Wstęp do rachunku prawdopodobieństwa
6. Funkcje zmiennej losowej. Zmienne losowe wielowymiarowe. Wartość oczekiwana, wariancja i kowariancja.

Zadanie 1. *Zmienna losowa X posiada rozkład podany w poniższej tabeli:*

k	0	1	2	3
$\mathbb{P}(X=k)$	0,4	0,3	0,2	0,1

a) *Oblicz wartość oczekiwaną oraz wariancję zmiennej losowej X.*

Bezpośrednio ze wzoru liczymy wartość oczekiwaną:

$$\mathbb{E}X = \sum_{k=0}^{3} k \cdot \mathbb{P}(X=k) = 0 \cdot 0{,}4 + 1 \cdot 0{,}3 + 2 \cdot 0{,}2 + 3 \cdot 0{,}1 = 1.$$

Do obliczenia wariancji skorzystamy ze wzoru:

$$\mathrm{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

W tym celu musimy wyznaczyć $\mathbb{E}X^2$:

$$\mathbb{E}X^2 = \sum_{k=0}^{3} k^2 \cdot \mathbb{P}(X=k) = 0^2 \cdot 0{,}4 + 1^2 \cdot 0{,}3 + 2^2 \cdot 0{,}2 + 3^2 \cdot 0{,}1 = 2.$$

Teraz wystarczy podstawić te wartości do wzoru powyżej skąd otrzymujemy:

$$\mathrm{Var}X = 2 - 1^2 = 1.$$

b) *Znajdź rozkład zmiennej losowej $Y = |2-X|$ i używając go znajdź jej wartość oczekiwaną.*

Na początek ustalmy zbiór wartości jakie może przyjmować zmienna losowa Y. Mamy:

$$|2-0| = 2, \quad |2-1| = 1, \quad |2-2| = 0, \quad |2-3| = 1.$$

Zatem zbiorem atomów zmiennej losowej Y jest zbiór $\{0,1,2\}$, przy czym

$$\begin{aligned}
\mathbb{P}(Y=0) &= \mathbb{P}(X=2) = 0{,}2\,,\\
\mathbb{P}(Y=1) &= \mathbb{P}(X=1) + \mathbb{P}(X=3) = 0{,}4\,,\\
\mathbb{P}(Y=2) &= \mathbb{P}(X=0) = 0{,}4\,.
\end{aligned}$$

Inny sposób przedstawienia funkcji masy prawdopodobieństwa tej zmiennej losowej to tabelka:

k	0	1	2
$\mathbb{P}(Y=k)$	0,2	0,4	0,4

Teraz możemy policzyć wartość oczekiwaną zmiennej losowej Y:

$$\mathbb{E}Y = 0 \cdot 0{,}2 + 1 \cdot 0{,}4 + 2 \cdot 0{,}4 = 1{,}2.$$

c) *Oblicz $\mathbb{E}|2-X|$ korzystając z „prawa leniwego statystyka”.*

Niech $f(X) = |2-X|$. Po podstawieniu do wzoru otrzymujemy:

$$\begin{aligned}
\mathbb{E}(f(X)) = \sum_{k=0}^{3} f(k) \cdot \mathbb{P}(X=k) &= f(0) \cdot 0{,}4 + f(1) \cdot 0{,}3 + f(2) \cdot 0{,}2 + f(3) \cdot 0{,}1\\
&= 2 \cdot 0{,}4 + 1 \cdot 0{,}3 + 0 \cdot 0{,}2 + 1 \cdot 0{,}1 = 1{,}2.
\end{aligned}$$

Zadanie 2. *Zmienna losowa X posiada rozkład podany w poniższej tabeli:*

i	-2	-1	0	1	2	7
$\mathbb{P}(X = i)$	0,18	0,16	0,28	0,23	0,14	0,01

Znajdź rozkład zmiennej losowej $Y = sgn(X)$ i oblicz $\mathbb{E}Y$ i $VarY$.

