RACHUNEK PRAWDOPODOBIEŃSTWA

WYKŁAD 3.

ZMIENNA LOSOWA JEDNOWYMIAROWA.

Zmienną losową X nazywamy funkcję (praktycznie każdą) przyporządkowującą zdarzeniom elementarnym liczby rzeczywiste.

$$X : \Omega \longrightarrow R$$

(dokładniej: przeciwobrazy zbiorów borelowskich powinny należeć do σ - ciała zdarzeń S, co w sposób równoważny można zapisać następująco $\bigwedge_{x \in R} \{\omega \in \Omega : X(\omega) < x\} \in S$).

Przykłady zmiennych losowych

Dla przestrzeni probabilistycznej – dwa rzuty kostką.

X – suma oczek, (wartości: 2, 3, ..., 12).

X – wynik rzutu o większej liczbie oczek (wartości: 1, 2, ..., 6).

Dla rzutu monetą aż wypadnie orzeł.

X – liczba rzutów monety **do** wypadnięcia pierwszego orła, (wartości: 0, 1, ...).

X – numer rzutu **w którym** wypadł pierwszy orzeł, (wartości: 1, 2, ...).

Dla losowego wyboru dwóch punktów w kole ośrodku O i promieniu R.

X – odległość między wylosowanymi punktami, (wartości: przedział [0, 2R]).

X – odległość środka odcinka utworzonego przez wylosowane punkty od punktu O , (wartości: przedział [0, R]).

Dla uproszczenia

- zapis $P(X < x)$ oznacza $P(\{\omega \in \Omega: X(\omega) < x\})$,
- zapis $P(X \leq x)$ oznacza $P(\{\omega \in \Omega: X(\omega) \leq x\})$,
- zapis $P(X = x)$ oznacza $P(\{\omega \in \Omega: X(\omega) = x\})$,
- zapis $P(X > x)$ oznacza $P(\{\omega \in \Omega: X(\omega) > x\})$,
- zapis $P(X \geq x)$ oznacza $P(\{\omega \in \Omega: X(\omega) \geq x\})$,

