Wstęp do rachunku prawdopodobieństwa
10. Nierówności probabilistyczne

Zadanie 1. *Czas pracy X pewnego urządzenia (liczony w dniach) ma rozkład dany dystrybuantą $F(x) = 1 - e^{-x/400}$ dla $x \geqslant 0$ i $F(x) = 0$ dla $x \leqslant 0$.*

(a) *Z jakim rozkładem prawdopodobieństwa mamy do czynienia?*

Na podstawie dystrybuanty widzimy , że zmienna losowa X ma rozkład wykładniczy z parametrem $\lambda = \frac{1}{400}$.

(b) *Znajdź prawdopodobieństwo, że urządzenie będzie działało przez rok (nieprzestępny).*

Urządzenie będzie działało przez rok (nieprzestępny), jeżeli będzie działało przez co najmniej 365 dni. Zatem musimy obliczyć $\mathbb{P}(X \geqslant 365)$, korzystając np. z dystrybuanty zmiennej losowej X:

$$\mathbb{P}(X \geqslant 365) = 1 - \mathbb{P}(X < 365) = 1 - \mathbb{P}(X \leqslant 365) = 1 - F(365) = 1 - (1 - e^{-365/400}) = e^{-73/80} \approx 0{,}4$$

(wykorzystaliśmy tutaj fakt, że podana dystrybuanta jest funkcją ciągłą).

(c) *Wyznacz gęstość f_X zmiennej losowej X.*

Gęstość f zmiennej losowej X wyznaczymy korzystając ze wzoru $f(x) = F'(x)$. Mamy zatem:

$$f(x) = \begin{cases} \frac{1}{400} e^{-x/400} & \text{dla } x \geqslant 0, \\ 0 & \text{w pozostałych przypadkach.} \end{cases}$$

(d) *Ile wynosi wartość oczekiwana oraz wariancja zmiennej losowej X?*

Ponieważ zmienna losowa X ma rozkład wykładniczy z parametrem $\lambda = \frac{1}{400}$, możemy od razu wyznaczyć jej wartość oczekiwaną oraz wariancję:

$$\mathbb{E}X = \frac{1}{\lambda} = 400, \quad \mathrm{Var}X = \frac{1}{\lambda^2} = 400^2 = 160000.$$

(e) *Oblicz dokładną wartość $\mathbb{P}(X \geqslant 1000)$, a następnie oszacuj to prawdopodobieństwo, korzystając z nierówności Markowa, a potem z nierówności Czebyszewa.*

Korzystając z dystrybuanty zmiennej losowej X otrzymujemy:

$$\mathbb{P}(X \geqslant 1000) = 1 - \mathbb{P}(X < 1000) = 1 - F(1000) = 1 - (1 - e^{-1000/400}) = e^{-5/2} \approx 0{,}082.$$

Następnie porównamy ten wynik z oszacowaniami, które dostajemy z nierówności Markowa oraz Czebyszewa-Bienaymé. Zwróćmy najpierw uwagę, że zmienna losowa X przyjmuje wartości ujemne z zerowym prawdopodobieństwem (bo $f(x) = 0$ dla $x < 0$). Oznacza to, że $\mathbb{P}(X \geqslant 1000) = \mathbb{P}(|X| \geqslant 1000)$ oraz $\mathbb{E}X = \mathbb{E}|X|$. Zatem z nierówności Markowa dostajemy:

$$\mathbb{P}(X \geqslant 1000) = \mathbb{P}(|X| \geqslant 1000) \leqslant \frac{\mathbb{E}|X|}{1000} = \frac{400}{1000} = \frac{2}{5} = 0{,}4.$$

