Wstęp do rachunku prawdopodobieństwa
Wykład 4: Zmienne losowe i ich rozkłady.
Funkcja masy prawdopodobieństwa. Dystrybuanta. Wartość oczekiwana.

Zmienne losowe

Przykład 1. Rozważmy eksperyment polegający na trzykrotnym rzucie monetą, zatem jako przestrzeń zdarzeń elementarnych przyjmijmy

$$\Omega = \{OOO, OOR, \ldots, RRR\},$$

gdzie wszystkie zdarzenia elementarne są równoprawdopodobne. Interesuje nas ile razy w tych trzech rzutach wypadnie orzeł. Możemy zatem zdefiniować funkcję X, która dla każdego zdarzenia elementarnego będzie zwracała liczbę orłów wyrzuconych w trzech rzutach monetą, tzn. X jest funkcją określoną na zbiorze Ω, która przyjmuje wartości w zbiorze $\{0, 1, 2, 3\}$ i zadana jest w następujący sposób:

$$X(\{OOO\}) = 3, \quad X(\{OOR\}) = 2, \quad X(\{ORO\}) = 2, \quad X(\{ROO\}) = 2,$$
$$X(\{RRR\}) = 0, \quad X(\{RRO\}) = 1, \quad X(\{ROR\}) = 1, \quad X(\{ORR\}) = 1.$$

Taką funkcję X nazywać będziemy zmienną losową.

Definicja 1. Zmienną losową *nazywamy funkcję (mierzalną)* $X : \Omega \to \mathbb{R}$ *działającą z przestrzeni probabilistycznej* $(\Omega, \mathcal{F}, \mathbb{P})$ *do zbioru liczb rzeczywistych* $\mathbb{R}$.

Dla $A \subseteq \mathbb{R}$ przyjmujemy oznaczenie $\mathbb{P}(X \in A) = \mathbb{P}\left(X^{-1}(A)\right) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$.

Definicja 2. *Zmienna losowa jest* **dyskretna**, *jeśli istnieje przeliczalny zbiór* $A = \{a_1, a_2, \ldots\}$ *(nazywany zbiorem atomów) taki, że* $\mathbb{P}(X = a_i) > 0$ *dla każdego* $a_i \in A$ *oraz*

$$\mathbb{P}(X \in A) = \sum_{a_i \in A} \mathbb{P}(X = a_i) = 1.$$

Przykład 2. Dla zmiennej losowej X z Przykładu 1. mamy:

$$\mathbb{P}(X \leqslant 1) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leqslant 1\}) = \mathbb{P}(\{RRR, RRO, ROR, ORR\}) = \frac{1}{2},$$
$$\mathbb{P}(X > 2) = \mathbb{P}(X = 3) = \mathbb{P}(\{OOO\}) = \frac{1}{8}.$$

Ponadto funkcja X jest przykładem zmiennej losowej dyskretnej, ponieważ z niezerowym prawdopodobieństwem przyjmuje jedynie wartości ze zbioru $\{0, 1, 2, 3\}$, a zatem te wartości stanowią jej zbiór atomów.

Rozkład prawdopodobieństwa zmiennej losowej

Definicja 3. Rozkładem prawdopodobieństwa *zmiennej losowej* X *nazywamy funkcję* μ_X, *która wszystkim zbiorom (borelowskim)* $B \subseteq \mathbb{R}$ *przypisuje wartość prawdopodobieństwa*

$$\mu_X(B) = \mathbb{P}\left(X^{-1}(B)\right).$$

W celu podania rozkładu prawdopodobieństwa zmiennej losowej dyskretnej, wystarczy podać wartości

$$p_i = \mathbb{P}(X = a_i)$$

dla wszystkich atomów $a_i \in A$. *Jest to tzw.* **funkcja masy prawdopodobieństwa** *tej zmiennej losowej. Czasem nazywamy ją po prostu* **rozkładem** *zmiennej losowej dyskretnej.*