Przypomnijmy, że funkcja $sgn : \mathbb{R} \to \mathbb{R}$ zdefiniowana jest następująco:

$$sgn(x) = \begin{cases} -1 & \text{dla } x < 0\,, \\ 0 & \text{dla } x = 0\,, \\ 1 & \text{dla } x > 0\,. \end{cases}$$

A zatem zbiorem atomów zmiennej losowej Y jest zbiór $\{-1, 0, 1\}$, przy czym mamy np. $Y = -1$ wtedy i tylko wtedy, gdy $X \in \{-2, -1\}$. Rozkład zmiennej losowej Y przedstawia się następująco:

i	-1	0	1
$\mathbb{P}(Y = i)$	0,34	0,28	0,38

Możemy teraz wyznaczyć wartość oczekiwaną oraz wariancję zmiennej losowej Y:

$$\mathbb{E}Y = (-1) \cdot 0{,}34 + 0 \cdot 0{,}28 + 1 \cdot 0{,}38 = 0{,}04\,,$$

$$\mathbb{E}Y^2 = (-1)^2 \cdot 0{,}34 + 0^2 \cdot 0{,}28 + 1^2 \cdot 0{,}38 = 0{,}72\,,$$

$$VarY = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = 0{,}72 - 0{,}0016 = 0{,}7184\,.$$

Zadanie 3. *Zmienna losowa* (X, Y) *ma następujący rozkład:*

$$\mathbb{P}(X = 0, Y = -1) = 1/2, \; \mathbb{P}(X = 1, Y = 1) = \mathbb{P}(X = -1, Y = 1) = 1/4\,.$$

Wyznacz rozkłady brzegowe obu zmiennych losowych i znajdź $\rho(X, Y)$*. Czy te zmienne losowe są niezależne?*

Zacznijmy od przedstawienia rozkładu wektora losowego (X, Y) za pomocą tabelki:

$X \backslash Y$	-1	1
-1	0	$1/4$
0	$1/2$	0
1	0	$1/4$

Zauważmy, że zmienna losowa X przyjmuje tylko wartości ze zbioru $\{-1, 0, 1\}$, zaś zmienna losowa Y odpowiednio $\{-1, 1\}$. W celu wyznaczenia rozkładu brzegowego zmiennej losowej X sumujemy wartości w wierszach, to znaczy:

$$\begin{aligned}
\mathbb{P}(X = -1) &= \mathbb{P}(X = -1, Y = -1) + \mathbb{P}(X = -1, Y = 1) = 0 + \frac{1}{4} = \frac{1}{4}\,, \\
\mathbb{P}(X = 0) &= \mathbb{P}(X = 0, Y = -1) + \mathbb{P}(X = 0, Y = 1) = \frac{1}{2} + 0 = \frac{1}{2}\,, \\
\mathbb{P}(X = 1) &= \mathbb{P}(X = 1, Y = -1) + \mathbb{P}(X = 1, Y = 1) = 0 + \frac{1}{4} = \frac{1}{4}\,.
\end{aligned}$$

Podobnie dla wyznaczenia rozkładu brzegowego zmiennej losowej Y sumujemy wartości w kolumnach, czyli:

$$\begin{aligned}
\mathbb{P}(Y = -1) &= \mathbb{P}(X = -1, Y = -1) + \mathbb{P}(X = 0, Y = -1) + \mathbb{P}(X = 1, Y - 1) = 0 + \frac{1}{2} + 0 = \frac{1}{2}\,, \\
\mathbb{P}(Y = 1) &= \mathbb{P}(X = -1, Y = 1) + \mathbb{P}(X = 0, Y = 1) + \mathbb{P}(X = 1, Y = 1) = \frac{1}{4} + 0 + \frac{1}{4} = \frac{1}{2}\,.
\end{aligned}$$

