Zadanie 1.

Niech N będzie zmienną losową z rozkładu jednostajnego na $\{0,\dots,19\}$. Podaj ile wynosi wartość oczekiwana EX, gdzie

$$X = \sum_{k=0}^{N} \binom{N-k}{k} (-1)^k.$$

(Tradycyjnie przyjmujemy $0! = 1$ oraz $\binom{m}{n} = 0$ dla $m < n$).

N ma rozkład jednostajny na $\{0, 1, \dots, 19\}$

Niech $X_n = \sum_{k=0}^{n} \binom{n-k}{k} (-1)^k$

Oczywiście:

$EX = \frac{1}{20}(X_0 + X_1 + \dots + X_{19})$

Więc zastanówmy się ile wynosi $(X_0 + X_1 + \dots + X_{19})$

$X_0 = 1$

$X_1 = \binom{1}{0} = 1$

$X_2 = \binom{2}{0} - \binom{1}{1} = 0$

$X_3 = \binom{3}{0} - \binom{2}{1} = 1 - 2 = -1$

$X_4 = \binom{4}{0} - \binom{3}{1} + \binom{2}{2} = 1 - 3 + 1 = -1$

$X_5 = \binom{5}{0} - \binom{4}{1} + \binom{3}{2} = 1 - 4 + 3 = 0$

$X_6 = \binom{6}{0} - \binom{5}{1} + \binom{4}{2} - \binom{3}{3} = 1 - 5 + 6 - 1 = 1$

$X_7 = \binom{7}{0} - \binom{6}{1} + \binom{5}{2} - \binom{4}{3} = 1 - 6 + 10 - 4 = 1$

$X_8 = \binom{8}{0} - \binom{7}{1} + \binom{6}{2} - \binom{5}{3} + \binom{4}{4} = 1 - 7 + 15 - 10 + 1 = 0$

zakładając, że ciąg $(1, 1, 0, -1, -1, 0)$ $|\Sigma = 0|$ dalej powtarza się cyklicznie,

dostajemy:

$\frac{1}{20}(0 + 0 + 0 + 1 + 1) = \frac{2}{20} = \frac{1}{10}$

Zadanie 2.

Rzucamy niezależnie symetryczną monetą. Jeśli wypadnie orzeł otrzymujemy 1 punkt, jeśli reszka 2 punkty. Początkowo mamy 0 punktów. Jakie jest prawdopodobieństwo, że w którymś momencie uzyskamy dokładnie n punktów (dla $n \geq 9$)?

(A) $\frac{1}{3} - \frac{1}{3}\left(-\frac{1}{2}\right)^n$

(B) $\frac{1}{2}$

(C) $\frac{2}{3} + \frac{1}{3}\left(-\frac{1}{2}\right)^n$

(D) $1 - \frac{1}{3}\left(1 + \left(-\frac{1}{2}\right)^n\right)$

(E) $\frac{1}{2} - \frac{2}{3}\left(-\frac{1}{2}\right)^n$

W warunkach egzaminacyjnych można policzyć szukane p-stwo dla $n = 1, 2$ i zauważyć, że jedynie odp. C zgadza się z rachunkami. Poniżej przedstawię pełne rozwiązanie.

Niech $p(n)$ będzie szukanym p-stwem. Wyrazimy $p(n)$ w terminach $p(n-1)$ i $p(n-2)$, co doprowadzi do równania rekurencyjnego, które potem rozwiążemy.

Rozważmy skończony ciąg rzutów, który kończy się uzyskaniem dokładnie n punktów. Zdarzenie to można rozbić na dwa rozłączne podzdarzenia w zależności od tego, czy w ostatnim rzucie wyrzuciliśmy reszkę czy orła.