zapis $P(x_1 < X < x_2)$ oznacza

$P(\{\omega\in\Omega: x_1 < X(\omega) < x_2\})$,

- zapis $P(x_1 \le X < x_2)$ oznacza

  $P(\{\omega\in\Omega: x_1 \le X(\omega) < x_2\})$,

- zapis $P(x_1 < X \le x_2)$ oznacza

  $P(\{\omega\in\Omega: x_1 < X(\omega) \le x_2\})$,

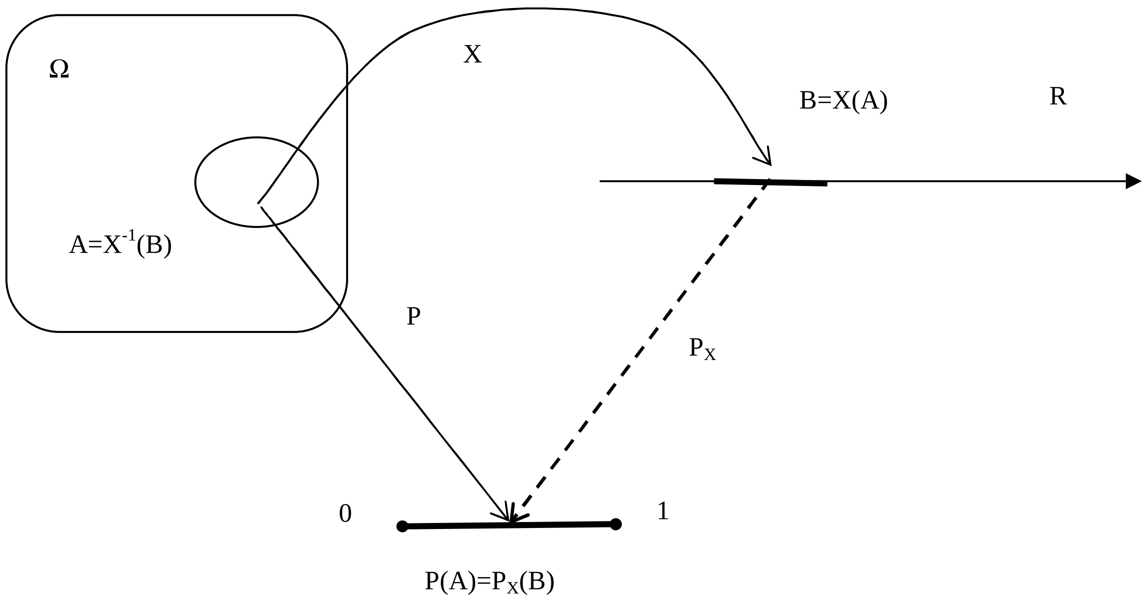
- zapis $P(x_1 \le X \le x_2)$ oznacza

  $P(\{\omega\in\Omega: x_1 \le X(\omega) \le x_2\})$, Zdarzeniom są przyporządkowane podzbiory zbioru R, musimy tym podzbiorom przyporządkować odpowiadające im prawdopodobieństwa. Przyporządkowanie to nazywamy **rozkładem prawdopodobieństwa** zmiennej losowej X i oznaczamy P_X.

$$P_X(B) = P\left(X^{-1}(B)\right) \quad \text{dla} \quad B \in \mathrm{B}(R),$$

$\mathrm{B}(R)$ - zbiory borelowskie

Tak określone P_X spełnia aksjomaty prawdopodobieństwa.


Ω
X
B=X(A)
R
A=X^-1(B)
P
P_X
0
1
P(A)=P_X(B)

Zmienne losowe równe sobie z prawdopodobieństwem 1 będziemy traktować jako nieodróżnialne.

Dystrybuantą zmiennej losowej X nazywamy funkcję $F: R \longrightarrow R$ określoną wzorem:

$$F(x) = P(X < x) = P_X((-\infty, x))$$

Własności dystrybuanty:

a) F jest funkcją niemalejącą,

b) F jest funkcją lewostronnie ciągłą,

c) $F(-\infty) = 0; \quad F(\infty) = 1$,

d) dystrybuanta zmiennej losowej wyznacza jednoznacznie jej rozkład,

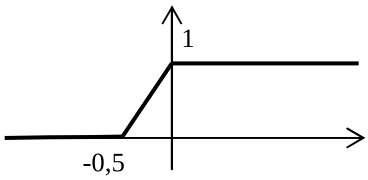
e) $P(a \leq X < b) = F(b) - F(a); \quad a < b$

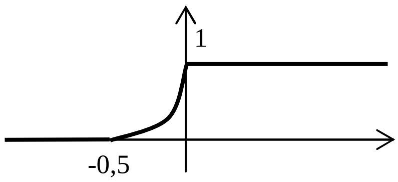
f) $P(X = a) = F(a^+) - F(a);$ gdzie $F(a^+)$ oznacza granicę prawostronną, (jeśli a jest punktem ciągłości dystrybuanty to $P(X = a) = 0$).

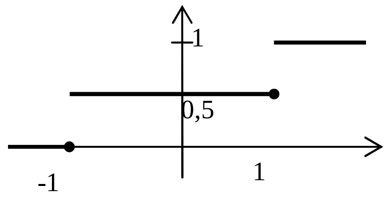
Uwaga
Jeśli funkcja rzeczywista spełnia własności a), b), c) to jest dystrybuantą pewnej zmiennej losowej, jej rozkład jest wyznaczony jednoznacznie co oznacza, że rozkłady zmiennych losowych określających taką dystrybuantę mają takie samo prawdopodobieństwo dla wszystkich zbiorów borelowskich.

Przykład.

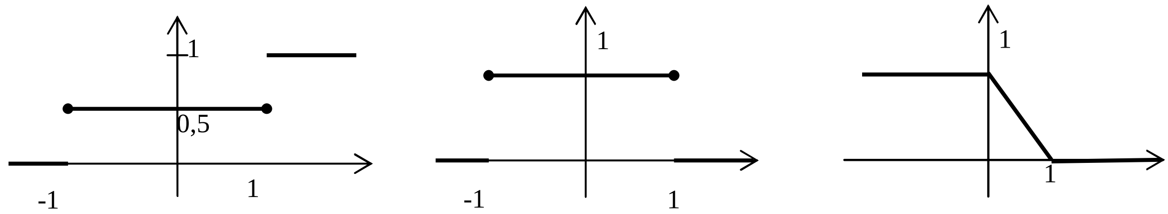
Poniższe funkcje są dystrybuantami.







