

Dystrybuanty i gęstości — definicje, fakty, przykłady

4 grudnia 2017

1 Dystrybuanty

Definicja 1.1. *Dystrybuantą rozkładu prawdopodobieństwa zmiennej losowej X , o wartościach w $\mathbb{R}$, nazywamy funkcję*

$$F(x) := P(X \leq x) .$$

Fakt 1. *Powyżej zdefiniowana dystrybuanta jest prawostronnie ciągła i posiada lewostronne granice, tj.*

$$(\forall x_0) \quad \lim_{x \rightarrow x_0^+} F(x) = F(x_0) \quad (1)$$

oraz

$$(\forall x_0) (\exists g \in [0, 1]) \quad \lim_{x \rightarrow x_0^-} F(x) = g . \quad (2)$$

UWAGA:

Istnieje inna, alternatywna definicja dystrybuanty:

$$F(x) = P(X < x) .$$

Wtedy dystrybuanta jest lewostronnie ciągła i posiada prawostronne granice, tj. „+” i „-” zamieniają się rolami w równaniach (1) oraz (2).

Fakt 2. *Każda dystrybuanta $F : \mathbb{R} \rightarrow \mathbb{R}$ (prawostronnie ciągła) spełnia poniższe własności:*

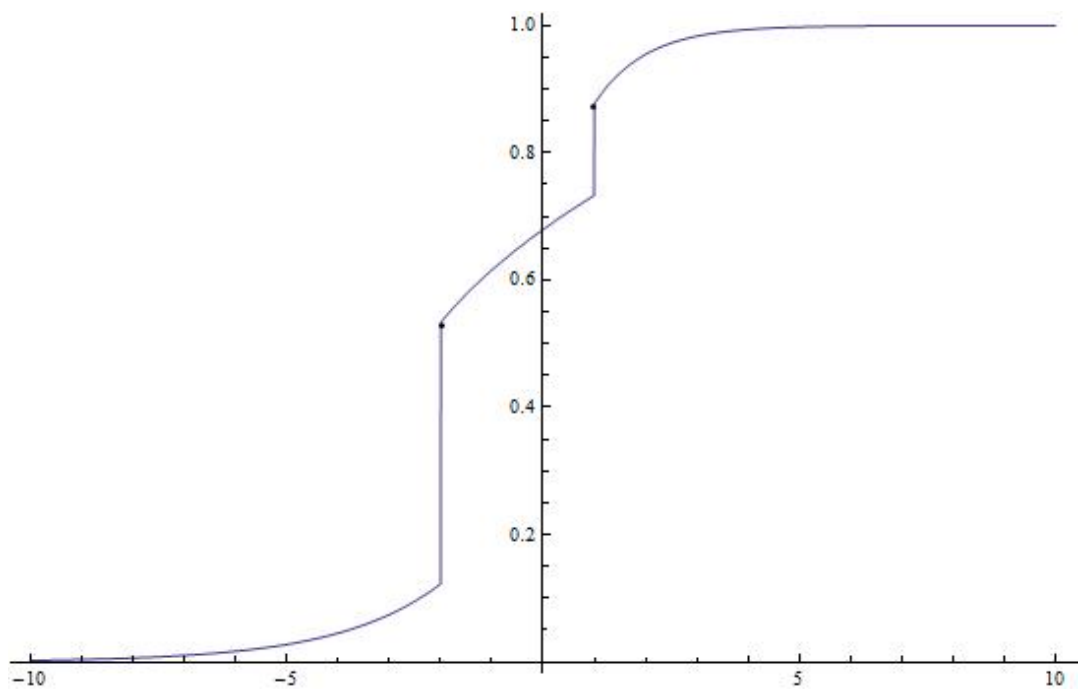
1. $\lim_{x \rightarrow -\infty} F(x) = 0$,
2. $\lim_{x \rightarrow \infty} F(x) = 1$,
3. F jest niemalejąca,
4. F jest prawostronnie ciągła i posiada lewostronne granice.

UWAGA:

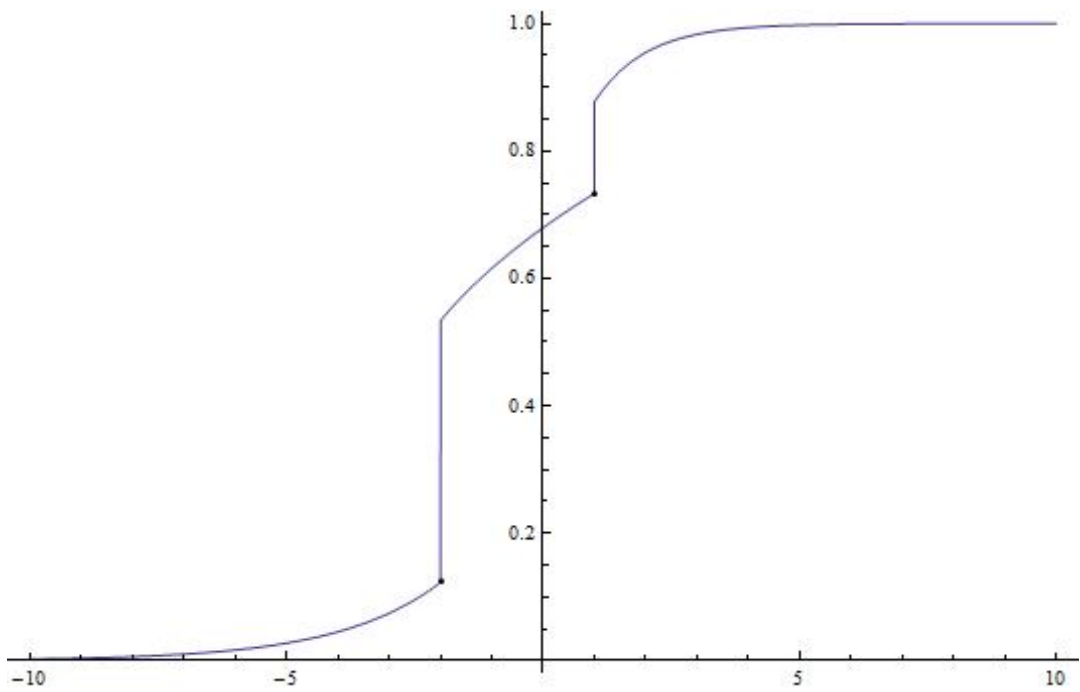
Dla dystrybuant lewostronnie ciągłych, w ostatniej własności należy zamienić miejscami fragmenty „lewo-” z „prawo-”.