Oszacujemy teraz to prawdopodobieństwo korzystając z nierówności Czebyszewa-Bienaymé. Zanim to zrobimy zauważmy, że zdarzenie $X - \mathbb{E}X \geqslant 600$ jest podzbiorem zdarzenia $|X - \mathbb{E}X| \geqslant 600$. Mamy zatem:

$$\mathbb{P}(X \geqslant 1000) = \mathbb{P}(X - \mathbb{E}X \geqslant 1000 - 400) \leqslant \mathbb{P}(|X - \mathbb{E}X| \geqslant 600) \leqslant \frac{\mathrm{Var}X}{600^2} = \frac{160000}{360000} = \frac{4}{9} \approx 0{,}44.$$

Jak widać w tym przypadku oba oszacowania są dalekie od właściwego wyniku i porównywalnie dobre, choć oszacowanie z nierówności Markowa jest nieznacznie lepsze.

Zadanie 2. *Na pewnym poznańskim skrzyżowaniu zdarza się średnio* 0,9 *wypadków rocznie. Oblicz dokładnie prawdopodobieństwo, że w przyszłym roku na tym skrzyżowaniu zajdą co najmniej trzy wypadki, a następnie oszacuj to prawdopodobieństwo używając nierówności Markowa i nierówności Czebyszewa.*

Niech X będzie zmienną losową oznaczającą liczbę wypadków w przyszłym roku. Musimy najpierw ustalić, jaki rozkład ma ta zmienna losowa. Wiemy, że wypadki na skrzyżowaniu zdarzają się tylko od czasu do czasu, a ich średnia to 0,9 wypadków na rok, więc możemy przyjąć, że jest to zdarzenie „rzadkie". Co za tym idzie możemy założyć, że X ma rozkład Poissona z parametrem 0,9 (bo przyjmujemy, że średnia liczba wypadków jest wartością oczekiwaną $\mathbb{E}X = \lambda$ zmiennej losowej X). W dalszej części zadania skorzystamy z własności rozkładu Poissona.

Policzymy najpierw dokładnie prawdopodobieństwo, że na tym skrzyżowaniu zajdą co najmniej trzy wypadki. Zauważmy, że dla $X \sim \text{Po}(0{,}9)$ mamy:

$$\mathbb{P}\left(X = k\right) = \frac{0{,}9^k}{k!} e^{-0{,}9}.$$

Zatem

$$\begin{aligned}\mathbb{P}\left(X \geqslant 3\right) &= 1 - \mathbb{P}\left(X < 3\right) = 1 - \left(\mathbb{P}\left(X = 0\right) + \mathbb{P}\left(X = 1\right) + \mathbb{P}\left(X = 2\right)\right) \\ &= 1 - \left(\frac{(0{,}9)^0}{0!} e^{-0{,}9} + \frac{(0{,}9)^1}{1!} e^{-0{,}9} + \frac{(0{,}9)^2}{2!} e^{-0{,}9}\right) = 1 - 2{,}305 e^{-0{,}9} \approx 0{,}063.\end{aligned}$$

Teraz oszacujemy to prawdopodobieństwo z nierówności Markowa. Podobnie jak w poprzednim zadaniu możemy założyć, że zmienna losowa X nie przyjmuje ujemnych wartości, zatem $\mathbb{P}\left(X \geqslant 3\right) = \mathbb{P}\left(|X| \geqslant 3\right)$ oraz $\mathbb{E}|X| = \mathbb{E}X = 0{,}9$. W związku z tym mamy następujące oszacowanie:

$$\mathbb{P}\left(X \geqslant 3\right) \leqslant \frac{0{,}9}{3} = 0{,}3.$$

Zanim skorzystamy z nierówności Czebyszewa-Bienaymé zauważmy, że wariancja zmiennej losowej X dana jest wzorem $\text{Var}X = \lambda = 0{,}9$. A zatem:

$$\mathbb{P}\left(X \geqslant 3\right) = \mathbb{P}\left(X - \mathbb{E}X \geqslant 3 - 0{,}9\right) \leqslant \mathbb{P}\left(|X - \mathbb{E}X| \geqslant 2{,}1\right) \leqslant \frac{\text{Var}X}{2{,}1^2} = \frac{0{,}9}{4{,}41} = \frac{10}{49} \approx 0{,}204.$$

Widzimy, że po raz kolejny oszacowania z nierówności nie dały nam aż tak dobrych wyników, natomiast w tym przypadku lepsze okazało się oszacowanie z nierówności Czebyszewa-Bienaymé.

