# Wstęp do rachunku prawdopodobieństwa
## 7. Parametry rozkładów dyskretnych

### Dodatek A. Zadania na ćwiczenia

**Zadanie A.1.** *Wiedząc, że* $\mathbb{E}(X) = 1$ *i* $Var(X) = 5$, *znajdź* $\mathbb{E}((2+X)^2)$ *oraz* $Var(4+3X)$.

Skorzystamy z liniowości wartości oczekiwanej:

$$\mathbb{E}((2+X)^2) = \mathbb{E}(4+4X+X^2) = 4 + 4\mathbb{E}X + \mathbb{E}(X^2).$$

Wiemy, że $\mathbb{E}X = 1$, ale musimy jeszcze obliczyć $\mathbb{E}X^2$. W tym celu wykorzystamy wzór na wariancję:

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

Przekształcając ten wzór otrzymujemy:

$$\mathbb{E}(X^2) = \mathrm{Var}X + (\mathbb{E}X)^2 = 5 + 1^2 = 6.$$

Zatem mamy:

$$\mathbb{E}((2+X)^2) = 4 + 4 + 6 = 14.$$

Obliczymy teraz $\mathrm{Var}(4+3X)$ korzystając z własności wariancji:

$$\mathrm{Var}(4+3X) = \mathrm{Var}(3X) = 3^2 \cdot \mathrm{Var}(X) = 9 \cdot 5 = 45.$$

**Zadanie A.2.** *Roztargniona sekretarka włożyła losowo 10 zaadresowanych listów do 10 zaadresowanych kopert (każdy list do innej koperty). Oblicz wartość oczekiwaną i wariancję liczby listów, które trafiły do swoich adresatów.*

Niech $X$ będzie zmienną losową oznaczającą liczbę listów, które trafiły do swoich adresatów. W celu obliczenia wartości oczekiwanej i wariancji tej zmiennej losowej skorzystamy ze zmiennych losowych indykatorowych, które definiujemy w następujący sposób: dla $i = 1, 2, \ldots, 10$ zmienna losowa $X_i$ jest równa 1, jeśli $i$-ty list trafił do swojego adresata oraz 0 w przeciwnym przypadku. Zauważmy, że wtedy

$$X = X_1 + X_2 + \ldots + X_{10}.$$

Ponadto, dla tak zdefiniowanych zmiennych losowych indykatorowych nie jest trudno wyznaczyć ich rozkład i policzyć ich wartość oczekiwaną oraz wariancję. Rzeczywiście, dla każdego $i = 1, 2, \ldots, 10$ mamy:

$$\mathbb{P}(X_i = 1) = \frac{1}{10} \quad \text{oraz} \quad \mathbb{P}(X_i = 0) = \frac{9}{10},$$
$$\mathbb{E}X_i = 1 \cdot \mathbb{P}(X_i = 1) = \frac{1}{10},$$
$$\mathbb{E}(X_i^2) = 1^2 \cdot \mathbb{P}(X_i = 1) = \frac{1}{10},$$
$$\mathrm{Var}X_i = \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2 = \frac{1}{10} - \frac{1}{100} = \frac{9}{100}.$$

Możemy teraz skorzystać z liniowości wartości oczekiwanej aby obliczyć wartość oczekiwaną zmiennej losowej $X$:

$$\mathbb{E}X = \mathbb{E}(X_1 + \ldots + X_{10}) = \mathbb{E}X_1 + \ldots + \mathbb{E}X_{10} = 10 \cdot \frac{1}{10} = 1.$$

