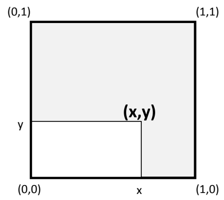
Egzamin z Rachunku Prawdopodobieństwa, 8 lutego 2013r.
Wersja I

1. Pewien artysta współczesny tworzy instalacje w następujący sposób: wybiera losowo punkt (X, Y) z blachy $[0,1] \times [0,1]$, a następnie odcina lewy dolny róg blachy (obszar poniżej i na lewo od punktu (X, Y)) i wykorzystuje go jako podpórkę, zaś do pozostałej części blachy przykleja pióra. Jaka jest średnio powierzchnia obklejona piórami? (*4 pkt*)
Rozw. Warto zrobić następujący pomocniczy rysunek:



Interesuje nas średnie pole szarego obszaru. Jeśli przez (x, y) oznaczymy współrzędne wylosowanego punktu, to pole szarego obszaru wynosi $1 - xy$. Chcąc obliczyć średnie pole musimy najpierw wyznaczyć rozkład zmiennej (X, Y) – jest to rozkład jednostajny na kwadracie, a więc o gęstości $g(x, y) = \mathbf{1}_{[0,1]}(x) \cdot \mathbf{1}_{[0,1]}(y)$. A zatem średnie pole szarego obszaru

$$\mathbb{E}(1 - XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (1 - xy)\mathbf{1}_{[0,1]}(x) \cdot \mathbf{1}_{[0,1]}(y)dxdy = \int_0^1 \int_0^1 (1 - xy)dxdy = 1 - \frac{x^2}{2}\bigg|_0^1 \cdot \frac{y^2}{2}\bigg|_0^1 = \frac{3}{4}.$$

Uwaga. NIE WOLNO pisać, że średnie pole szarego obszaru $= 1 - \mathbb{E}X \cdot \mathbb{E}Y$, o ile nie skomentuje się najpierw faktu, iż zmienne X i Y są niezależne.

Wyznaczyć rozkład zmiennej $X + Y$ (*4 pkt*).
Rozw. Rozkład można wyznaczać bądź wprost z definicji zmiennej (X, Y), bądź korzystając z faktu, iż zmienne X i Y są niezależne i gęstość sumy jest splotem gęstości.
METODA 1. Ze splotu. Zmienne X i Y są niezależne o rozkładach jednostajnych na $[0,1]$, a więc

$$g_{X+Y}(t) = g_X * g_Y(t) = \int_{-\infty}^{\infty} \mathbf{1}_{[0,1]}(x) \cdot \mathbf{1}_{[0,1]}(t - x)dx = \int_0^1 \mathbf{1}_{[0,1]}(t - x)dx = \int_0^1 \mathbf{1}_{[t-1,t]}(x)dx.$$

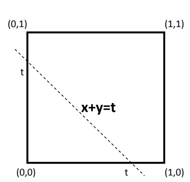
To ostatnie wyrażenie zależy od wartości t: gdy $t \notin [0,2]$, całka równa jest 0; gdy $t \in [0,1]$, całka równa jest $\int_0^t dx$, gdy $t \in [1,2]$ całka równa jest $\int_{t-1}^1 dx$, a więc ostatecznie:

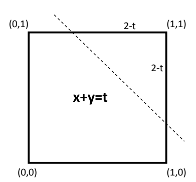
$$g_{X+Y}(t) = \begin{cases} t & t \in [0,1] \\ 2 - t & t \in (1,2] \\ 0 & t \notin [0,2] \end{cases}.$$

METODA 2. Z definicji. Wyznaczymy dystrybuantę zmiennej $X + Y$, czyli

$$\mathbb{P}(X + Y \leqslant t) = \iint_{x+y \leqslant t} \mathbf{1}_{[0,1]}(x) \cdot \mathbf{1}_{[0,1]}(y)dxdy.$$