Podsumowując, rozkłady brzegowe obu zmiennych losowych przedstawione są w poniższych tabelkach:

x	-1	0	1
$\mathbb{P}(X = x)$	$1/4$	$1/2$	$1/4$

y	-1	1
$\mathbb{P}(Y = y)$	$1/2$	$1/2$

Przypomnijmy następnie wzór na współczynnik korelacji wektora losowego (X, Y):

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}X}\sqrt{\text{Var}Y}},$$

gdzie wzór na kowariancję wektora losowego (X, Y) przedstawia się następująco:

$$\text{Cov}(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}XY - \mathbb{E}X\,\mathbb{E}Y.$$

A zatem na początek musimy wyznaczyć kowariancję. Mamy:

$$\mathbb{E}XY = (-1) \cdot 1 \cdot \frac{1}{4} + 1 \cdot 1 \cdot \frac{1}{4} = 0\,.$$

Proszę zauważyć, że w powyższych obliczeniach pomijamy wszystkie przypadki, gdzie $\mathbb{P}(X = x, Y = y) = 0$, $x = 0$ lub $y = 0$. Podobnie mamy:

$$\mathbb{E}X = (-1) \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} = 0\,.$$

Ostatecznie otrzymujemy:

$$\text{Cov}(X, Y) = 0 - 0 \cdot \mathbb{E}Y = 0\,.$$

Ponieważ kowariancja się zeruje, nie musimy już wyznaczać wariancji zmiennych losowych X i Y (w definicji współczynnika korelacji w liczniku mamy wówczas 0). Wystarczy zauważyć, że obie zmienne losowe mają co najmniej dwa atomy, a zatem nie mają rozkładu jednopunktowego, czyli $\text{Var}X > 0$ i $\text{Var}Y > 0$. Ostatecznie, współczynnik korelacji wynosi

$$\rho(X, Y) = 0.$$

Mogłoby się wydawać, że skoro $\mathrm{Cov}(X,Y) = 0$, to zmienne losowe X, Y są niezależne. Okazuje się jednak, że w tym przypadku nie jest to prawdą (proszę pamiętać, że prawdziwa jest jedynie implikacja w drugą stronę!). Aby pokazać, że zmienne losowe X, Y nie są niezależne, wystarczy wskazać parę wartości (x, y), dla której

$$\mathbb{P}\left(X = x, Y = y\right) \neq \mathbb{P}\left(X = x\right)\mathbb{P}\left(Y = y\right).$$

Najlepszym sposobem na znalezienie takiego kandydata jest sprawdzenie tych par (x, y), dla których

$$\mathbb{P}\left(X = x, Y = y\right) = 0.$$

Weźmy zatem $x = -1, y = -1$. Mamy:

$$\mathbb{P}\left(X = -1, Y = -1\right) = 0 \neq \frac{1}{8} = \frac{1}{4} \cdot \frac{1}{2} = \mathbb{P}\left(X = -1\right)\mathbb{P}\left(Y = -1\right),$$

skąd wnioskujemy, że rzeczywiście zmienne losowe X i Y nie są niezależne.

Zadanie 4. *Rzucamy symetryczną kostką n razy. Czy liczba wszystkich wyrzuconych jedynek X_n i liczba wszystkich wyrzuconych dwójek Y_n są niezależnymi zmiennymi losowymi? (Spróbuj najpierw zgadnąć odpowiedź i uzasadnić ją bez dokonywanie obliczeń, a następnie przeprowadź formalny dowód żeby potwierdzić swoją tezę.)*

Intuicja podpowiada, że zmienne losowe X_n i Y_n nie są niezależne. Rzeczywiście, jeśli np. zmienna losowa X_n przyjmuje wartość n, oznacza to, że w każdym rzucie wypadła jedynka, a zatem automatycznie zmienna losowa Y_n musi przyjąć wartość zero. A zatem wartość zmiennej losowej X_n w niektórych przypadkach może determinować wartość zmiennej losowej Y_n.

