X)] = E(X^2 + \frac{1}{4}X^3) = EX^2 + \frac{1}{4}EX^3 = EX^2 = 2$

$EX^2 EY = 2 \cdot 1$

$Cov(X^2, Y) = 2 - 2 = 0$

TRZECIA METODA

tw. Wicka

$X \sim N(0,2)$, $Y \sim N(1; 0,5)$

$W = Y - 1 \sim N(0; 0,5)$

$Y = W + 1$

$Cov(X^2 Y) = EX^2 Y - EX^2 EY$

$EX^2 Y = EX^2(W+1) = EX^2 W + EX^2 = 2$

$Cov(X^2 Y) = 2 - 2 = 0$

Zadanie 7.

Niech $S_n := \left\{\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}\right\}, n \geq 6$. Dla $\{a_1, \ldots, a_m\} = S \subseteq S_n$ (oczywiście $1 \leq m \leq n-1$) niech $P(S)$ oznacza iloczyn wszystkich elementów z S, tzn. $P(S) = \prod_{i=1}^{m} a_i$. Zdefiniujmy

$$T_{parz} = \sum_{\substack{S \subseteq S_n: \\ |S| \text{ jest parzyste}}} P(S),$$

tzn. T_{parz} jest sumą wszystkich takich iloczynów dla wszystkich podzbiorów S o parzystej liczbie elementów. Ile wynosi T_{parz}?

Zdefiniujmy funkcję tworzącą:

$$F(x) = \left(1+\frac{x}{2}\right)\left(1+\frac{x}{3}\right)\left(1+\frac{x}{4}\right)\cdot\ldots\cdot\left(1+\frac{x}{n}\right)$$

Wtedy $F(x)$ to jest wielomian o takiej własności, że współczynnik przy x^k to jest $\sum_{\{S \subset S_n,\, |S|=k\}} P(S)$

Jak już mamy powyższe, to wystarczy zauważyć, że suma współczynników wielomianu $f(x)$ stojących przy x^m gdzie m są parzyste $= (f(1)+f(-1))/2$

Jak więc szukana w zadaniu wartość to $(F(1)+F(-1))/2$ czyli:

$$\left(1+\frac{1}{2}\right)\left(1+\frac{1}{3}\right)\cdot\ldots\cdot\left(1+\frac{1}{n}\right) + \left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\cdot\ldots\cdot\left(1-\frac{1}{n}\right) =$$

$$= \frac{3}{2}\cdot\frac{4}{3}\cdot\frac{5}{4}\cdot\ldots\cdot\left(\frac{n+1}{n}\right) + \frac{1}{2}\cdot\frac{2}{3}\cdot\ldots\cdot\left(\frac{n-1}{n}\right) =$$

$$= \frac{1}{2}(n+1) + \frac{1}{n} = \left(n^2+n+2\right)/2n$$

Po podzieleniu przez 2 dostajemy odpowiedź:

$$\frac{n^2+n+2}{4n}$$

Prawdopodobnie w zadaniu chodziło o niepuste zbiory S, wtedy musimy odjąć od powyższego wyrażenia $P(\emptyset) = 1$, wtedy dostajemy:

$$\frac{n^2-3n+2}{4n}$$

Zadanie 9.

Niech $X_1, \ldots, X_n, n \geq 3$ będą niezależnymi zmiennymi losowymi o rozkładzie

$$f(x) = \begin{cases} \dfrac{\alpha c^\alpha}{x^{\alpha+1}}, & x \geq c, \\ 0, & \text{w przeciwnym przypadku,} \end{cases}$$

gdzie parametr $c > 0$ jest znany, a parametr $\alpha > 0$ nie jest znany. Niech $\hat{\alpha}_n$ oznacza estymator parametru α otrzymany metodą największej wiarygodności.
Ile wynosi $Var(\hat{\alpha})$?

$$f(x) = \alpha c^\alpha x^{-\alpha-1} \mathbb{1}(x \geq c)$$

$$L(x_i; \alpha) = \prod_{i=1}^{n} \alpha c^\alpha x_i^{-\alpha-1} = \alpha^n c^{n\alpha} \prod_{i=1}^{n} x_i^{-\alpha-1}$$

$$L(x_i; \alpha) = n \ln(\alpha) + n\alpha \ln(c) - (\alpha+1) \sum_{i=1}^{n} \ln(x_i)$$

$$L' = \frac{n}{\alpha} + n \ln(c) - \sum_{i=1}^{n} \ln(x_i) := 0$$

$$\frac{n}{\alpha} = \sum_{i=1}^{n} \ln(x_i) - n \ln(c) = \sum_{i=1}^{n} \ln\left(\frac{x_i}{c}\right)$$

$$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln\left(\frac{x_i}{c}\right)}$$