Poniższe funkcje nie są dystrybuantami.



Nie sp. wł. b). Nie sp. wł. a), b), c). Nie sp. wł. a), c).

Przykład.

Dla jakich wartości A, B, C, D funkcja

$$F(x) = \begin{cases} A & \text{dla} \quad x \le 0 \\ Bx^2 & \text{dla} \quad 0 < x \le 1 \\ C(2-x)+1 & \text{dla} \quad 1 < x \le 2 \\ D & \text{dla} \quad x > 2 \end{cases}$$

jest dystrybuantą pewnej zmiennej losowej?

$A = 0$, $D = 1$ z własności c). Lewostronna ciągłość jest zapewniona. Należy sprawdzić własność a). Aby w przedziałach funkcje były niemalejące, musi być $B \geq 0$, $C \leq 0$; oprócz tego należy sprawdzić tą własność dla

$x = 1$ i $x = 2$.

Dla $x = 2$,   C - dowolne.

Dla $x = 1$ $B \leq C(2 - 1) + 1$, $B \leq 1$.

Zatem ostatecznie $0 \leq B \leq 1$ - dowolne, $0 \geq C \geq B - 1$.

Zmienna losowa jest **skokowa** (**dyskretna**) jeśli zbiór wszystkich jej wartości jest skończony lub przeliczalny.

Rozkład zmiennej losowej skokowej często określamy za pomocą **funkcji prawdopodobieństwa:**

$$\boxed{P(X = x_k) = p_k}$$

(własność: $\sum_k p_k = 1;\ p_k > 0$)

Liczby p_k nazywamy **skokami**, a wartości x_k **punktami skokowymi**.

Znając funkcję prawdopodobieństwa zmiennej losowej skokowej można wyznaczyć jej dystrybuantę

$$F(x) = \sum_{\substack{k \\ x_k < x}} p_k$$

oraz jej rozkład prawdopodobieństwa

$$P(X \in B) = \sum_{\substack{k \\ x_k \in B}} p_k$$

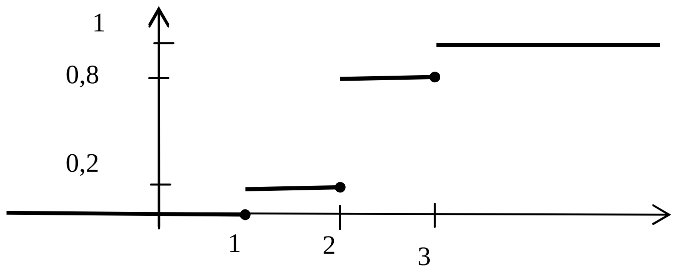
Przykład

Funkcja rozkładu prawdopodobieństwa zmiennej losowej X dana jest w tabeli:

x_k	1	2	3
p_k	0,2	0,6	0,2

Jej dystrybuanta ma postać

$$F(x) = \begin{cases} 0 & \text{dla} \quad x \le 1 \\ 0{,}2 & \text{dla} \quad 1 < x \le 2 \\ 0{,}8 & \text{dla} \quad 2 < x \le 3 \\ 1 & \text{dla} \quad x > 3 \end{cases}$$



Zauważmy, że punkty skokowe są punktami nieciągłości dystrybuanty a skoki wyznaczają przyrosty dystrybuanty (jej skoki) w tych punktach.