Udowodnijmy teraz formalnie, że te zmienne losowe nie są niezależne. W tym celu wystarczy rozważyć zdarzenie $\{X_n = n, Y_n = n\}$. Z jednej strony zauważmy, że

$$\mathbb{P}\left(X_n = n,\, Y_n = n\right) = 0,$$

ponieważ nie możemy w n rzutach otrzymać jednocześnie n razy jedynkę i n razy dwójkę. Z drugiej strony

$$\mathbb{P}\left(X_n = n\right) = \mathbb{P}\left(Y_n = n\right) = \left(\frac{1}{6}\right)^n.$$

Stąd otrzymujemy, że

$$\mathbb{P}\left(X_n = n, Y_n = n\right) = 0 \neq \left(\frac{1}{6}\right)^{2n} = \left(\frac{1}{6}\right)^n \cdot \left(\frac{1}{6}\right)^n = \mathbb{P}\left(X_n = n\right)\mathbb{P}\left(Y_n = n\right),$$

a zatem zmienne losowe X_n i Y_n nie są niezależne.

Zadanie 5. *W pierwszej urnie znajduje się 5 kul białych i 3 czarne, a w drugiej 2 białe i 2 czarne. Z pierwszej urny losujemy jedną kulę, a z drugiej również jedną kulę. Niech X będzie zmienną losową oznaczającą liczbę kul białych wśród dwóch wylosowanych kul, a Y liczbę wylosowanych kul czarnych.*

a) *Znajdź rozkład łączny wektora losowego (X, Y).*

Zauważmy, że skoro losujemy dwie kule, to zmienne losowe X i Y mają ten sam zbiór atomów równy $\{0, 1, 2\}$. Co więcej, ponieważ w urnach mamy tylko białe i czarne kule, prawdą jest, że $X + Y = 2$. Mamy zatem:

$$\mathbb{P}(X = 0, Y = 0) = \mathbb{P}(X = 0, Y = 1) = \mathbb{P}(X = 1, Y = 0) =$$
$$\mathbb{P}(X = 1, Y = 2) = \mathbb{P}(X = 2, Y = 1) = \mathbb{P}(X = 2, Y = 2) = 0\,.$$

Dla pozostałych wartości rozkład łączny wygląda następująco:

$$\mathbb{P}(X = 0, Y = 2) = \frac{\binom{3}{1}\binom{2}{1}}{\binom{8}{1}\binom{4}{1}} = \frac{3}{16}\,,$$

$$\mathbb{P}(X = 1, Y = 1) = \frac{\binom{3}{1}\binom{2}{1} + \binom{5}{1}\binom{2}{1}}{\binom{8}{1}\binom{4}{1}} = \frac{1}{2}\,,$$

$$\mathbb{P}(X = 2, Y = 0) = \frac{\binom{5}{1}\binom{2}{1}}{\binom{8}{1}\binom{4}{1}} = \frac{5}{16}\,.$$

Możemy też przedstawić rozkład łączny w tabelce, co ułatwi nam dalsze obliczenia i pozwoli od razu stwierdzić jak wyglądają rozkłady brzegowe zmiennych losowych występujących w wektorze losowym (X, Y).

X \ Y	0	1	2
0	0	0	3/16
1	0	1/2	0
2	5/16	0	0

b) *Czy zmienne losowe X i Y są niezależne?*

Zmienne losowe X i Y nie są niezależne. Mamy np.

$$\mathbb{P}(X = 0, Y = 0) = 0 \neq \frac{3}{16} \cdot \frac{5}{16} = \mathbb{P}(X = 0)\,\mathbb{P}(Y = 0)\,.$$

c) *Oblicz $\mathbb{E}X$ i $\mathbb{E}Y$ korzystając z rozkładów brzegowych zmiennych losowych X i Y.*

