2. ZMIENNE LOSOWE

Definicja. Niech $(\Omega, \mathcal{F}, \mathcal{P})$ będzie przestrzenią probabilistyczną. *Zmienną losową* nazywamy dowolną funkcję $X : \Omega \to \mathbb{R}$ spełniającą warunek

$$\{\omega : X(\omega) \leqslant a\} \in \mathcal{F} \text{ dla każdego } a \in \mathbb{R}. \tag{1}$$

O funkcji X spełniającej (1) mówimy, że jest mierzalna. Zatem, innymi słowy, zmienna losowa to dowolna mierzalna funkcja $X : \Omega \to \mathbb{R}$.

Przykład 2.1. Gra polega na rzucie symetryczną kostką do gry. Jeśli wypadnie 6 oczek, to wygrywamy 10 zł, jeśli 4 lub 5 oczek, to wygrywamy 2 zł, w pozostałych przypadkach tracimy 4 zł. Niech X oznacza wysokoć naszej wygranej.
Mamy $\Omega = \{1, 2, 3, 4, 5, 6\}$ i

$$X(1) = X(2) = X(3) = -4,\ X(4) = X(5) = 2,\ X(6) = 10.$$

Skoro $\mathcal{F} = 2^\Omega$, to warunek (1) jest na pewno spełniony i X jest zmienną losową. Prawdopodobieństwa otrzymania kolejnych, możliwych wartości dla zmiennej losowej X opiszemy tabelką:

x_k	-4	2	10
p_k	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Tabelka ta podaje rozkład X. Zauważmy, że musimy mieć $p_k > 0$ dla wszystkich k i $\sum_k p_k = 1$.

Przykład 2.2. Wykładowca zawsze spóźnia się na wykład i rozpoczyna go w losowym momencie pomiędzy 10:15 a 10:30. Niech X oznacza opóźnienie rozpoczęcia wykładu liczone w minutach. Wtedy X przyjmuje nieprzeliczalną liczbę wartości (wszystkie liczby z przedziału $(0, 15)$) i jest zmienną losową. Rzeczywiście, możemy przyjąć $\Omega = (15, 30)$, $\mathcal{F} = \mathcal{B}(15, 30)$,

$$\{\omega : X(\omega) \leqslant a\} = \begin{cases} \emptyset & \text{dla} \quad a \leqslant 0, \\ (15, 15 + a] & \text{dla} \quad a \in (0, 15), \\ \Omega & \text{dla} \quad a \geqslant 15 \end{cases}$$

i widzimy, że warunek $\{\omega : X(\omega) \leqslant a\} \in \mathcal{F}$ jest spełniony dla każdego $a \in \mathbb{R}$, bo, bezpośrednio z definicji σ-ciała

$$\Omega \in \mathcal{F} \ \Rightarrow\ \emptyset = \Omega \setminus \Omega \in \mathcal{F},$$

a z definicji borelowskiego σ-ciała

$$\begin{aligned} &(15 + a, 30) \in \mathcal{B}(15, 30) \ \Rightarrow \\ &(15, 15 + a] = (15, 30) \setminus (15 + a, 30) \in \mathcal{B}(15, 30). \end{aligned}$$

Wyróżniamy trzy rodzaje zmiennych losowych.

1. **Zmienne losowe dyskretne** (inaczej zwane **skokowymi**).
   Przyjmują one skończoną bądź przeliczalną liczbę wartości. Rozkład tych o skończonej liczbie wartości wygodnie jest opisywać za pomocą tabelki, tak jak to zrobiliśmy w przykładzie 2.1.