Zadanie 3. *Pięć stacji pomiarowych na pięciu różnych kontynentach rejestruje wysokoenergetyczne cząstki kosmiczne. Liczba cząstek zarejestrowanych rocznie przez każdą ze stacji ma rozkład Poissona ze średnią* 3,6*. Znajdź prawdopodobieństwo, że dokładnie trzy z pięciu stacji zarejestruje więcej niż 3 cząstki w następnym roku.*

Zastanówmy się najpierw jakie są szanse na to, że jedna stacja zarejestruje więcej niż 3 cząstki w następnym roku. Niech Y będzie zmienną losową zliczająca cząstki zarejestrowane przez pojedynczą stację w następnym roku. Wiemy, że Y ma rozkład Poissona z parametrem $\lambda = 3{,}6$ (bo średnia liczba zarejestrowanych cząstek to 3,6). Zatem dla $k = 0, 1, 2, \ldots$ mamy:

$$\mathbb{P}\left(Y = k\right) = \frac{3{,}6^k}{k!} e^{-3{,}6}.$$

Korzystając z tych informacji obliczamy interesujące nas prawdopodobieństwo:

$$\begin{aligned}\mathbb{P}\left(Y > 3\right) &= 1 - \mathbb{P}\left(Y \leqslant 3\right) = 1 - \left(\mathbb{P}\left(Y = 0\right) + \mathbb{P}\left(Y = 1\right) + \mathbb{P}\left(Y = 2\right) + \mathbb{P}\left(Y = 3\right)\right) \\ &= 1 - e^{-3{,}6}\left(\frac{3{,}6^0}{0!} + \frac{3{,}6}{1!} + \frac{3{,}6^2}{2!} + \frac{3{,}6^3}{3!}\right) = 1 - 18{,}856e^{-3{,}6} \approx 0{,}485.\end{aligned}$$

W zadaniu pytamy o szanse na to, że dokładnie trzy z pięciu stacji zarejestrują więcej niż 3 cząstki. O tym doświadczeniu losowym możemy myśleć w następujący sposób: jest to pięciokrotne powtórzenie eksperymentu losowego (po jednym powtórzeniu na każdą stację), w którym interesuje nas czy zajdzie sukces (czyli stacja zarejestruje więcej niż 3 cząstki w następnym roku) czy też porażka (czyli stacja zarejestruje co najwyżej 3 cząstki w przyszłym roku). Ponadto możemy założyć, że stacje rejestrują cząstki niezależnie. Zatem mamy do czynienia ze schematem Bernoulliego, gdzie mamy $n = 5$ prób, a prawdopodobieństwo sukcesu w jednej próbie to $p = \mathbb{P}\left(Y > 3\right)$. Oznacza to, że szukane prawdopodobieństwo jest równe:

$$\tau_3 = \binom{5}{3} p^3 (1-p)^2 \approx \binom{5}{3} (0{,}485)^3 (1 - 0{,}485)^2 \approx 0{,}302.$$

Zadanie 4. *W pewnym urzędzie czas obsługi petenta (liczony w minutach) jest zmienną losową X o parametrach $\mathbb{E}X = 45$ i $VarX = 15$. Zakładamy, że czasy obsługi różnych petentów są niezależne.*

(a) *Załóżmy, że w urzędzie pojawiło się 50 petentów. Oszacuj z dołu szansę, że średni czas ich obsługi (czyli średnia arytmetyczna z czasów ich obsługi) będzie wynosił od 35 do 55 minut (nierówności ostre)?*