Przykład 3. Rozkład zmiennej losowej X z Przykładu 1. zadany jest funkcją masy prawdopodobieństwa:

$$p_0 = \mathbb{P}(X = 0) = \frac{1}{8}, \quad p_1 = \mathbb{P}(X = 1) = \frac{3}{8}, \quad p_2 = \mathbb{P}(X = 2) = \frac{3}{8}, \quad p_3 = \mathbb{P}(X = 3) = \frac{1}{8}.$$

Innym wygodnym sposobem podania rozkładu zmiennej losowej jest tabelka:

x	0	1	2	3
$\mathbb{P}(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Przykład 4. Podane na Wykładzie 2. (Przykład 5. i 6.) schemat Bernoulliego (zwany też **rozkładem Bernoulliego** lub **rozkładem dwumianowym**, oznaczanym przez $Bin(n, p)$) i **rozkład geometryczny** (oznaczany przez $Ge(p)$) stanowią przykłady ważnych rozkładów prawdopodobieństwa. Dla przypomnienia, rozkład Bernoulliego dotyczy liczby sukcesów w n niezależnych próbach (tzw. próbach Bernoulliego) z prawdopodobieństwem sukcesu w pojedyńczej próbie równym p. Jeśli przez X oznaczymy zmienną losową równą liczbie sukcesów w n próbach Bernoulliego, to rozkład zmiennej losowej X dany jest wzorem:

$$\tau_k = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Sprawdźmy, że jest to rzeczywiście rozkład prawdopodobieństwa. W tym celu musimy pokazać, że prawdopodobieństwa poszczególnych atomów sumują się do 1. Do tego celu wykorzystamy poniższą tożsamość, która zachodzi dla dowolnych $a, b \in \mathbb{R}$ oraz $n \in \mathbb{N}$:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Mamy zatem:

$$\sum_{k=0}^{n} \mathbb{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Wnioskujemy stąd, że jest to rzeczywiście dobrze określony rozkład prawdopodobieństwa.

Rozkład geometryczny z kolei dotyczy wystąpienia pierwszego sukcesu w niezależnych próbach Bernoulliego przy prawdopodobieństwie sukcesu w pojedyńczej próbie równym p. Jeśli przez Y oznaczymy zmienną losową równą liczbie prób potrzebnych do osiągnięcia pierwszego sukcesu, to rozkład zmiennej losowej Y dany jest wzorem:

$$\sigma_k = \mathbb{P}(Y = k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

Ponownie sprawdźmy, czy jest to dobrze zadany rozkład prawdopodobieństwa. W tym celu posłużymy się wzorem na **sumę szeregu geometrycznego**:

$$\sum_{k=0}^{\infty} \alpha \rho^k = \frac{\alpha}{1-\rho}, \quad \text{przy założeniu } |\rho| < 1.$$

Zatem otrzymujemy

$$\sum_{k=1}^{\infty} \mathbb{P}(Y = k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{k=0}^{\infty} (1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1.$$

Spostrzegawczy studenci powinni zauważyć, że zmienna losowa X z Przykładu 1. ma właśnie rozkład dwumianowy z parametrami $n = 3$ i $p = 1/2$, co zapisujemy również jako $X \sim Bin(3, 1/2)$.

Dystrybuanta zmiennej losowej

Definicja 4. Dystrybuantą *zmiennej losowej X nazywamy funkcję $F_X : \mathbb{R} \to \mathbb{R}$ daną wzorem:*

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}\left(X^{-1}((-\infty, x])\right).$$

W szczególności dla zmiennej losowej dyskretnej o zbiorze atomów $A = \{a_1, a_2, \ldots\}$ mamy:

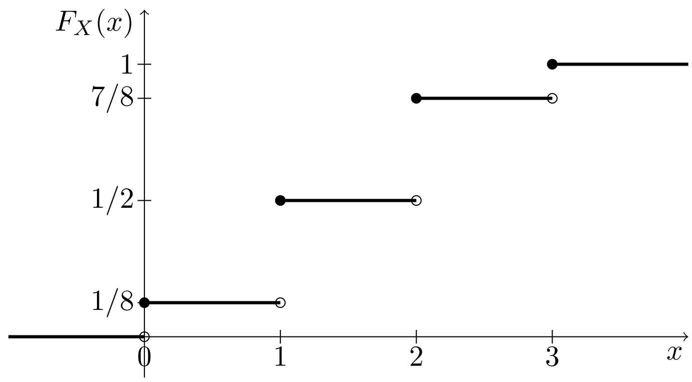
$$F_X(x) = \sum_{a_i \in A, a_i \leqslant x} \mathbb{P}(X = a_i).$$

Zazwyczaj dystrybuantę będziemy podawać albo za pomocą wzoru funkcji F_X, albo rysując jej wykres.