Zauważmy, iż jeśli $t < 0$ to szukane prawdopodobieństwo równe jest 0; jeśli $t > 2$, to szukane prawdopodobieństwo równe jest 1. Dla pozostałych t mamy dwa możliwe przypadki: gdy $t \in [0,1]$ (rys. po lewej) oraz gdy $t \in [1,2]$ (rys. po prawej):





W tym pierwszym przypadku całka równa jest polu trójkąta poniżej prostej $x+y=t$, a więc $\frac{t^2}{2}$; w tym drugim przypadku całka równa jest polu kwadratu pomniejszonego o pole trójkąta nad prostą $x+y=t$: $1-\frac{(2-t)^2}{2}$. Mamy więc:

$$F_{X+Y}(t) = \begin{cases} 0 & t<0 \\ \frac{t^2}{2} & t\in[0,1] \\ 1-\frac{(2-t)^2}{2} & t\in(1,2] \\ 1 & t>2 \end{cases}$$

2. Dwuwymiarowa zmienna losowa (X,Y) ma rozkład z gęstością $g(x,y)=(x+y)1_{\{0\leqslant x\leqslant 1\}}1_{\{0\leqslant y\leqslant 1\}}$. Wyznaczyć rozkłady zmiennych X i Y (*4 pkt*). Obliczyć kowariancję zmiennych X i Y (*4 pkt*). Czy zmienne X i Y są niezależne? Odpowiedź uzasadnić (*2 pkt*).
Rozw. Rozkład X jest ciągły i ma gęstość

$$g_X(x)=\int_{-\infty}^{\infty} g(x,y)\,dy=\int_0^1 (x+y)1_{\{0\leqslant x\leqslant 1\}}\,dy=(x+\frac{1}{2})1_{\{0\leqslant x\leqslant 1\}}.$$

Podobnie, rozkład Y jest ciągły i ma gęstość $g_Y(y)=(y+\frac{1}{2})1_{\{0\leqslant y\leqslant 1\}}$. W celu obliczenia kowariancji zmiennych X i Y liczymy

$$\mathbb{E}XY=\int_0^1\int_0^1 xy(x+y)\,dy\,dx=\int_0^1 (\frac{1}{2}x^2y^2+\frac{1}{3}xy^3)\Big|_{y=0}^1\,dx=\int_0^1 \frac{1}{2}x^2+\frac{1}{3}x\,dx=\frac{1}{3},$$

$$\mathbb{E}X=\int_0^1 x(x+\frac{1}{2})\,dx=\frac{1}{3}+\frac{1}{4}=\frac{7}{12}$$

i podobnie $\mathbb{E}Y=\frac{7}{12}$. Wobec tego $\mathrm{Cov}(X,Y)=\mathbb{E}XY-\mathbb{E}X\mathbb{E}Y=\frac{1}{3}-(\frac{7}{12})^2=-\frac{1}{144}$. Zmienne X, Y nie są niezależne, albowiem ich kowariancja jest różna od 0.

3. Dwuwymiarowa zmienna losowa (X,Y) ma rozkład z gęstością $g(x,y)=\frac{e^{-y}}{y}1_{\{0<x<y\}}$. Obliczyć $\mathbb{E}(X\,|\,Y)$ (*4 pkt*), $\mathbb{E}(X^2\sin(Y)\,|\,Y)$ (*4 pkt*), oraz $\mathbb{E}X$ (*2 pkt*).
Rozw. Dla $y>0$,

$$\mathbb{E}(X\,|\,Y=y)=\frac{\int_{-\infty}^{\infty} xg(x,y)\,dx}{\int_{-\infty}^{\infty} g(x,y)\,dx}=\frac{\int_0^y x\,dx\cdot\frac{e^{-y}}{y}}{\int_0^y 1\,dx\cdot\frac{e^{-y}}{y}}=\frac{\frac{1}{2}y^2}{y}=\frac{y}{2}.$$