2. **Zmienne losowe absolutnie ciągłe**.
   Przyjmują one nieprzeliczalną liczbę wartości i ich rozkład daje się opisać przez *funkcję gęstości* $f(x)$, czyli nieujemną funkcję, taką że

   $$P(X \in A) = \int_A f(x)dx \qquad \text{dla dowolnego } A \in \mathcal{B}(\mathbb{R}).$$

   Każda funkcja $f : \mathbb{R} \to \mathbb{R}$ spełniająca warunki:

   (a) $f(x) \geqslant 0$ dla wszystkich $x \in \mathbb{R}$,

   (b) $\int_{\mathbb{R}} f(x)dx = 1$

   jest gęstością pewnej absolutnie ciągłej zmiennej losowej.

   Zauważmy, że dla zmiennych losowych absolutnie ciągłych mamy

   $$P(X = a) = \int_{\{a\}} f(x)dx = 0,$$

   pomimo, że zdarzenie $\{X = a\}$ nie musi być niemożliwe.

   W przykładzie 2.2 mamy zmienną losową absolutnie ciągłą o funkcji gęstości

   $$f(x) = \begin{cases} \frac{1}{15} & \text{dla} \quad x \in (0, 15) \\ 0 & \text{dla} \quad x \notin (0, 15). \end{cases}$$

3. **Zmienne losowe singularne**.

Przykłady rozkładów (i zmiennych losowych) dyskretnych

1). <u>Rozkład dwupunktowy</u>: Wykonujemy doświadczenie, które może zakończyć się jedynie sukcesem (z prawdopodobieństwem $p \in (0, 1)$) albo porażką. Zmienna losowa X przyjmuje wartość 1, gdy zajdzie sukces, oraz wartość 0, gdy zajdzie porażka. Wtedy rozkład X opisuje poniższa tabelka:

x_k	0	1
p_k	$1-p$	p

,

co równoważnie można zapisać następująco:

$$P(X = 0) = 1 - p \ \text{ i } \ P(X = 1) = p.$$

2). Rozkład dwumianowy: $X \sim binom(n, p)$.

Powtarzamy n-krotnie doświadczenie, które może zakończyć się jedynie sukcesem (z prawdopodobieństwem $p \in (0, 1)$) albo porażką. Zakładamy, że kolejne powtórzenia doświadczenia przebiegają niezależnie od pozostałych. Wtedy X to liczba uzyskanych sukcesów.

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

3). Rozkład Poissona: $X \sim Pois(\lambda)$, gdzie $\lambda > 0$, jeśli

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{dla } k = 0, 1, \ldots.$$

4). Rozkład geometryczny: $X \sim Ge(p)$, gdzie $p \in (0, 1)$.

Powtarzamy doświadczenie, które może zakończyć się sukcesem (z prawdopodobieństwem p) albo porażką. Zakładamy, że kolejne powtórzenia doświadczenia przebiegają niezależnie od pozostałych. Wtedy X to liczba porażek poprzedzających pierwszy sukces.

$$P(X = k) = p(1-p)^k \text{ dla } k = 0, 1, \ldots.$$

5). Rozkład ujemny dwumianowy: $X \sim nb(p, r)$, gdzie $p \in (0, 1)$, $r \in \{1, 2, \ldots\}$.

Powtarzamy doświadczenie, które może zakończyć się sukcesem (z prawdopodobieństwem p) albo porażką. Zakładamy, że kolejne powtórzenia doświadczenia przebiegają niezależnie od pozostałych. Wtedy X to liczba porażek poprzedzających r-ty sukces.

$$P(X = k) = \binom{r+k-1}{k} p^r (1-p)^k \text{ dla } k = 0, 1, \ldots.$$

Przykłady rozkładów (i zmiennych losowych) absolutnie ciągłych

1). Rozkład jednostajny na przedziale (a, b), oznaczany $U(a, b)$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{dla} \quad x \in (a, b) \\ 0 & \text{dla} \quad x \notin (a, b). \end{cases}$$

2). Rozkład wykładniczy z parametrem $\lambda > 0$, oznaczany $Exp(\lambda)$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla} \quad x > 0 \\ 0 & \text{dla} \quad x \leqslant 0. \end{cases}$$