- Uzyskanie wyniku $n \geqslant 2$ punktów, gdy w ostatnim rzucie dostaliśmy orła ma p-stwo $\frac{1}{2} p(n-1)$.
- Uzyskanie wyniku $n \geqslant 2$ punktów, gdy w ostatnim rzucie dostaliśmy reszkę ma p-stwo $\frac{1}{2} p(n-2)$

Jak więc dla $n \geqslant 2$ mamy:

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

Ponadto $p(1) = \frac{1}{2}$ (orzeł w pierwszym rzucie) i $p(2) = \frac{3}{4}$ (reszka w pierwszym rzucie lub dwa orły w dwóch pierwszych rzutach)

Równanie rekurencyjne na $p(n)$ wraz z wartościami brzegowymi $p(1)$, $p(2)$ jednoznacznie wyznacza wartości $p(n)$, więc formalnie do zakończenia dowodu wystarczy sprawdzić, że wzór podany w odpowiedzi C spełnia równanie rekurencyjne, oraz zgadza się z wartościami $p(1), p(2)$.

Istnieją też ogólne metody na znajdowanie rozwiązań. Do rozwiązania dostajemy równanie charakterystyczne:

$$x^2 = \frac{1}{2}x + \frac{1}{2},$$

którego pierwiastki to 1 oraz $-\frac{1}{2}$, więc z ogólnej teorii rozwiązanie ma postać:

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

gdzie stałe α, β należy dobrać tak żeby spełnione były warunki brzegowe na $p(1), p(2)$.

Zadanie 3.

Niech $\{X_i\}_{i=0}^{\infty}$ będzie łańcuchem Markowa na przestrzeni stanów $E=\{0,1,2\}$ z macierzą przejść

$$\mathbf{P}=\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

Rozkład początkowy jest jednostajny $Pr(X_0=k)=\frac{1}{3}$ dla $k\in\{0,1,2\}$.
Zdefiniujmy

$$\begin{aligned} Z_0 &= 0, \\ Z_k &= (X_k - X_{k-1}) \bmod 3, \qquad \text{dla } k\geq 3, \end{aligned}$$

(przyjmujemy typowo: $(-1) \bmod 3 = 2$ oraz $(-2) \bmod 3 = 1$).

Ile wynosi EZ_n?

Wyznaczmy rozkład zm. X_k.

Wiemy, że X_0 ma rozkład $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

Bezpośrednio z macierzy przejścia otrzymujemy

$$P(X_k=0)=p(0\to 0)P(X_{k-1}=0)+p(1\to 0)P(X_{k-1}=1)+p(2\to 0)P(X_{k-1}=2)$$

$$P(X_k=1)=p(0\to 1)P(X_{k-1}=0)+p(1\to 1)P(X_{k-1}=1)+p(2\to 1)P(X_{k-1}=2)$$

$$P(X_k=2)=p(0\to 2)P(X_{k-1}=0)+p(1\to 2)P(X_{k-1}=1)+p(2\to 2)P(X_{k-1}=2)$$

$$P(X_k=0)=\tfrac{1}{2}P(X_{k-1}=0)+\tfrac{1}{4}P(X_{k-1}=1)+\tfrac{1}{4}P(X_{k-1}=2)$$

$$P(X_k=1)=\tfrac{1}{4}P(X_{k-1}=0)+\tfrac{1}{2}P(X_{k-1}=1)+\tfrac{1}{4}P(X_{k-1}=2)$$

$$P(X_k=2)=\tfrac{1}{4}P(X_{k-1}=0)+\tfrac{1}{4}P(X_{k-1}=1)+\tfrac{1}{2}P(X_{k-1}=2)$$

$$P(X_k=1)-P(X_k=2)=\tfrac{1}{4}P(X_{k-1}=1)-\tfrac{1}{4}P(X_{k-1}=2)$$

Jak więc

$$P(X_k=1)-P(X_k=2)=\ldots=\left(\tfrac{1}{4}\right)^k\left(P(X_0=1)-P(X_0=2)\right)=0$$

$P(X_k=1)=P(X_k=2)$ co trochę upraszcza równania

Odejmijmy inną parę równań:

$P(X_k=0)-P(X_k=1) = \frac{1}{4}P(X_{k-1}=0) - \frac{1}{4}P(X_{k-1}=1)$, więc analogicznie jak powyżej:

$P(X_k=0)=P(X_k=1)$,

czyli X_k ma rozkład $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ dla każdego k.

Ogólny fakt: jeśli macierz przejścia łańcucha Markowa jest symetryczna, to rozkład jednostajny na przestrzeni stanów jest rozkładem stacjonarnym.