Rozkład $\ln\left(\frac{X}{c}\right)$: $P\left(\ln\left(\frac{X}{c}\right) < x\right) = P\left(X < c \exp(x)\right)$

$$F_X(x) = \int_c^x \alpha c^\alpha u^{-\alpha-1} du = \alpha c^\alpha \frac{u^{-\alpha}}{-\alpha} \Big|_c^x = -c^\alpha u^{-\alpha} \Big|_c^x =$$

$$= -c^\alpha x^{-\alpha} + c^\alpha c^{-\alpha} = 1 - c^\alpha x^{-\alpha} = 1 - \left(\frac{c}{x}\right)^\alpha$$

$$P(X < c \exp(x)) = 1 - \left(\frac{c}{ce^x}\right)^\alpha = 1 - e^{-\alpha x} \sim Exp(\alpha) \quad \Gamma(n, \alpha)$$

$$EX = n \int_0^\infty \frac{1}{x} \frac{\alpha^n e^{-\alpha x} x^{n-1}}{\Gamma(n)} dx = \frac{n\alpha^n}{\Gamma(n)} \int_0^\infty e^{-\alpha x} x^{n-1-1} dx =$$

$$= \frac{n\alpha^n}{\Gamma(n)} \cdot \frac{\Gamma(n-1)}{\alpha^{n-1}} = \frac{n\alpha^n \cancel{\Gamma(n-1)}}{(n-1)\cancel{\Gamma(n-1)}\alpha^{n-1}} = \frac{n}{n-1}\alpha$$

$$EX^2 = n^2 \int_0^\infty \frac{1}{x^2} \frac{\alpha^n e^{-\alpha x} x^{n-1}}{\Gamma(n)} dx = \frac{n^2\alpha^n}{\Gamma(n)} \int_0^\infty e^{-\alpha x} x^{n-2-1} dx =$$

$$= \frac{n^2\alpha^n}{\Gamma(n)} \frac{\Gamma(n-2)}{\alpha^{n-2}} = \frac{n^2\alpha^n \cancel{\Gamma(n-2)}}{\alpha^{n-2}(n-1)(n-2)\cancel{\Gamma(n-2)}} = \frac{n^2\alpha^2}{(n-1)(n-2)}$$

$$Var(X) = \frac{n^2\alpha^2}{(n-1)(n-2)} - \frac{n^2\alpha^2}{(n-1)^2} = \frac{n^2\alpha^2(n-1) - n^2\alpha^2(n-2)}{(n-1)^2(n-2)} =$$

$$= \frac{n^2\alpha^2}{(n-1)^2(n-2)}$$

Zadanie 10.

Wysokość szkód X dla nieznanych parametrów $\alpha \in (0,1), \beta \in (0,1)$ ma następujący rozkład

$$Pr(X = x|\alpha,\beta) = \begin{cases} \alpha, & x = 0, \\ (1-\alpha)\beta(1-\beta)^{x-1}, & x = 1,2,\ldots, \end{cases}$$

Zaobserwowano $X_1, X_2, \ldots, X_n$ z czego wiadomo, iż k razy zaobserwowano 0 (tzn. $X_i = 0$ dla $i \in A \subset \{1,\ldots,n\}, |A| = k$).
Załóżmy, że α i β są niezależne oraz, że α ma rozkład a priori $Beta(a_1,a_2)$, $a_1,a_2 \in (0,1)$, natomiast β ma rozkład a priori $Beta(b_1,b_2), b_1,b_2 \in (0,1)$. Zmienna o rozkładzie $Beta(v,w)$ ma gęstość:

$$f(x) = \frac{x^{v-1}(1-x)^{w-1}}{B(v,w)}, x \in (0,1), v \in (0,1), w \in (0,1),$$

gdzie $B(v,w) = \dfrac{\Gamma(v)\Gamma(w)}{\Gamma(v+w)}$, a $\Gamma(z) = \displaystyle\int_0^\infty x^{z-1}e^{-x}dx$.

Jaki rozkład a posteriori ma β ?

$$P(\beta,\alpha|X=x) = \frac{P(X_1=x_1,\ldots,X_n=x_n|\alpha,\beta)\,P_\alpha(\alpha)\,P_\beta(\beta)}{P(X)}$$

$$= C \cdot \left[\prod_{i=1}^{n} P(X_i = x_i|\alpha,\beta)\right] P_\alpha(\alpha)\,P_\beta(\beta)$$

$$= C \cdot \alpha^k \prod_{i=1}^{n-k} (1-\alpha)\beta(1-\beta)^{x_i-1} \cdot \frac{\alpha^{a_1-1}(1-\alpha)^{a_2-1}}{B(a_1,a_2)} \cdot \frac{\beta^{b_1-1}(1-\beta)^{b_2-1}}{B(b_1,b_2)} =$$

$$= C\,\beta^{n-k}(1-\beta)^{\sum x_i-(n-k)} \cdot \beta^{b_1-1}(1-\beta)^{b_2-1}\,\alpha^{k+a_1-1}(1-\alpha)^{a_2-1}$$

$$P(\beta|X=x) = C\beta^{n-k}(1-\beta)^{\sum x_i-(n-k)} \cdot \beta^{b_1-1}(1-\beta)^{b_2-1} \cdot \int_0^1 \frac{\alpha^{k+a_1-1}(1-\alpha)^{a_2-1}}{B(k+a_1,\,a_2)}d\alpha =$$

$$= C\,\beta^{n-k+b_1-1}(1-\beta)^{\sum X_i-(n-k)+b_2-1}$$

Stąd:

$$\beta \sim Beta\left(n-k+b_1,\ \sum_{i=1}^{n-k} X_i-(n-k)+b_2\right)$$