3). Rozkład normalny o parametrach μ i σ^2, oznaczany $\mathcal{N}(\mu, \sigma^2)$, czyli rozkład o gęstości

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \quad x \in \mathbb{R}, \text{ gdzie } \mu \in \mathbb{R}, \ \sigma > 0.$$

Wykres tej gęstości to tzw. *krzywa gaussa.*

Własności rozkładu normalnego:

- Rozkład normalny $\mathcal{N}(\mu, \sigma^2)$ jest rozkładem symetrycznym względem μ.
- Jeśli $X \sim \mathcal{N}(\mu, \sigma^2)$, to $U = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$. Rozkład $\mathcal{N}(0,1)$ nazywamy *standardowym rozkładem normalnym.*
- Ogólniej jeśli $X \sim \mathcal{N}(\mu, \sigma^2)$ i a, b są liczbami rzeczywistymi, to $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

4). Rozkład gamma z parametrem kształtu a i drugim parametrem β, oznaczany $Gamma(a, \beta)$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \frac{\beta^a}{\Gamma(a)} x^{a-1} \exp(-\beta x) & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}, \ a > 0, \beta > 0,$$

gdzie $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$.

5). Rozkład chi-kwadrat o n stopniach swobody, który będziemy oznaczać symbolem $\chi^2_{[n]}$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} \exp(-x/2) & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}.$$

Rozkład $\chi^2_{[n]}$ to to samo co rozkład $Gamma(a = \frac{n}{2}, \beta = \frac{1}{2})$.

6). Rozkład t-Studenta o n stopniach swobody, który będziemy oznaczać symbolem $t_{[n]}$, czyli rozkład o gęstości

$$f(x) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \frac{1}{\sqrt{n\pi}} \frac{1}{(1 + x^2/n)^{(n+1)/2}} \quad \text{dla } x \in \mathbb{R}.$$

7). Rozkład F-Snedecora o stopniach swobody m i n, który będziemy oznaczać symbolem $F_{[m,n]}$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} \frac{x^{(m-2)/2}}{\left(1+\frac{m}{n}x\right)^{(m+n)/2}} & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}.$$

Wielkości służące do opisu zmiennych losowych:

1). Dystrybuanta : $F(x) = P(X \leqslant x)$.

Jeśli X jest zmienną losową dyskretną, to jej dystrybuanta jest funkcją schodkową, prawostronnie ciągłą.

Jeśli natomiast X jest zmienną losową absolutnie ciągłą, to jej dystrybuantę wyznaczamy korzystając ze wzoru

$$F(t) = P(X \leqslant t) = \int_{-\infty}^{t} f(x)dx$$

i otrzymujemy funkcję ciągłą.

Twierdzenie 2. 1. Dystrybuanta ma następujące własności:

1. jest funkcją niemalejącą,
2. jest co najmniej prawostronnie ciągła,
3. $\lim_{x\to-\infty} F(x) = 0$ i $\lim_{x\to\infty} F(x) = 1$.

Prawdziwe jest także twierdzenie odwrotne do powyższego.

Twierdzenie 2. 2. Każda funkcja $F : \mathbb{R} \to \mathbb{R}$, mająca własności 1.-3. z poprzedniego twierdzenia, jest dystrybuantą pewnej zmiennej losowej.

2). Ogon dystrybuanty (funkcja przeżycia) :

$$\overline{F}(x) = 1 - F(x) = P(X > x).$$

3). Kwantyl rzędu $\alpha \in (0, 1)$:

$$q_\alpha \text{ to najmniejsza liczba spełniająca } F(q_\alpha) \geqslant \alpha.$$

W szczególności kwantyl rzędu $\alpha = \frac{1}{2}$ to **mediana** med, kwantyl rzędu $\alpha = \frac{1}{4}$ to **dolny (pierwszy) kwartyl** Q_1, zaś kwantyl rzędu $\alpha = \frac{3}{4}$ to **górny (trzeci) kwartyl** Q_3.