Niech Y będzie zmienną losową oznaczającą łączny czas obsługi 50 petentów. Chcemy oszacować prawdopodobieństwo, że średni czas obsługi będzie wynosił od 35 do 55 minut, czyli $\mathbb{P}\left(35 < \frac{Y}{50} < 55\right)$. Znamy tylko parametry związane z czasem obsługi pojedynczego petenta, więc przedstawimy zmienną losową Y jako sumę pomocniczych zmiennych losowych $X_1, \ldots, X_{50}$, gdzie dla $i = 1, \ldots, 50$ zmienna losowa X_i oznacza czas obsługi i-tego petenta.

Musimy teraz obliczyć parametry zmiennej losowej Y. Zgodnie z treścią zadania każda ze zmiennych losowych X_i ma wartość oczekiwaną równą 45 i wariancję równą 15. Zatem z liniowości wartości oczekiwanej otrzymujemy:

$$\mathbb{E}Y = \mathbb{E}X_1 + \ldots + \mathbb{E}X_{50} = 50 \cdot 45 = 2250.$$

Ponadto zakładamy, że czasy obsługi petentów są niezależne, skąd otrzymujemy:

$$\mathrm{Var}Y = \mathrm{Var}(X_1 + \ldots + X_{50}) = \mathrm{Var}X_1 + \ldots + \mathrm{Var}X_{50} = 50 \cdot 15 = 750.$$

Teraz możemy już oszacować interesujące nas prawdopodobieństwo korzystając z nierówności Czebyszewa-Bienaymé:

$$\begin{aligned}\mathbb{P}\left(35 < \frac{Y}{50} < 55\right) &= \mathbb{P}\left(1750 < Y < 2750\right) = \mathbb{P}\left(1750 - 2250 < Y - \mathbb{E}Y < 2750 - 2250\right)\\ &= \mathbb{P}\left(-500 < Y - \mathbb{E}Y < 500\right) = \mathbb{P}\left(|Y - \mathbb{E}Y| < 500\right) = 1 - \mathbb{P}\left(|Y - \mathbb{E}Y| \geqslant 500\right)\\ &\geqslant 1 - \frac{\mathrm{Var}Y}{500^2} = 1 - \frac{750}{250000} = 1 - \frac{3}{1000} = 0{,}997.\end{aligned}$$

(b) *Oszacuj, ilu petentów w urzędzie wystarczy, by z prawdopodobieństwem co najmniej* 0,9 *średni czas obsługi należał do przedziału* $(44, 46)$.

Podobnie jak w poprzednim podpunkcie niech zmienna losowa Y będzie łącznym czasem obsługi petentów, a zmienne losowe $X_1, \ldots, X_n$ będą czasami obsługi poszczególnych petentów. Wtedy $Y = X_1 + \ldots + X_n$, a dla każdego $i = 1, \ldots, n$ mamy $\mathbb{E}X_i = 45$ oraz $\mathrm{Var}X_i = 15$. W związku z tym parametry zmiennej losowej Y wynoszą:

$$\mathbb{E}Y = \mathbb{E}X_1 + \ldots + \mathbb{E}X_n = n \cdot 45 = 45n$$

oraz

$$\mathrm{Var}Y = \mathrm{Var}(X_1 + \ldots + X_{50}) = \mathrm{Var}X_1 + \ldots + \mathrm{Var}X_n = 15n.$$