Z tabelki rozkładu łącznego łatwo odczytamy rozkłady brzegowe zmiennych losowych X i Y:

x	0	1	2
$\mathbb{P}(X = x)$	3/16	1/2	5/16

y	0	1	2
$\mathbb{P}(Y = y)$	5/16	1/2	3/16

A zatem:

$$\mathbb{E}X = 0 \cdot \frac{3}{16} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{5}{16} = \frac{9}{8}\,,$$

$$\mathbb{E}Y = 0 \cdot \frac{5}{16} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{3}{16} = \frac{7}{8}\,.$$

d) *Oblicz $\mathbb{E}X$ przedstawiając X w postaci sumy $X = X_1 + X_2$, gdzie X_i oznacza liczbę kul białych wyciągniętych z i-tej urny.*

Zauważmy, że zmienne losowe X_1 i X_2 przyjmują tylko wartości 0 lub 1, przy czym:

$$\mathbb{P}\left(X_1 = 1\right) = \frac{\binom{5}{1}}{\binom{8}{1}} = \frac{5}{8}, \ \ \mathbb{P}\left(X_2 = 1\right) = \frac{\binom{2}{1}}{\binom{4}{1}} = \frac{1}{2}.$$

Jak nietrudno zauważyć, mamy wówczas:

$$\mathbb{E}X_1 = \mathbb{P}\left(X_1 = 1\right) = \frac{5}{8}, \ \ \mathbb{E}X_2 = \mathbb{P}\left(X_2 = 1\right) = \frac{1}{2}.$$

Z liniowości wartości oczekiwanej zachodzi:

$$\mathbb{E}X = \mathbb{E}(X_1 + X_2) = \mathbb{E}X_1 + \mathbb{E}X_2 = \frac{9}{8}.$$

e) *Co można powiedzieć o rozkładzie zmiennej losowej $Z = X + Y$?*

Zmienna losowa Z jest zawsze równa 2, czyli posiada dokładnie jeden atom oraz $\mathbb{P}\left(Z = 2\right) = 1$.

f) *Znajdź współczynnik korelacji $\rho(X, Y)$. Czy można to zrobić prościej korzystając z podpunktu e) i nie używając rozkładu łącznego (X, Y)?*

Zanim przejdziemy do rachunków, zauważmy, że na podstawie „kształtu" tabelki rozkładu łącznego możemy przewidzieć, że współczynnik korelacji będzie równy -1, to znaczy istnieje zależność liniowa pomiędzy zmiennymi losowymi X i Y. Rzeczywiście, jak zauważyliśmy zachodzi równanie $X+Y = 2$, a zatem $Y = 2-X$, co implikuje, że współczynnik korelacji wynosi $\rho(X, Y) = -1$.

Policzymy teraz ten współczynnik rachunkowo. Korzystając z tego, że $Y = 2 - X$, oraz z liniowości wartości oczekiwanej, wyznaczamy kolejno:

$$\mathbb{E}XY = \mathbb{E}(X(2 - X)) = \mathbb{E}(2X - X^2) = 2\,\mathbb{E}X - \mathbb{E}X^2\,,$$

$$\mathbb{E}X\,\mathbb{E}Y = \mathbb{E}X\,\mathbb{E}(2 - X) = \mathbb{E}X \cdot (2 - \mathbb{E}X) = 2\mathbb{E}X - (\mathbb{E}X)^2\,.$$

Po odjęciu drugiego wiersza od pierwszego, otrzymujemy:

$$\mathrm{Cov}(X, Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = (\mathbb{E}X)^2 - \mathbb{E}X^2 = -\mathrm{Var}X\,.$$

Ponadto, korzystając znów z liniowości wartości oczekiwanej mamy:

$$\begin{aligned}\mathrm{Var}Y &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = \mathbb{E}(X - 2)^2 - (\mathbb{E}(2 - X))^2 = \mathbb{E}(X^2 - 4X + 4) - (2 - \mathbb{E}X)^2 \\ &= \mathbb{E}X^2 - 4\,\mathbb{E}X + 4 - (4 - 4\,\mathbb{E}X + (\mathbb{E}X)^2) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathrm{Var}X\,.\end{aligned}$$

Ostatecznie:

$$\rho(X, Y) = \frac{-\mathrm{Var}X}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}X}} = -1\,.$$