4). Wartość oczekiwana (wartość średnia, wartość przeciętna) :

$$EX = \begin{cases} \sum_k x_k p_k & \text{dla zmiennych losowych dyskretnych,} \\ \int_{\mathbb{R}} x f(x)dx & \text{dla zmiennych losowych absolutnie ciągłych,} \end{cases}$$

przy czym mówimy, że EX nie istnieje, gdy $\sum\limits_{k:x_k\leqslant 0} x_k p_k = -\infty$ i $\sum\limits_{k:x_k\geqslant 0} x_k p_k = \infty$ (w przypadku zmiennych losowych dyskretnych) lub $\int_{-\infty}^{0} x f(x)dx = -\infty$ i $\int_0^\infty x f(x)dx = \infty$ (w przypadku zmiennych losowych absolutnie ciągłych).

Własności wartości oczekiwanej: jeśli $a \in \mathbb{R}$, to

- $E(a) = a$,
- $E(aX) = aE(X)$, jeśli tylko EX istnieje (przy czym przyjmujemy tu, że $0 \cdot \pm\infty = 0$),
- $E(X \pm Y) = EX \pm EY$, jeśli tylko EX i EY istnieją i wyrażenie $EX \pm EY$ ma sens (tzn. nie prowadzi do $\infty - \infty$ lub $-\infty + \infty$),
- jeśli $X \leqslant Y$ i EX oraz EY istnieją, to $EX \leqslant EY$.

5). Wariancja : $Var(X) = E\left((X - EX)^2\right) = E(X^2) - (EX)^2$, gdzie

$$E(X^2) = \begin{cases} \sum_k x_k^2 p_k & \text{dla zmiennych losowych dyskretnych,} \\ \int_{\mathbb{R}} x^2 f(x)dx & \text{dla zmiennych losowych absolutnie ciągłych} \end{cases}$$

i ogólniej, jeśli g jest funkcją taką, że $g(X)$ jest zmienną losową (w szczególności, jeśli g jest funkcją ciągłą), to

$$E(g(X)) = \begin{cases} \sum_k g(x_k)p_k & \text{dla zmiennych losowych dyskretnych,} \\ \int_{\mathbb{R}} g(x)f(x)dx & \text{dla zmiennych losowych absolutnie ciągłych,} \end{cases}$$

przy czym mówimy, że $E(g(X))$ nie istnieje, gdy $\sum\limits_{k:g(x_k)\leqslant 0} g(x_k)p_k = -\infty$ i $\sum\limits_{k:g(x_k)\geqslant 0} g(x_k)p_k = \infty$ (w przypadku zmiennych losowych dyskretnych) lub $\int_{\mathbb{R}} g^-(x)f(x)dx = \int_{\mathbb{R}} g^+(x)f(x)dx = \infty$, gdzie

$$g^-(x) = \begin{cases} 0 & \text{gdy} \quad g(x) > 0 \\ -g(x) & \text{gdy} \quad g(x) \leqslant 0 \end{cases} \quad \text{i } g^+(x) = \begin{cases} 0 & \text{gdy} \quad g(x) < 0 \\ g(x) & \text{gdy} \quad g(x) \geqslant 0 \end{cases}$$

(w przypadku zmiennych losowych absolutnie ciągłych).

Własności wariancji:

- jeśli $a, b \in \mathbb{R}$ i $Var(X)$ istnieje, to $Var(aX + b) = a^2 Var(X)$; w szczególności $Var(b) = 0$;
- dla dowolnej zmiennej losowej X, takiej że $Var(X)$ istnieje, mamy $Var(X) \geqslant 0$.

6). Odchylenie standardowe : $\sigma_X = \sqrt{Var(X)}$.