Naszym celem jest znalezienie takiej wartości liczby n, dla której prawdopodobieństwo $\mathbb{P}\left(\frac{Y}{n} \in (44, 46)\right)$ będzie równe co najmniej 0,9. W tym celu najpierw oszacujemy to prawdopodobieństwo korzystając z nierówności Czebyszewa-Bienaymé:

$$\begin{aligned}\mathbb{P}\left(\frac{Y}{n} \in (44, 46)\right) &= \mathbb{P}\left(44 < \frac{Y}{n} < 46\right) = \mathbb{P}\left(44n < Y < 46n\right) = \mathbb{P}\left(44n - 45n < Y - \mathbb{E}Y < 46n - 45n\right)\\ &= \mathbb{P}\left(-n < Y - \mathbb{E}Y < n\right) = \mathbb{P}\left(|Y - \mathbb{E}Y| < n\right) = 1 - \mathbb{P}\left(|Y - \mathbb{E}Y| \geqslant n\right) \geqslant 1 - \frac{\mathrm{Var}Y}{n^2}\\ &= 1 - \frac{15n}{n^2} = 1 - \frac{15}{n}.\end{aligned}$$

Zatem wystarczy dobrać n spełniające poniższą nierówność

$$1 - \frac{15}{n} \geqslant 0{,}9.$$

Otrzymujemy $n \geqslant 150$. Zauważmy, że zgodnie z obliczeniami powyżej oznacza to, że jeśli w urzędzie pojawi się 150 (lub więcej) petentów, ich średni czas obsługi będzie mieścił się w przedziale $(44, 46)$ z prawdopodobieństwem co najmniej 0,9. Co więcej, to prawdopodobieństwo będzie rosło wraz z n.

Zadanie 5. *Rozkład zmiennej losowej X zadany jest dystrybuantą*

$$F(x) = \begin{cases} 0 & \text{dla } x < -1 \\ \frac{(x+1)^2}{4} & \text{dla } -1 \leqslant x \leqslant 1 \\ 1 & \text{dla } x > 1 \end{cases}$$

Oszacuj prawdopodobieństwo, że zmienna losowa X odstaje od swojej wartości oczekiwanej o co najmniej

a) $\frac{2}{3}$,
b) $\frac{1}{9}$.

Zacznijmy od obliczenia parametrów zmiennej losowej X. Podana dystrybuanta jest funkcją ciągłą, więc zmienna losowa X też jest ciągła. To oznacza, że możemy wyznaczyć jej gęstość $f(x)$. Pamiętając, że $f(x) = F'(x)$ otrzymujemy:

$$f(x) = \begin{cases} \frac{x+1}{2} & \text{dla } -1 \leqslant x \leqslant 1, \\ 0 & \text{w pozostałych przypadkach.} \end{cases}$$

Zatem wartość oczekiwana tej zmiennej losowej wynosi

$$\mathbb{E}X = \int_{-\infty}^{\infty} tf(t)dt = \int_{-1}^{1} \frac{t^2+t}{2}dt = \left[\frac{t^3}{6} + \frac{t^2}{4}\right]_{-1}^{1} = \frac{5}{12} - \frac{1}{12} = \frac{1}{3}.$$

Aby wyznaczyć wariancję zmiennej losowej X policzymy najpierw $\mathbb{E}X^2$:

$$\mathbb{E}X^2 = \int_{-\infty}^{\infty} t^2 f(t)dt = \int_{-1}^{1} \frac{t^3+t^2}{2}dt = \left[\frac{t^4}{8} + \frac{t^3}{6}\right]_{-1}^{1} = \frac{7}{24} + \frac{1}{24} = \frac{1}{3}.$$

Zatem

$$\mathrm{Var}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{1}{3} - \left(\frac{1}{3}\right)^2 = \frac{2}{9}.$$

Możemy teraz oszacować prawdopodobieństwo, że zmienna losowa X odstaje od swojej wartości oczekiwanej o co najmniej $\frac{2}{3}$, czyli $\mathbb{P}\left(|X - \mathbb{E}X| \geqslant \frac{2}{3}\right)$. W tym celu skorzystamy z nierówności Czebyszewa-Bienaymé:

$$\mathbb{P}\left(|X - \mathbb{E}X| \geqslant \frac{2}{3}\right) \leqslant \frac{\mathrm{Var}X}{\left(\frac{2}{3}\right)^2} = \frac{\frac{2}{9}}{\frac{4}{9}} = \frac{1}{2}.$$

