

Zmienna losowa, jej dystrybucja i gęstość

Krótkie praktyczne przypomnienie pojęć z rachunku prawdopodobieństwa

Ω – przestrzeń zdarzeń elementarnych

\mathcal{F} – σ -ciało zdarzeń (ustalona rodzina podzbiorów zbioru Ω będąca σ -ciałem)

$\mathcal{B}(\mathbb{R})$ – σ -ciało borelowskich podzbiorów \mathbb{R} – wszystkie „sensowne” podzbiory \mathbb{R} , jakie jesteśmy sobie w stanie wyobrazić, w szczególności wszystkie przedziały

Definicja 1 Zmienną losową nazywamy mierzalną funkcję $X: \Omega \rightarrow \mathbb{R}$ (mierzalną tzn. taką, że dla każdego $B \in \mathcal{B}(\mathbb{R})$ mamy $X^{-1}[B] = \{\omega \in \Omega: X(\omega) \in B\} \in \mathcal{F}$).

Jest to definicja formalna. Ze względu na jej teoretyczny charakter nie będziemy się nią posługiwać. Do naszych rozważań wystarczy nam intuicyjne rozumienie pojęcia zmiennej losowej.

Przez **zmienną losową** będziemy rozumieli wielkość losową, którą można zmierzyć i pomiar ten wyrazić za pomocą liczby rzeczywistej.

Czasem będziemy się też odwoływać do pojęcia **wartości (realizacji) zmiennej losowej**. Będzie ono dla nas oznaczać wartość zmiennej losowej zmierzoną w konkretnym przypadku.

Zmienna losowa	Wartość zmiennej losowej
liczba oczek wyrzucona na kostce do gry	liczba oczek wyrzucona na kostce do gry w konkretnym rzucie
wzrost kobiety	wzrost ustalonej kobiety
masa mężczyzny	masa ustalonego mężczyzny
temperatura powietrza	temperatura powietrza zmierzona w danym miejscu i czasie
liczba prosiąt w miocie	liczba prosiąt w miocie z konkretnej ciąży u konkretnej lochy
ciśnienie skurczowe krwi obwodowej	ciśnienie skurczowe zmierzone w danym momencie u konkretnej osoby

Zmienne losowe zazwyczaj oznaczamy wielkimi literami a ich wartości odpowiadającymi im małymi literami.

Definicja 2 Dystrybucją zmiennej losowej X nazywamy funkcję $F_X: \mathbb{R} \rightarrow [0, 1]$ zadaną wzorem: $F_X(t) = P(X \leq t)$.

Twierdzenie 1 (Charakteryzacja dystrybucji) Funkcja $F: \mathbb{R} \rightarrow [0, 1]$ jest dystrybucją pewnej zmiennej losowej X wtedy i tylko wtedy, gdy

- funkcja F jest niemalejąca,
- funkcja F jest prawostronnie ciągła,
- $\lim_{t \rightarrow -\infty} F(t) = 0$,
- $\lim_{t \rightarrow \infty} F(t) = 1$.

Fakt 1 Dla dowolnych $a, b \in \mathbb{R}$, przy czym $a \leq b$, zachodzi:

$$\begin{aligned}P(X \leq a) &= F_X(a), & P(X < a) &= \lim_{t \rightarrow a^-} F_X(t), \\P(X \geq a) &= 1 - \lim_{t \rightarrow a^-} F_X(t), & P(X > a) &= 1 - F_X(a), \\P(a < X \leq b) &= F_X(b) - F_X(a), & P(a \leq X \leq b) &= F_X(b) - \lim_{t \rightarrow a^-} F_X(t),\end{aligned}$$

$$P(a \leq X < b) = \lim_{t \rightarrow b^-} F_X(t) - \lim_{t \rightarrow a^-} F_X(t), \quad P(a < X \leq b) = \lim_{t \rightarrow b^-} F_X(t) - F_X(a).$$

Jeśli dystrybuanta F_X jest funkcją ciągłą, to

$$P(X \leq a) = P(X < a) = F_X(a), \quad P(X \geq a) = P(X > a) = 1 - F_X(a),$$

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b) = F_X(b) - F_X(a).$$

W szczególności jeśli dystrybuanta F_X jest funkcją ciągłą, to $P(X = a) = 0$.