$Z_k = (X_k - X_{k-1})$ - skoro wiemy, że X_{k-1} ma rozkład jednostajny $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, to sprawdzamy wszystkie 9 możliwości

$$E(Z_k) = \frac{1}{3}\left(0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 0\cdot\frac{1}{2} + 1\cdot\frac{1}{4} + 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 0\cdot\frac{1}{2}\right) =$$

$$= \frac{1}{3}\cdot\frac{9}{4} = \frac{3}{4}$$

Zadanie 4.

Niech $X_1, \ldots, X_n, n \geq 4$ będą niezależnymi zmiennymi losowymi o rozkładzie $U(0, \theta)$, $\theta > 0$. Zdefiniujmy $T = \max_{1 \leq i \leq n} X_i$. Znajdź wartość $\alpha > 0$, które minimaluzuje błąd średniokwadratory MSE estymatora αT parametru θ.

(Dla estymatora $\hat{\theta}$ parametru θ MSE definiujemy następująco: $\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$).

$$F_X(x) = \frac{x}{\theta}$$

$$F_T(t) = \left[F_X(t)\right]^n = \left(\frac{t}{\theta}\right)^n$$

$$f_T(t) = \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1} = \frac{n}{\theta^n} t^{n-1}$$

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = E\left[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2\right] = \alpha^2 ET^2 - 2\theta\alpha ET + \theta^2$$

$$ET = \frac{n}{\theta^n}\int_0^\theta t^n \, dt = \frac{n}{\theta^n} \frac{t^{n+1}}{n+1}\Bigg|_0^\theta = \frac{n}{n+1}\frac{\theta^{n+1}}{\theta^n} = \frac{n\theta}{n+1}$$

$$ET^2 = \frac{n}{\theta^n}\int_0^\theta t^{n+1} \, dt = \frac{n}{\theta^n} \frac{t^{n+2}}{n+2}\Bigg|_0^\theta = \frac{n}{n+2}\frac{\theta^{n+2}}{\theta^n} = \frac{n\theta^2}{n+2}$$

$$MSE(\hat{\theta}) = \alpha^2 \frac{n\theta^2}{n+2} - \alpha 2\theta \frac{n\theta}{n+1} + \theta^2 = \alpha^2 \frac{n\theta^2}{n+2} - \alpha\frac{2n\theta^2}{n+1} + \theta^2$$

$$\frac{\partial MSE(\hat{\theta})}{\partial \alpha} = \alpha \frac{2n\theta^2}{n+2} - \frac{2n\theta^2}{n+1} := 0$$

$$\frac{\alpha}{n+2} = \frac{1}{n+1}$$

$$\alpha = \frac{n+2}{n+1}$$

Zadanie 5.

Mamy ciąg zmiennych losowych $X_1,\ldots,X_n, n \geq 4$ takich, że $EX_i = \sqrt{i}, VarX_i = 1, i = 1,\ldots,n$ oraz $Cov(X_i,X_j) = \rho$ dla $i \neq j$. Zmienne losowe $I_1,\ldots,I_n$ są wzajemnie niezależne, a także są niezależne od ciągu $X_1,\ldots,X_n$, mają rozkład $P(I_i = 0) = P(I_i = 1) = 0.5$.

Oblicz $Var\left(\sum_{i=1}^{n} I_i X_i\right)$.

Wiemy, że

$$Var(Y_1 + \ldots + Y_n) = \sum_k Var(Y_k) + 2\sum_{i<j} Cov(Y_i, Y_j)$$

W naszym przypadku

$$Var(I_1X_1 + \ldots + I_nX_n) = \sum Var(I_kX_k) + 2\sum_{i<j} Cov(I_iY_i, I_jY_j) =$$

Czyli musimy policzyć $Var(I_kX_k)$, oraz $Cov(I_jX_i, I_jX_j)$.