Ponieważ $\mathbb{P}(Y>0)=1$, to $\mathbb{E}(X\,|\,Y)=\frac{Y}{2}$. (Można też zauważyć, że rozkładem warunkowym X pod warunkiem $Y=y$, dla $y>0$, jest rozkład jednostajny na odcinku $(0,y)$.)
Podobnie jak wyżej obliczymy, że $\mathbb{E}(X^2\,|\,Y)=\frac{1}{3}Y^2$, a zatem

$$\mathbb{E}(X^2\sin(Y)\,|\,Y)=\mathbb{E}(X^2\,|\,Y)\sin(Y)=\frac{1}{3}Y^2\sin(Y).$$

W celu obliczenia $\mathbb{E}X$, możemy skorzystać z tego, że rozkład Y ma gęstość $g_Y(y)=\int_{-\infty}^{\infty} g(x,y)\,dx=\int_0^y \frac{e^{-y}}{y}\,dy1_{\{y>0\}}=e^{-y}1_{\{y>0\}}$, czyli Y ma rozkład wykładniczy z parametrem 1. Następnie,

$$\mathbb{E}X=\mathbb{E}(\mathbb{E}(X\,|\,Y))=\mathbb{E}(Y/2)=\frac{1}{2}.$$

4. Liczba głównych wygranych w Lotto (trafień szóstek) w pojedynczym losowaniu jest zmienną losową o rozkładzie Poissona z parametrem $\lambda=1/4$ (czyli przyjmuje wartość $k\in\{0,1,2,\ldots\}$ z prawdopodobieństwem $\frac{\lambda^k}{k!}e^{-\lambda}$). Przypuśćmy, że każdy z posiadaczy wygrywającego losu odbiera swoją nagrodę z prawdopodobieństwem 90%. Niech X oznacza liczbę głównych wygranych, jakie padną w najbliższym losowaniu, natomiast Y — liczbę głównych wygranych, które zostaną odebrane. Opisać rozkład warunkowy $Y|X=x$ (*2 pkt*). Obliczyć $\mathbb{E}(Y|X)$ (*3 pkt*) oraz $\mathbb{E}Y$ (*2 pkt*).
Rozw. Dla $x\in\{0,1,2,\ldots\}$, rozkładem warunkowym Y pod warunkiem $X=x$ jest rozkład Bernoulliego z parametrami $n=x$ i $p=0,9$. Wobec tego $\mathbb{E}(Y|X=x)$, dla $x\in\{0,1,2,\ldots\}$ jest średnią z rozkładu Bernoulliego z parametrami $n=x$ i $p=0,9$, czyli $0,9x$. Stąd $\mathbb{E}(Y|X)=0,9X$. Uwaga. To NIE JEST

to samo, co (błędne) stwierdzenie, że rozkład warunkowy Y pod warunkiem $X = x$ jest $0,9X$!
W końcu,

$$\mathbb{E}Y = \mathbb{E}(\mathbb{E}(Y \mid X)) = \mathbb{E}(0,9X) = 0,9\lambda = \frac{9}{40}.$$

5. Wykonujemy ciąg rzutów kostką do gry. Definiujemy:

(a) zmienne losowe X_n następująco: $X_n = 1$, jeśli w n-tym rzucie wypadła szóstka, i $X_n = 0$ w przeciwnym przypadku. Zbadać zbieżność ciągu $\frac{X_1+\ldots+X_n}{6n}$ prawie na pewno i wyznaczyć granicę, jeśli istnieje. (*4 pkt*)
Rozw. Zmienne X_1, X_2, $\ldots$ są niezależne, o identycznych rozkładach t.że $\mathbb{P}(X_n = 1) = \frac{1}{6}$, $\mathbb{P}(X_n = 0) = \frac{5}{6}$, a więc t. że $\mathbb{E}X_n$ istnieje i równe jest $\frac{1}{6}$. A zatem, ciąg X_1, X_2, $\ldots$ spełnia założenia MPWL, czyli $\frac{X_1+\ldots+X_n}{n} \xrightarrow{p.n.} \mathbb{E}X_1 = \frac{1}{6}$, a zatem ciąg z treści zadania jest zbieżny prawie na pewno, i jego granica wynosi

$$\frac{X_1 + \ldots + X_n}{6n} \xrightarrow{p.n.} \frac{1}{6 \cdot 6} = \frac{1}{36}.$$