W drugim przypadku szacujemy prawdopodobieństwo, że zmienna losowa X odstaje od swojej wartości oczekiwanej o co najmniej $\frac{1}{9}$, czyli $\mathbb{P}\left(|X - \mathbb{E}X| \geqslant \frac{1}{9}\right)$. Korzystając ponownie z nierówności Czebyszewa-Bienaymé dostajemy:

$$\mathbb{P}\left(|X - \mathbb{E}X| \geqslant \frac{1}{9}\right) \leqslant \frac{\mathrm{Var}X}{\left(\frac{1}{9}\right)^2} = \frac{\frac{2}{9}}{\frac{1}{81}} = 18.$$

Proszę zwrócić uwagę, że w tym przypadku otrzymaliśmy oszacowanie gorsze od oszacowania trywialnego

$$\mathbb{P}\left(|X - \mathbb{E}X| \geqslant \frac{1}{9}\right) \leqslant 1.$$

Zadanie 6. $X_1, X_2, \ldots, X_{20}$ *są zmiennymi losowymi (niekoniecznie niezależnymi) o rozkładzie Poissona* $Po(1)$*.*

(a) *Użyj nierówności Markowa do oszacowania z góry* $\mathbb{P}(X \geqslant 30)$*, dla* $X = \sum_i^{20} X_i$*.*

Musimy najpierw obliczyć $\mathbb{E}X$. Każda ze zmiennych losowych $X_1, \ldots, X_{20}$ ma rozkład Poissona z parametrem $\lambda = 1$, więc ich wartości oczekiwane również wynoszą 1. Zatem z liniowości wartości oczekiwanej otrzymujemy:

$$\mathbb{E}X = \mathbb{E}X_1 + \ldots + \mathbb{E}X_{20} = 20 \cdot 1 = 20.$$

Zauważmy jeszcze, że zmienna losowa X jest sumą zmiennych losowych o rozkładach Poissona, a więc przyjmujących wartości ujemne z zerowym prawdopodobieństwem. Zatem zmienna losowa X także przyjmuje wartości ujemne z zerowym prawdopodobieństwem. Możemy teraz oszacować szukane prawdopodobieństwo korzystając z nierówności Markowa:

$$\mathbb{P}(X \geqslant 30) = \mathbb{P}(|X| \geqslant 30) \leqslant \frac{\mathbb{E}|X|}{30} = \frac{20}{30} = \frac{2}{3}.$$

(b) *Załóżmy teraz, że* $X_1, X_2, \ldots, X_{20}$ *są niezależne. Czy można uzyskać lepsze oszacowanie prawdopodobieństwa* $\mathbb{P}(X \geqslant 30)$ *korzystając z nierówności Czebyszewa–Bienaymé?*

Wariancja każdej ze zmiennych losowych $X_1, \ldots, X_{20}$ wynosi $\lambda = 1$. Zatem jeśli założymy, że zmienne losowe $X_1, \ldots, X_{20}$ są niezależne, to:

$$\mathrm{Var}X = \mathrm{Var}X_1 + \ldots + \mathrm{Var}X_{20} = 20 \cdot 1 = 20.$$

Możemy teraz oszacować szukane prawdopodobieństwo z nierówności Czebyszewa-Bienaymé:

$$\mathbb{P}(X \geqslant 30) = \mathbb{P}(X - \mathbb{E}X \geqslant 30 - 20) \leqslant \mathbb{P}(|X - \mathbb{E}X| \geqslant 10) \leqslant \frac{\mathrm{Var}X}{10^2} = \frac{20}{100} = \frac{1}{5}.$$

Po raz kolejny oszacowanie z nierówności Czebyszewa-Bienaymé okazuje się być lepsze niż oszacowanie z nierówności Markowa.