Dla przykładowych rozkładów zmiennych losowych mamy:

- jeśli $X \sim binom(1, p)$, to $EX = p$ i $Var(X) = p(1-p)$;
- jeśli $X \sim binom(n, p)$, to $EX = np$ i $Var(X) = np(1-p)$;
- jeśli $X \sim Pois(\lambda)$, to $EX = Var(X) = \lambda$;
- jeśli $X \sim Ge(p)$, to $EX = \frac{1-p}{p}$ i $Var(X) = \frac{1-p}{p^2}$;
- jeśli $X \sim nb(p, r)$, to $EX = \frac{(1-p)r}{p}$ i $Var(X) = \frac{(1-p)r}{p^2}$;
- jeśli $X \sim U(a, b)$, to $EX = \frac{a+b}{2}$ i $Var(X) = \frac{(b-a)^2}{12}$;
- jeśli $X \sim Exp(\lambda)$, to $EX = \frac{1}{\lambda}$ i $Var(X) = \frac{1}{\lambda^2}$;
- jeśli $X \sim \mathcal{N}(\mu, \sigma^2)$, to $EX = \mu$ i $Var(X) = \sigma^2$;
- jeśli $X \sim Gamma(a, \beta)$, to $EX = \frac{a}{\beta}$ i $Var(X) = \frac{a}{\beta^2}$;
- jeśli $X \sim t_{[n]}$, to $EX = 0$ dla $n > 1$ i $Var(X) = \frac{n}{n-2}$ dla $n > 2$;
- jeśli $X \sim F_{[m,n]}$, to $EX = \frac{n}{n-2}$ dla $n > 2$ i $Var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ dla $n > 4$.

FUNKCJE W PAKIECIE R:

rozkład	$P(X=x)$ dla rozkładów dyskretnych lub $f(x)$ dla rozkładów ciągłych
$binom(n,p)$	dbinom(x, size=n, prob=p)
$Pois(\lambda)$	dpois(x, lambda= λ)
$Ge(p)$	dgeom(x, prob=p)
$nb(p,r)$	dnbinom(x, size=r, prob=p)
$U(a,b)$	dunif(x, min=a, max=b)
$Exp(\lambda)$	dexp(x, rate= λ)
$\mathcal{N}(\mu,\sigma^2)$	dnorm(x, mean= μ, sd= σ)
$Gamma(a,\beta)$	dgamma(x, shape=a, rate= β)
$\chi^2_{[n]}$	dchisq(x, n=n)
$t_{[n]}$	dt(x, n=n)
$F_{[m,n]}$	df(x, n=n, m=m)

Ponadto dla rozkładu dwumianowego:

- dystrybuanta: $P(X \leqslant x)$ =pbinom(q=x, size=n, prob=p, lower.tail=TRUE);
- ogon dystrybuanty: $P(X > x)$ =pbinom(q=x, size=n, prob=p, lower.tail=FALSE);
- kwantyl rzędu $\alpha \in (0,1)$: q_α, czyli najmniejsza liczba spełniająca $P(X \leqslant q_\alpha) \geqslant \alpha$:

  $$q_\alpha = \text{qbinom}(p=\alpha,\ \text{size=n, prob=p, lower.tail=TRUE});$$

- qbinom($p=\alpha$, size=n, prob=p, lower.tail=FALSE) wyznaczy najmniejszą liczbę c_α spełniającą $P(X > c_\alpha) \leqslant \alpha$;
- rbinom(n, size, prob) wygeneruje n liczb losowych z rozkładu *binom*(size,prob)

i analogicznie dla pozostałych rozkładów (fragmenty napisane na niebiesko można pominąć).

Przykład 2.3. Pojemność wyprodukowanej partii kondensatorów jest zmienną losową o rozkładzie normalnym z wartością przeciętną 500pF i odchyleniem standardowym 25pF.
a). Oblicz prawdopodobieństwo tego, że losowo wybrany kondensator z tej partii ma pojemność większą niż 450pF i jednocześnie mniejszą niż 800pF.
b). Wyznacz wartość pojemności, której nie przekracza 75% kondensatorów z tej partii.