(b) zmienne losowe Y_n następująco: $Y_n = 1$, jeśli w pierwszych n rzutach choć raz wypadła szóstka, oraz $Y_n = 0$ w przeciwnym przypadku. Zbadać zbieżność ciągu Y_n według prawdopodobieństwa i wyznaczyć granicę, jeśli istnieje. (*4 pkt*)
Rozw. Zmienne Y_n mają tę własność, iż $\mathbb{P}(Y_n = 0) = \mathbb{P}(Y_{n-1} = 0) \cdot \frac{1}{2} = \frac{1}{2^n} \to 0$, zaś $\mathbb{P}(Y_n = 1) = 1 - \frac{1}{2^n} \to 1$. A zatem, można się spodziewać, iż granica według prawdopodobieństwa dla tego ciągu istnieje, i równa jest 1. Istotnie, dla dowolnego $\varepsilon > 0$ mamy

$$\lim_{n\to\infty} \mathbb{P}(|Y_n - 1| > \varepsilon) \leqslant \lim_{n\to\infty} \mathbb{P}(Y_n = 0) = \lim_{n\to\infty} \frac{1}{2^n} = 0,$$

a więc

$$Y_n \xrightarrow{\mathbb{P}} 1.$$

6. Przypuśćmy, że co trzeci klient kasy w pewnym banku przychodzi do oddziału wypłacić gotówkę, zaś pozostali dokonują wpłat. Wyznaczyć przybliżone prawdopodobieństwo, że spośród 162 (niezależnych) klientów odwiedzających bank pewnego dnia, co najwyżej 66 wypłaci gotówkę (*4 pkt*).
Rozw. Oznaczmy przez X_n – zmienną losową równą 1, jeśli n-ty klient wypłaca gotówkę, i 0 w przeciwnym przypadku, oraz $S_n = X_1 + \ldots X_n$. Zmienne X_n są niezależne, o rozkładzie dwupunktowym: $\mathbb{P}(X_n = 1) = \frac{1}{3} = 1 - \mathbb{P}(X_n = 0)$. Można więc do tych zmiennych stosować tw. de Moivre'a - Laplace'a. Przybliżyć mamy $\mathbb{P}(S_{162} \leqslant 66)$, co przekształcimy do postaci, w której można skorzystać z tw. de Moivre'a -Laplace'a:

$$\mathbb{P}(S_{162} \leqslant 66) = \mathbb{P}\left(\frac{S_{162} - 162 \cdot \frac{1}{3}}{\sqrt{162 \cdot \frac{1}{3} \cdot \frac{2}{3}}} \leqslant \frac{66 - 54}{\sqrt{36}}\right) \approx \Phi\left(\frac{12}{6}\right) = \Phi(2) \approx 0,977 = 0,003.$$