Definicja 3 Jeśli dla danej zmiennej losowej X istnieje funkcja $f_X: \mathbb{R} \rightarrow [0, \infty)$ taka że

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad \int_B f_X(t) dt = P(X \in B),$$

$$(\text{równoważnie:} \quad \forall t \in \mathbb{R} \quad \int_{-\infty}^t f_X(t) dt = P(X \leq t) = F_X(t) \quad)$$

to funkcję f_X nazywamy gęstością zmiennej losowej X .

Fakt 2 (Własności gęstości) Jeśli zmienna losowa X ma gęstość f_X , to wówczas

- $\lim_{t \rightarrow -\infty} f_X(t) = \lim_{t \rightarrow \infty} f_X(t) = 0$,
- $\int_{-\infty}^{\infty} f_X(t) dt = 1$,
- dla dowolnego $t \in \mathbb{R}$ zachodzi $F_X(t) = \int_{-\infty}^t f_X(u) du$.

Twierdzenie 2 Jeśli dystrybuanta F_X jest różniczkowalna w całym zbiorze \mathbb{R} lub też wszędzie poza borelowskim zbiorem $N \subseteq \mathbb{R}$ o tej własności, że $\int_{\mathbb{R}} \mathbf{1}_N(x) dx = 0$ (gdzie $\mathbf{1}_N(x) = 1$ wtedy i tylko wtedy, gdy $x \in N$) i $\int_{\mathbb{R} \setminus N} F'_X(x) dx = 1$, to zmienna losowa X posiada gęstość i dla $x \in \mathbb{R} \setminus N$ możemy przyjąć, że $f_X(x) = F'_X(x)$. Na zbiorze N gęstość f_X możemy określić dowolnie.

Fakt 3

- Jeśli dystrybuanta F_X nie jest ciągła, to nie istnieje gęstość (patrz: druga kropka powyżej).
- Jeśli dystrybuanta F_X jest ciągła a także różniczkowalna poza co najwyżej skończoną liczbą punktów (lub nawet co najwyżej przeliczalną liczbą [izolowanych] punktów), to możemy przyjąć, że $f_X = F'_X$ w tych punktach, w których F_X jest różniczkowalna. W punktach nieróżniczkowalności funkcji F_X gęstość f_X możemy określić dowolnie.
- Jeśli zmienna losowa ma gęstość, to zmiana wartości gęstości na dowolnym zbiorze $Q \subseteq \mathbb{R}$ o tej własności, że $\int_{\mathbb{R}} \mathbf{1}_Q(x) dx = 0$, nie ma wpływu na bycie gęstością i na rozkład zadany przez tę gęstość (bo zmiana ta nie zmienia całki).

Warto zwrócić uwagę, że każdy skończony bądź przeliczalny zbiór $Q \subseteq \mathbb{R}$ złożony z punktów izolowanych ma tę własność, że $\int_{\mathbb{R}} \mathbf{1}_Q(x) dx = 0$.

Każda zmienna losowa ma dystrybuantę, ale nie każda ma gęstość!

Definicja 4 Mówimy, że zmienne losowe X i Y są niezależne, jeśli

$$\forall s, t \in \mathbb{R} \quad P(X \leq s, Y \leq t) = P(X \leq s) \cdot P(Y \leq t) = F_X(s) \cdot F_Y(t).$$

Definicja 5 Mówimy, że zmienne losowe X i Y mają taki sam rozkład, jeśli mają te same dystrybuanty tzn. $F_X(t) = F_Y(t)$ dla każdego $t \in \mathbb{R}$.

Fakt 4 Jeśli zmienne losowe X i Y mają gęstości f_X i f_Y i gęstości te są sobie równe lub też równe poza borelowskim zbiorem $N \subseteq \mathbb{R}$ o tej własności, że $\int_{\mathbb{R}} \mathbf{1}_N(x) dx = 0$, to zmienne losowe X i Y mają taki sam rozkład.