$$Var(I_kX_k) = E(I_kX_k)^2 - (EI_k)^2(EX_k)^2 = \frac{1}{2}EX_k^2 - \frac{1}{4}(EX_k)^2 = \frac{1}{2}(1+k) - \frac{1}{4}k = \frac{1}{2} + \frac{1}{4}k$$

$$Cov(I_iX_i, I_jX_j) = E(I_iI_jX_iX_j) - E(I_iX_i)E(I_jX_j) =$$

$$= \frac{1}{4}E(X_iX_j) - \frac{1}{4}E(X_i)E(X_j) = \frac{1}{4}Cov(X_i, X_j) = \frac{1}{4}\rho$$

Tak więc:

$$Var(I_1X_1 + \ldots + I_nX_n) = \sum_{k=1}^{n}\left(\frac{1}{2} + \frac{1}{4}k\right) + \frac{1}{4}n(n-1)\rho =$$

$$= \frac{1}{4}\left(2n + \frac{1}{2}n(n+1) + n(n-1)\rho\right) = \frac{1}{8}n\left(4 + (n+1) + 2(n-1)\rho\right) =$$

$$= \frac{1}{8}n\left(5 + n + 2\rho(n-1)\right)$$

Zadanie 6.

Niech (X,Y) będzie wektorem losowym o łącznej gęstości

$$f(x,y) = \begin{cases} 6x & \text{dla } x \geq 0, y \geq 0, x+y \leq 1, \\ 0, & \text{w przeciwnym przypadku,} \end{cases}$$

Zdefiniujmy zmienną losową $T = \dfrac{X}{X+Y}$. Ile wynosi $VarT$?

$$E\left[\frac{X}{X+Y}\right] = \int_0^1 \int_0^{1-x} \frac{x}{x+y} 6x \, dy \, dx = 6 \int_0^1 x^2 \int_0^{1-x} \frac{1}{x+y} \, dy \, dx =$$

$$= 6 \int_0^1 x^2 \ln(x+y) \Big|_0^{1-x} dx = 6 \int_0^1 x^2 (-\ln(x)) \, dx =$$

$$= -6 \int_0^1 \left(\frac{x^3}{3}\right)' \ln x \, dx = -6 \left[\frac{x^3 \ln x}{3} \Big|_0^1 - \int_0^1 \frac{x^2}{3} dx \right] =$$

$$= -2 \left[-\frac{x^3}{3} \right]_0^1 = \frac{2}{3}$$

$$E\left[\left(\frac{x}{x+y}\right)^2\right] = \int_0^1 \int_0^{1-x} \frac{6x^3}{(x+y)^2} \, dy \, dx = 6 \int_0^1 x^3 \int_0^{1-x} (x+y)^{-2} \, dy \, dx =$$

$$= 6 \int_0^1 x^3 (x+y)^{-1} \Big|_0^{1-x} dx = 6 \int_0^1 x^3 \left(1 - \frac{1}{x}\right) dx =$$

$$= 6 \int_0^1 x^3 - x^2 \, dx = 6 \left[\frac{x^4}{4} - \frac{x^3}{3} \right]_0^1 = 6 \cdot \left(\frac{1}{4} - \frac{1}{3}\right) = \frac{1}{2}$$

$$Var(T) = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

Zadanie 8.

Wektor (X,Y,Z) ma trójwymiarowy rozkład normalny $N(\boldsymbol{\mu},\Sigma)$ ze średnią $\boldsymbol{\mu}=(\mu,4\mu,3\mu)$, gdzie parametr $\mu \neq 0$ nie jest znany, a macierz kowariancji jest następująca:

$$\Sigma = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Interesują nas estymatory nieznanego parametru μ postaci

$$\hat{\mu} = aX + bY + cZ, \quad a,b,c \in \mathbb{R}.$$

Rozpatrujemy tylko estymatory *nieobciążone*. Ile wynosi najmniejsza wariancja takiego estymatora?