Zmienna losowa X o dystrybuancie *F* jest **ciągła** jeśli jej dystrybuanta da się przedstawić w postaci

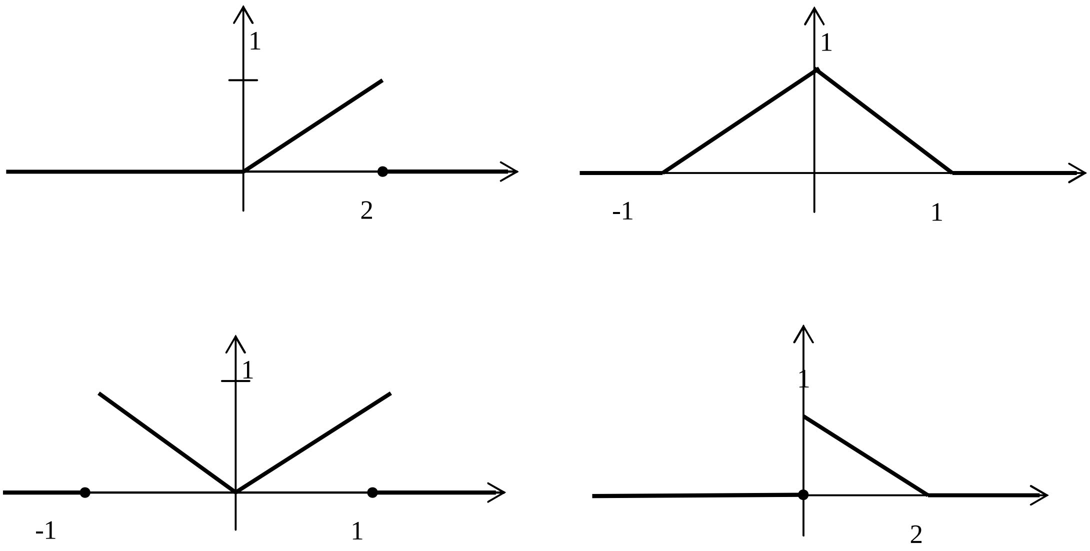
$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad x \in R$$

gdzie f jest funkcją spełniającą warunki:

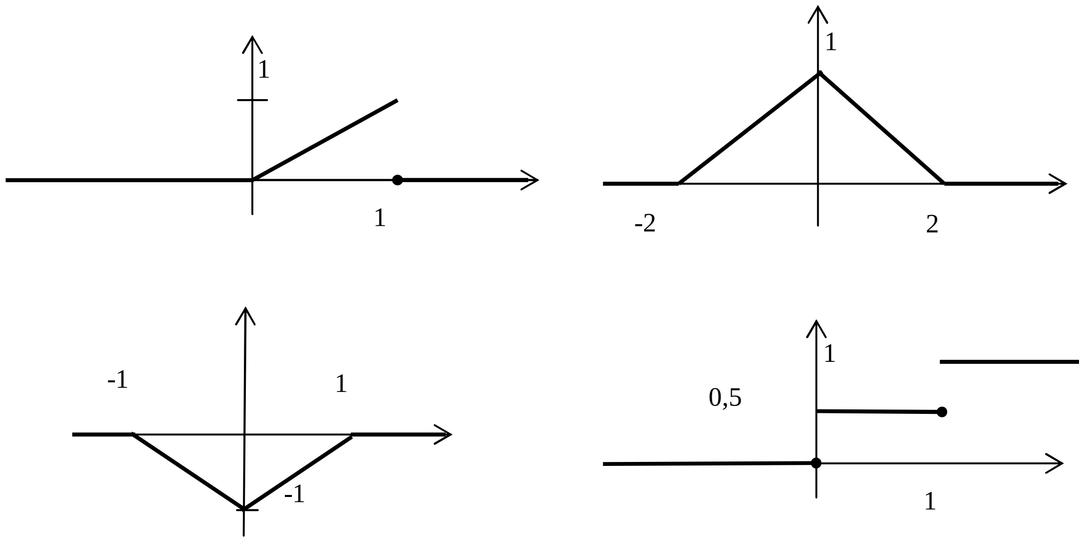
$$f(x) \geq 0; \quad x \in R; \qquad \int_{-\infty}^{\infty} f(t)dt = 1$$

i nazywamy ją **gęstością prawdopodobieństwa** zmiennej losowej X.

Przykład.



Te funkcje są gęstościami prawdopodobieństwa pewnych zmiennych losowych ciągłych.



Te funkcje nie są gęstościami prawdopodobieństwa.