Przykład 5. Dla zmiennej losowej X z Przykładu 1. wzór dystrybunaty przedstawia się następująco:

$$F_X(x) = \begin{cases} 0 \text{ dla } x < 0; \\ \frac{1}{8} \text{ dla } 0 \leqslant x < 1; \\ \frac{1}{2} \text{ dla } 1 \leqslant x < 2; \\ \frac{7}{8} \text{ dla } 2 \leqslant x < 3; \\ 1 \text{ dla } x \geqslant 3. \end{cases}$$

Z kolei wykres tej dystrybuanty wygląda jak poniżej:



Proszę zwrócić uwagę, że w przypadku zmiennej dyskretnej wykresem funkcji jest tzw. funkcja schodkowa, gdzie „skoki” odpowiadają dokładnie atomom tej zmiennej, tzn. “skoki” występują dla tych wartości x, które są atomami, natomiast „wysokość” skoku równa jest prawodpodobieństwu tego atomu, czyli ma wartość $\mathbb{P}\left(X = x\right)$.

Wartość oczekiwana zmiennej losowej

Jednym z ważniejszych parametrów dotyczących zmiennych losowych jest tzw. wartość oczekiwana zmiennej losowej. Intuicyjnie rzecz ujmując, jest to średnia wartość przyjmowana przez zmienną losową, czyli taka, której należy się spodziewać.

Definicja 5. **Wartością oczekiwaną** *(lub* **wartością średnią***) dyskretnej zmiennej losowej X o zbiorze atomów A nazywamy liczbę*

$$\mathbb{E}X = \sum_{a_i \in A} a_i \mathbb{P}\left(X = a_i\right),$$

o ile suma szeregu po prawej stronie powyższego równania istnieje. Jeśli szereg po prawej stronie powyższego równania nie jest bezwględnie zbieżny, mówimy, że wartość oczekiwana zmiennej losowej X nie istnieje.

Przykład 6. Zmienna losowa X z Przykładu 1. ma wartość oczekiwaną równą:

$$\mathbb{E}X = 0 \cdot \mathbb{P}\left(X = 0\right) + 1 \cdot \mathbb{P}\left(X = 1\right) + 2 \cdot \mathbb{P}\left(X = 2\right) + 3 \cdot \mathbb{P}\left(X = 3\right) = 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2}.$$

Jak już zauważyliśmy, zmienna losowa X ma rozkład dwumianowy $Bin(3, 1/2)$. Okazuje się, że nieprzypadkowo $\mathbb{E}X = 3 \cdot 1/2$. Pokażemy, że wartość oczekiwana zmiennej losowej Y o rozkładzie dwumianowym $Bin(n, p)$ wynosi

$$\mathbb{E}Y = np.$$

Przypomnijmy, że zbiorem atomów zmiennej losowej Y jest zbiór $\{0, 1, \ldots, n\}$ oraz

$$\mathbb{P}(Y = k) = \tau_k = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Mamy zatem:

$$\begin{aligned}
\mathbb{E}Y &= \sum_{k=0}^{n} k\cdot\mathbb{P}(Y=k) = \sum_{k=1}^{n} k\cdot\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=1}^{n} k\cdot\frac{n!}{k!(n-k!)}p^k(1-p)^{n-k}\\
&= np\cdot\sum_{k=1}^{n}\frac{(n-1)!}{(k-1)!(n-k)!}p^{k-1}(1-p)^{n-k} = np\cdot\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}\\
&= np\cdot\sum_{k=0}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k} = np\cdot(p+(1-p))^{n-1} = np\,.
\end{aligned}$$

Przykład 7 (trudny!). Rozważamy zmienną losową X o rozkładzie geometrycznym z prawdopodobieństwem sukcesu równym p, czyli $X \sim Ge(p)$. Ile wynosi wartość oczekiwana zmiennej losowej X?