Chcemy wyznaczyć najmniejszą wariancję estymatora $\hat{\mu} = aX + bY + cZ$ parametru μ. Ma on być nieobciążony, zatem:

$$E\hat{\mu} = \mu = a\mu + 4b\mu + 3c\mu \iff a + 4b + 3c = 1 \iff a = 1 - 4b - 3c$$

Wyznaczmy teraz ogólny wzór na wariancję tego estymatora. Skorzystamy z następujących wzorów:

- $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{1\le i < j \le n} Cov(X_i, X_j)$
- $Var(aX) = a^2 Var(X)$
- $Cov(aX, bY) = ab\,Cov(X,Y)$

Zatem mamy (w przejściu $*$ skorzystamy z wynikającej z nieobciążoności estymatora równości):

$$Var(\hat{\mu}) = Var(aX) + Var(bY) + Var(cZ) + 2Cov(aX,bY) + 2Cov(aX,cZ) +$$
$$+ 2Cov(bY,cZ) =$$
$$= 1a^2 + 4b^2 + 2c^2 + 1\cdot 2ab + 0\cdot 2ac + 1\cdot 2bc =$$
$$= a^2 + 2ab + 4b^2 + 2c^2 + 2bc \overset{*}{=}$$
$$\overset{*}{=} 1 - 4b - 3c - 4b + 16b^2 + 12bc - 3c + 12bc + 9c^2 + 2b - 8b^2 -$$
$$- 6bc + 4b^2 + 2c^2 + 2bc =$$
$$= 12b^2 + 11c^2 + 20bc - 6b - 6c + 1 := f(b,c)$$

Wyznaczmy minimum funkcji $f(b, c)$:

- $\frac{\partial f}{\partial b} = 24b + 20c - 6 = 0 \iff c = 0.3 - 1.2b$

- $\frac{\partial f}{\partial c} = 22c + 20b - 6 = 0 \iff c = \frac{6}{22} - \frac{20}{22}b$

Po rozwiązaniu układu równań otrzymamy punkt stacjonarny $(b, c) = \left(\frac{3}{32}, \frac{3}{16}\right)$. Wyznaczając drugie pochodne, można pokazać, że jest to minimum lokalne (i zarazem globalne). Wstawiając otrzymane wartości do funkcji f, otrzymujemy:

$$f\left(\frac{3}{32}, \frac{3}{16}\right) = \frac{5}{32}$$

Zadanie 9.

Szkody $X_1, \ldots, X_n, n \geq 3$ są niezależnymi zmiennymi losowymi o rozkładzie normalnym ze *znaną* średnią μ_0 i nieznaną wariancją σ^2. Oznaczmy $\xi = \frac{1}{\sigma^2}$. Załóżmy, że rozkład *a priori* parametru ξ to rozkład $\Gamma(a,b)$ (tj. rozkład gamma) ze znanymi parametrami a oraz b. Wtedy rozkład *a posteriori* ξ pod warunkiem $X_1 = x_1, \ldots, X_n = x_n$ ma również rozkład gamma z parametrami a_1, b_1, tj. $\Gamma(a_1, b_1)$. Ile wynosi parametr a_1 ?

Zmienna o rozkładzie $\Gamma(\alpha,\beta)$ $(\alpha,\beta > 0)$ ma gęstość

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0,$$

gdzie

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

$$X \sim N(\mu_0, \sigma^2)$$

$$\varepsilon = \frac{1}{\sigma^2} \sim \Gamma(a, b) \qquad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{(x-\mu_0)^2}{2\sigma^2}\right\}$$

$$P(\varepsilon \mid X_1 = x_1, \ldots, X_n = x_n) = \frac{P(X_1 = x_1, \ldots, X_n = x_n \mid \varepsilon = \varepsilon) f_\varepsilon(\varepsilon)}{P(X)} =$$

$$= \prod_{i=1}^n \frac{\sqrt{\varepsilon}}{\sqrt{2\pi}} \exp\left\{-\frac{\varepsilon(x_i-\mu_0)^2}{2}\right\} \cdot \frac{b^a}{\Gamma(a)} \varepsilon^{a-1} e^{-b\varepsilon} \cdot \frac{1}{P(X)} =$$

$$= C \left(\sqrt{\varepsilon}\right)^n \exp\left\{-\frac{\varepsilon}{2}\sum_{i=1}^n (x_i-\mu_0)^2\right\} \varepsilon^{a-1} \exp(-b\varepsilon) =$$

$$= C\, \varepsilon^{a+\frac{n}{2}-1} \exp\left\{-\left(\frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)^2 + b\right)\varepsilon\right\}$$

Czyli $a_1 = a + \frac{n}{2}$