Własności zmiennej losowej ciągłej:

a) $\boxed{P(X < a) = \int_{-\infty}^{a} f(x)dx = F(a)}$,

b) $\boxed{\begin{array}{l} P(a \le X \le b) = P(a < X \le b) = P(a \le X < b) = \\ = P(a < X < b) = \int_{a}^{b} f(x)dx = F(b) - F(a) \end{array}}$

c) $P(X > b) = \int_{b}^{\infty} f(x)dx = 1 - F(b)$,

d) $P(X = a) = 0, \text{ dla dowolnego } a \in R$;

(brak punktów skokowych),

e) F jest funkcją ciągłą i prawie wszędzie różniczkowalną $F'(x) = f(x)$ (równość zachodzi dla punktów ciągłości gęstości). Wyznaczając gęstość przez różniczkowanie dystrybuanty, w punktach w których F nie jest różniczkowalna można przyjąć, że gęstość jest równa zero.

Przykład.

Wyznaczymy wartości c dla której funkcja

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

jest gęstością pewnej zmiennej losowej ciągłej?

Aby gęstość była nieujemna i

$$\int_{-\infty}^{\infty} f(x)dx = 1,$$

musi być c > 0 i pole odpowiedniego trójkąta prostokątnego równe 1. Stąd c = 2.

Dystrybuanta tej zmiennej losowej ma postać

dla $x \in (-\infty\,;\,0]$ $$F(x) = \int_{-\infty}^{x} 0dt = 0$$

dla $x \in (0\,,1]$ $$F(x) = \int_{-\infty}^{0} 0dt + \int_{0}^{x} 2tdt = x^2$$

dla $x \in (1\,,\infty]$ $$F(x) = \int_{-\infty}^{0} 0dt + \int_{0}^{1} 2tdt + \int_{1}^{x} 0dt = 1$$

Ostatecznie

$$F(x)=\begin{cases}0 & \text{dla} \quad x\in(-\infty;\,0] \\ x^2 & \text{dla} \quad x\in(0,1] \\ 1 & \text{dla} \quad x\in(1,\infty]\end{cases}$$

Obliczymy prawdopodobieństwo

$P(0{,}25 \le X \le 0{,}75)$.

Sposób I. Za pomocą gęstości

$$P(0{,}25 \le X \le 0{,}75) = \int_{0,25}^{0,75} 2x\,dx = x^2\Big|_{0,25}^{0,75} = 0{,}5$$

Sposób II. Za pomocą dystrybuanty

$$P(0{,}25 \le X \le 0{,}75) = F(0{,}75) - F(0{,}25) = 0{,}5$$

Twierdzenie Lebesgue'a o **rozkładzie dystrybuanty**.
Każdą dystrybuantę F można przedstawić w postaci

$$F = c_1 F_1 + c_2 F_2 + c_3 F_3$$

gdzie

$$c_1 + c_2 + c_3 = 1, \quad c_1, c_2, c_3 \geq 0$$

F_1 - dystrybuanta skokowa,
F_2 - dystrybuanta ciągła,
F_3 - dystrybuanta osobliwa,

Uwagi o rozkładzie funkcji zmiennej losowej.

Jeśli X - skokowa, o funkcji prawdopodobieństwa

$P(X = x_i) = p_i$,

g - dowolna to funkcja prawdopodobieństwa

zmiennej losowej $Y = g(X)$ ma postać:

$g(x_1)$	$g(x_2)$	...	$g(x_k)$
p_1	p_2	...	p_k

Po uporządkowaniu rosnąco wartości $g(x_i)$ i **zsumowaniu** odpowiednich prawdopodobieństw.

Dokładniej

$$P(Y=y)=P(g(X)=y)=P\left(\bigcup_{\{i:g(x_i)=y\}}(X=x_i)\right)=$$

$$=\sum_{\{i:g(x_i)=y\}}P(X=x_i)=\sum_{\{i:g(x_i)=y\}}p_i$$

Przykład.

X - zmienna losowa skokowa o funkcji prawdopodobieństwa:

-4	-2	-1	0	1	2
0,4	0,1	0,1	0,1	0,1	0,2

wyznaczymy funkcję prawdopodobieństwa zmiennej losowej Y = sgnX .

sgn(-4) = sgn(-2) = sgn(-1) = -1.
sgn(0) = 0.
sgn(1) = sgn(2) = 1.
Zatem funkcja prawdopodobieństwa zmiennej losowej Y jest następująca

-1	0	1
0,6	0,1	0,3

X - dana zmienna losowa ciągła o **gęstości** f.