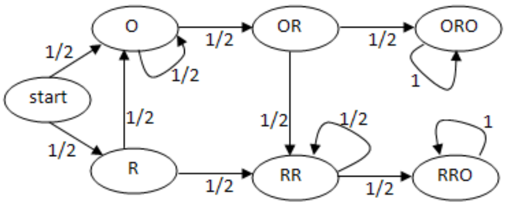
Przypuśćmy teraz, że klient wypłacający pobiera z kasy średnio 4000 PLN, z odchyleniem standardowym 600PLN, zaś klient wpłacający zostawia w kasie średnio 2000 PLN, z odchyleniem standardowym 800 PLN. Wyznaczyć przybliżone prawdopodobieństwo, że po dniu, w którym przy kasie było 100 klientów wpłacających i 100 klientów wypłacających gotówkę, kasa została uszczuplona o co najmniej 197 tys PLN. (*5 pkt*).
Rozw. Aby móc w prosty sposób zastosować centralne twierdzenie graniczne w sformułowaniu z wykładu do przybliżenia szukanego prawdopodobieństwa, warto stworzyć „sztuczne" zmienne o identycznych rozkładach. Jeśli więc przez X_i będziemy oznaczać wypłaty klientów wypłacających, a przez Y_i wpłaty klientów wpłacających, to zdefiniujmy pomocnicze zmienne $Z_i = X_i - Y_i$. Zmienne te mają następujące własności: $\mathbb{E}Z_i = \mathbb{E}(X_i - Y_i) = 4000 - 2000 = 2000$, zaś $VarZ_i = Var(X_i - Y_i) = VarX_i + VarY_i = 600^2 + 800^2 = 1000^2$ (z niezależności klientów). Szukane w zadaniu prawdopodobieństwo to $\mathbb{P}(Z_1 + \ldots + Z_{100} \geqslant 197000)$. Ponieważ zmienne Z_i są niezależne o identycznych rozkładach

posiadających wartość oczekiwaną i wariancję, więc można stosować do nich CTG; przekształcimy więc odpowiednio szukane prawdopodobieństwo:

$$\mathbb{P}\left(\frac{Z_1+\ldots+Z_{100}-100\cdot 2000}{\sqrt{100}\cdot 1000} \geqslant \frac{197000-200000}{10000}\right) \approx 1-\Phi\left(-\frac{3000}{10000}\right) = \Phi(0,3).$$

7. Jacek i Placek grają w następującą grę: rzucają symetryczną monetą, aż wypadnie ciąg Orzeł-Reszka-Orzeł, kiedy to wygrywa Jacek, lub ciąg Reszka-Reszka-Orzeł, kiedy to wygrywa Placek. Jaka jest szansa, że najpóźniej po pierwszych pięciu rzutach Jacek wygra? (*3 pkt*)
Rozw. Łatwo sprawdzić (zob. np. graf niżej), że sytuacje, które mogą doprowadzić do takiego wyniku odpowiadają ciągom: ORO, OORO, RORO, OORO, OOORO oraz ROORO. Prawdopodobieństwo, że wyrzucony zostanie jeden z nich wynosi $\frac{1}{8}+\frac{1}{16}+\frac{1}{16}+\frac{1}{32}+\frac{1}{32}=\frac{10}{32}$.

Jaka jest szansa wygranej Jacka w ogóle? (*5 pkt*)
Rozw. Najprościej sytuację opisać przy pomocy łańcucha Markowa, którego graf wyglądać może np. tak:



Wówczas pytanie o to, jaka jest szansa wygranej Jacka jest pytaniem o prawdopodobieństwo, iż łańcuch znajdujący się w stanie początkowym „Start" znajdzie się w stanie ORO zanim znajdzie się w stanie RRO. Jeśli przez p_i *oznaczymy prawdopodobieństwo, że ze stanu i-tego łańcuch dotrze do stanu ORO przed dotarciem do stanu RRO, to możemy zapisać układ równań:*

$$\begin{cases} p_{start} = \frac{1}{2}p_O + \frac{1}{2}p_R \\ p_O = \frac{1}{2}p_O + \frac{1}{2}p_{OR} \\ p_R = \frac{1}{2}p_O + \frac{1}{2}p_{RR} \\ p_{OR} = \frac{1}{2}p_{RR} + \frac{1}{2}p_{ORO} \\ p_{RR} = \frac{1}{2}p_{RR} + \frac{1}{2}p_{RRO} \\ p_{ORO} = 1 \\ p_{RRO} = 0 \end{cases},$$

którego rozwiązanie spełnia: $p_{RR}=0$, $p_{OR}=\frac{1}{2}$, $p_O=\frac{1}{2}$, $p_R=\frac{1}{4}$, *oraz interesujące nas* $p_{start}=\frac{3}{8}$.