Pozostaje nam do obliczenia $\mathrm{Var}X$. Skorzystamy w tym celu z twierdzenia o wariancji sumy zmiennych losowych. Musimy na początku policzyć kowariancję zmiennych losowych indykatorowych. Niech $1 \leqslant i < j \leqslant 10$. Proszę zwrócić uwagę, że dla $i \neq j$ wektor losowy $(X_i, X_j)$ ma tylko cztery atomy: $(0,0)$, $(0,1)$, $(1,0)$ i $(1,1)$. Co więcej, do wyznaczenia $\mathrm{Cov}(X_i, X_j)$ potrzebne jest nam jedynie $\mathbb{P}(X_i = 1, X_j = 1)$. Mamy zatem:

$$\mathbb{P}(X_i = 1, X_j = 1) = \mathbb{P}(X_i = 1)\,\mathbb{P}(X_j = 1 | X_i = 1) = \frac{1}{10} \cdot \frac{1}{9} = \frac{1}{90},$$
$$\mathbb{E}(X_i X_j) = 1 \cdot 1 \cdot \mathbb{P}(X_i = 1, X_j = 1) = \frac{1}{90},$$
$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}X_i \mathbb{E}X_j = \frac{1}{90} - \frac{1}{100} = \frac{1}{900}.$$

Ostatecznie otrzymujemy:

$$\mathrm{Var}X = \mathrm{Var}\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} \mathrm{Var}X_i \; + \; 2 \sum_{1 \leqslant i < j \leqslant 10} \mathrm{Cov}(X_i, X_j) = 10 \cdot \frac{9}{100} + 2 \cdot \binom{10}{2} \cdot \frac{1}{900} = 1.$$

**Zadanie A.3.** *Rzucamy 100 razy trzema kostkami. Wyznacz wartość oczekiwaną i wariancję sumy wyrzuconych oczek.*

Niech $X$ będzie zmienną losową oznaczającą sumę wyrzuconych oczek. Na początku zauważmy, że jeśli interesuje nas tylko suma wyrzuconych oczek, to o 100 rzutach trzema kostkami możemy też myśleć jako o 300 rzutach jedną kostką. Zdefiniujmy zatem pomocnicze zmienne losowe $X_1, X_2, \ldots, X_{300}$, gdzie dla $i = 1, 2, \ldots, 300$ zmienna losowa $X_i$ oznacza liczbę oczek w $i$-tym rzucie. Wtedy

$$X = X_1 + X_2 + \ldots + X_{300}.$$

Podobnie jak w poprzednich zadaniach zaczniemy od wyznaczenia rozkładu i parametrów zmiennych losowych pomocniczych. Niech $i \in \{1, 2, \ldots, 300\}$. Wtedy:

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $\mathbb{P}(X_i = k)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

$$\begin{aligned} \mathbb{E}X_i &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3\frac{1}{2}\,, \\ \mathbb{E}(X_i^2) &= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = 15\frac{1}{6}\,, \\ \mathrm{Var}X_i &= \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2 = 15\frac{1}{6} - \left(3\frac{1}{2}\right)^2 = 2\frac{11}{12}\,. \end{aligned}$$

Możemy teraz skorzystać z liniowości wartości oczekiwanej aby wyznaczyć wartość oczekiwaną:

$$\mathbb{E}X = \mathbb{E}X_1 + \mathbb{E}X_2 + \ldots + \mathbb{E}X_{300} = 300 \cdot 3\frac{1}{2} = 1050.$$

Zauważmy ponadto, że kolejne rzuty kostką są od siebie niezależne, a więc także parami nieskorelowane, co oznacza, że wariancja sumy zmiennych losowych $X_i$ równa jest sumie wariancji, skąd otrzymujemy:

$$\mathrm{Var}X = \mathrm{Var}X_1 + \mathrm{Var}X_2 + \ldots + \mathrm{Var}X_{300} = 300 \cdot 2\frac{11}{12} = 875.$$

**Zadanie A.4.** *Niech $X$ będzie liczbą jedynek, a $Y$ liczbą dwójek otrzymanych w wyniku $n$ rzutów wyważoną kostką. Oblicz $\rho(X, Y)$.*

Ponownie skorzystamy ze zmiennych losowych indykatorowych. Dla $i = 1, 2, \ldots, n$ niech $X_i$ będzie zmienną losową, która przyjmuje wartość 1, gdy w $i$-tym rzucie wypada jedynka i 0 w przeciwnym przypadku, a $Y_i$ będzie zmienną losową, która przyjmuje wartość 1, gdy w $i$-tym rzucie wypada dwójka i 0 w przeciwnym przypadku. Wtedy:

$$X = X_1 + \ldots + X_n \quad \text{oraz} \quad Y = Y_1 + \ldots + Y_n.$$

Przypomnijmy wzór na współczynnik korelacji:

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

Zacznijmy od obliczenia wartości oczekiwanych i wariancji zmiennych losowych $X$ i $Y$, korzystając z ich przedstawienia za pomocą zmiennych losowych indykatorowych. Dla $i = 1, 2, \ldots, n$ mamy:

$$\begin{aligned}
\mathbb{P}\left(X_i = 1\right) &= \frac{1}{6},\\
\mathbb{E}X_i = 1 \cdot \mathbb{P}\left(X_i = 1\right) &= \frac{1}{6},\\
\mathbb{E}(X_i^2) = 1^2 \cdot \mathbb{P}\left(X_i = 1\right) &= \frac{1}{6},\\
\mathrm{Var}X_i = \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2 &= \frac{1}{6} - \frac{1}{36} = \frac{5}{36}.
\end{aligned}$$

Zatem:

$$\mathbb{E}X = \mathbb{E}X_1 + \ldots + \mathbb{E}X_n = \frac{n}{6}.$$

Ponadto zmienne losowe $X_1, \ldots, X_n$ są parami niezależne, a więc także parami nieskorelowane. To oznacza, że:

$$\mathrm{Var}X = \mathrm{Var}X_1 + \ldots + \mathrm{Var}X_n = \frac{5n}{36}.$$

W podobny sposób obliczamy odpowiednie parametry dla zmiennych losowych $Y_1, \ldots, Y_n$, otrzymując ostatecznie:

$$\mathbb{E}Y = \frac{n}{6} \quad \text{oraz} \quad \mathrm{Var}Y = \frac{5n}{36}.$$

Pozostaje nam do policzenia kowariancja $\mathrm{Cov}(X, Y)$. Wartość oczekiwaną $\mathbb{E}(XY)$ obliczymy korzystając ze zmiennych losowych indykatorowych i liniowości wartości oczekiwanej:

$$\mathbb{E}(XY) = \mathbb{E}(X_1 + \ldots + X_n)(Y_1 + \ldots + Y_n) = \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}X_i Y_j.$$

Okazuje się, że składniki tak otrzymanej sumy mogą przyjmować dwie wartości:

- Załóżmy najpierw, że $i = j$. Wtedy $\mathbb{P}\left(X_i = 1, Y_i = 1\right) = 0$ (bo w $i$-tym rzucie nie może jednocześnie wypaść jedynka i dwójka). Zatem:
$$\mathbb{E}(X_i Y_i) = 1 \cdot 1 \cdot \mathbb{P}\left(X_i = 1, Y_i = 1\right) = 0.$$

- Załóżmy teraz, że $i \neq j$. Wtedy $\mathbb{P}\left(X_i = 1, Y_j = 1\right) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$. Zatem:
$$\mathbb{E}(X_i Y_i) = 1 \cdot 1 \cdot \mathbb{P}\left(X_i = 1, Y_i = 1\right) = \frac{1}{36}.$$

Widzimy, że w sumie podwójnej powyżej niezerowe składniki otrzymujemy tylko w drugim przypadku, czyli dla $i \neq j$. Stąd:

$$\mathbb{E}(XY) = \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \frac{1}{36} = \frac{n(n-1)}{36}.$$

Możemy teraz obliczyć kowariancję $\mathrm{Cov}(X, Y)$:

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y = \frac{n(n-1)}{36} - \frac{n}{6} \cdot \frac{n}{6} = -\frac{n}{36}.$$

Ostatecznie otrzymujemy:

$$\rho(X,Y) = \frac{-\frac{n}{36}}{\sqrt{\frac{5n}{36}}\sqrt{\frac{5n}{36}}} = -\frac{1}{5}.$$

Jak możemy zinterpretować ten wynik? Ponieważ współczynnik korelacji jest ujemny, oznacza to, że wraz z rosnącymi wartościami zmiennej losowej $X$, wartości zmiennej losowej $Y$ powinny maleć, co oczywiście ma sens, bo im więcej jedynek wyrzucimy w $n$ rzutach kostką, tym mniej możemy wyrzucić dwójek.