Y = g(X) g - borelowska,

tzn. $g^{-1}(B) \in \mathcal{B}(R)$ dla $B \in \mathcal{B}(R)$,

Wyznaczyć gęstość g(y) zmiennej losowej Y.

1) Jeśli g - ściśle monotoniczna i różniczkowalna w przedziale (a, b) koncentracji X to:

$$g(y) = f\left(h(y)\right)\left|h^{'}(y)\right|$$

gdzie $h = g^{-1}$.

Należy pamiętać o przekształceniu przedziału koncentracji.

Przykład.

Jeśli X ma rozkład o gęstości

$$f(x)=\begin{cases}0 & \text{dla } x\le 0\\ e^{-x} & \text{dla } x>0\end{cases}\qquad Y=\sqrt{X}-2,$$

wtedy $h(y)=(y+2)^2$, $h'(y)=2(y+2)$,

g(0) = -2, g(∞) = ∞,

$$g(y)=\begin{cases}0 & \text{dla } x\le -2\\ 2(y+2)e^{-(y+2)^2} & \text{dla } x>-2\end{cases},$$

2) Jeśli g - przedziałami ściśle monotoniczna i różniczkowalna w przedziale (a, b) koncentracji X to:

$$g(y) = \sum_{i=1}^{k} f\left(h_i(y)\right)\left|h_i^{'}(y)\right|$$

gdzie h_i - funkcje odwrotne do g dla poszczególnych przedziałów,
k - liczba wartości funkcji odwrotnej odpowiadających danemu y.

Przykład.

Y = |X|, wtedy $g(y) = f(-y) + f(y) \quad y > 0$,

Przykład.

$Y = X^2$, wtedy

$$g(y) = f\left(-\sqrt{y}\right)\frac{1}{2\sqrt{y}} + f\left(\sqrt{y}\right)\frac{1}{2\sqrt{y}} \quad y > 0,$$

W niektórych zagadnieniach wyznaczania rozkładu funkcji zmiennej losowej najpierw wyznaczamy dystrybuantę rozkładu zmiennej losowej Y = g(X), wg schematu

$$F_Y(y) = P(Y < y) = P(g(X) < y) = P\left(X \in g^{-1}((-\infty, y)) < y\right)$$

następnie jeśli to możliwe, wyznaczamy funkcję prawdopodobieństwa (gdy jest to rozkład skokowy) lub gęstość (gdy jest to rozkład ciągły).

Przykład.

Jeśli X ma rozkład o gęstości

$$f(x) = \begin{cases} 0 & \text{dla } x \notin [0\,,\,3] \\ \dfrac{1}{3} & \text{dla } x \in [0\,,\,3] \end{cases}$$ (rozkład jednostajny na [0, 3])

$$Y = \max(2, X),$$

wtedy

$$F_Y(y) = P\big(Y < y\big) = P\big(\max(2, X) < y\big) =$$

$$= \begin{cases} 0 & \text{dla} \quad \mathrm{y} \le 2 \\ \dfrac{1}{3}y & \text{dla} \quad 2 < \mathrm{y} \le 3 \\ 1 & \text{dla} \quad \mathrm{y} > 3 \end{cases}$$

Nie jest to ani rozkład skokowy ani ciągły. Nie można więc wyznaczyć ani funkcji prawdopodobieństwa ani gęstości.
Jest to rozkład mieszany skokowo - ciągły i zgodnie z twierdzeniem o rozkładzie dystrybuanty powyższą dystrybuantę można przedstawić w postaci

$$F_Y = c_1 F_1 + c_2 F_2$$

gdzie $c_1 = 2/3$, $F_1(y) = \begin{cases} 0 & \text{dla} \quad y \le 2 \\ 1 & \text{dla} \quad y > 2 \end{cases}$,

$c_2 = 1/3$, $F_2(y) = \begin{cases} 0 & \text{dla} \quad y \le 2 \\ y - 2 & \text{dla} \quad 2 < y \le 3 \\ 1 & \text{dla} \quad y > 3 \end{cases}$,