Zbiorem atomów zmiennej losowej X jest zbiór $\{1, 2, \ldots\}$, przy czym

$$\mathbb{P}\left(X=k\right) = p(1-p)^{k-1}.$$

Stąd mamy:

$$\mathbb{E}X = \sum_{k=1}^{\infty} k\cdot\mathbb{P}\left(X=k\right) = \sum_{k=1}^{\infty} k\cdot p(1-p)^k = p\sum_{k=1}^{\infty} k\cdot(1-p)^{k-1}.$$

Jeżeli potraktujemy p jako zmienną formalną, to możemy **scałkować**, a potem **zróżniczkować** po p szereg po prawej otrzymując:

$$\mathbb{E}X = p\sum_{k=1}^{\infty}\left(\int k\cdot(1-p)^{k-1}dp\right)' = p\sum_{k=1}^{\infty}\left(-(1-p)^k\right)' = -p\left(\sum_{k=1}^{\infty}(1-p)^k\right)'.$$

Następnie, korzystając ze wzoru na sumę szeregu geometrycznego, mamy:

$$\mathbb{E}X = -p\left(\frac{1-p}{1-(1-p)}\right)' = -p\left(\frac{1-p}{p}\right)' = -p\left(\frac{1}{p}-1\right)' = -p\cdot\left(-\frac{1}{p^2}\right) = \frac{1}{p}.$$

A zatem wartość oczekiwana zmiennej losowej X o rozkładzie $Ge(p)$ wynosi $1/p$.

Przykład 8. Rozważmy eksperyment polegający na sześciokrotnym rzucie standardową kostką. Niech X oznacza liczbę szóstek, które wypadły w trakcie tego eksperymentu, a Y numer rzutu, w którym szóstka wypadła po raz pierwszy, przy czym jeśli szóstka nie wypadła ani razu przyjmujemy, że $Y = 0$. Ile wynosci $\mathbb{E}X$ i $\mathbb{E}Y$?

Zmienna losowa X ma rozkład **dwumianowy** $X \sim Bin(6, \frac{1}{6})$. Zatem

$$\mathbb{E}X = 6\cdot\frac{1}{6} = 1.$$

Rozkład zmiennej losowe Y nie jest żadnym z poznanych do tej pory rozkładów, więc musimy go wyznaczyć ręcznie.

y	0	1	2	3	4	5	6
$\mathbb{P}\left(Y=y\right)$	$\left(\frac{5}{6}\right)^6$	$\frac{1}{6}$	$\frac{1}{6}\cdot\frac{5}{6}$	$\frac{1}{6}\left(\frac{5}{6}\right)^2$	$\frac{1}{6}\left(\frac{5}{6}\right)^3$	$\frac{1}{6}\left(\frac{5}{6}\right)^4$	$\frac{1}{6}\left(\frac{5}{6}\right)^5$

Teraz liczymy wartość oczekiwaną zmiennej losowej Y odwołując się do definicji

$$\mathbb{E}Y = \frac{1}{6}\left(1+2\cdot\frac{5}{6}+3\cdot\left(\frac{5}{6}\right)^2+4\cdot\left(\frac{5}{6}\right)^3+5\cdot\left(\frac{5}{6}\right)^4+6\cdot\left(\frac{5}{6}\right)^5\right).$$

Gdyby dopuścić dowolną liczbę rzutów kostką i przez Z oznaczyć liczbę rzutów potrzebnych do wyrzucenia pierwszej szóstki, wtedy zmienna losowa Z miałaby rozkład **geometryczny** $Z \sim Ge(\frac{1}{6})$, a jej wartość oczekiwana wyniosłaby

$$\mathbb{E}Z = \frac{1}{\frac{1}{6}} = 6.$$