Fakt 3. *Będziemy wyróżniać 3 rodzaje dystrybuant:*

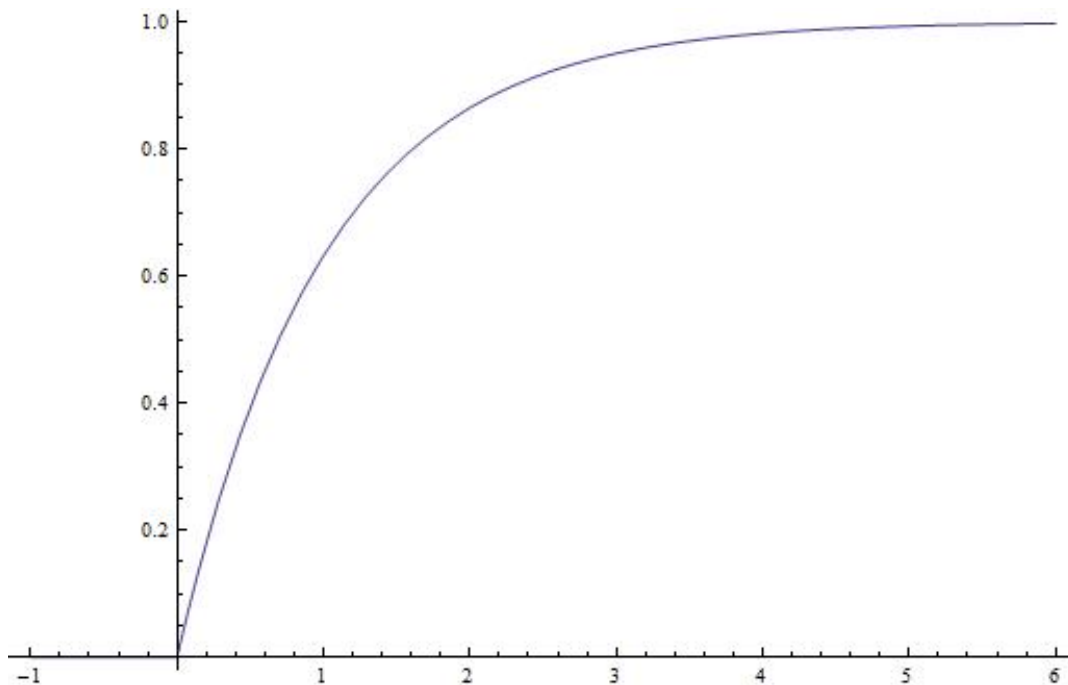
- Ciągłą (absolutnie)
- Skokową
- Ciągło-skokową



Rysunek 1: Przykład dystrybuanty prawostronnie ciągłej



Rysunek 2: Przykład dystrybuanty lewostronnie ciągłej



Rysunek 3: Przykład dystrybuanty ciągłej absolutnie

1.1 Dystrybuanta ciągła

Fakt 4. W pierwszym przypadku, istnieje funkcja f , którą nazywamy gęstością rozkładu prawdopodobieństwa. Spełnia ona warunek:

$$F(x) = \int_{-\infty}^x f(t)dt, \quad (3)$$

a w szczególności:

$$1 = \int_{-\infty}^{\infty} f(t)dt. \quad (4)$$

(Osoby znające całkę Stieltjesa, mogą zapisać powyższe całki jako $\int dF(t)$ w odpowiednich granicach.)

Definicja 1.2. Wartość oczekiwanej funkcji zmiennej losowej $\varphi(X)$ wyliczamy wtedy ze wzoru:

$$\mathbb{E}\varphi(X) = \int_{-\infty}^{\infty} \varphi(x)f(x)dx,$$

jak np.

$$\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx.$$

1.2 Dystrybuanta skokowa

W drugim przypadku istnieje ciąg punktów (x_i) , które są możliwymi wartościami zmiennej X . Ciąg ten może być skończony, ale nie musi (np. ile razy trzeba rzucać kostką sześcienną, by w końcu wyrzucić "6"?). Każdy skok dystrybuanty odpowiada wtedy prawdopodobieństwu p_i (p_i są odpowiednikami gęstości) wylosowania danej liczby x_i .

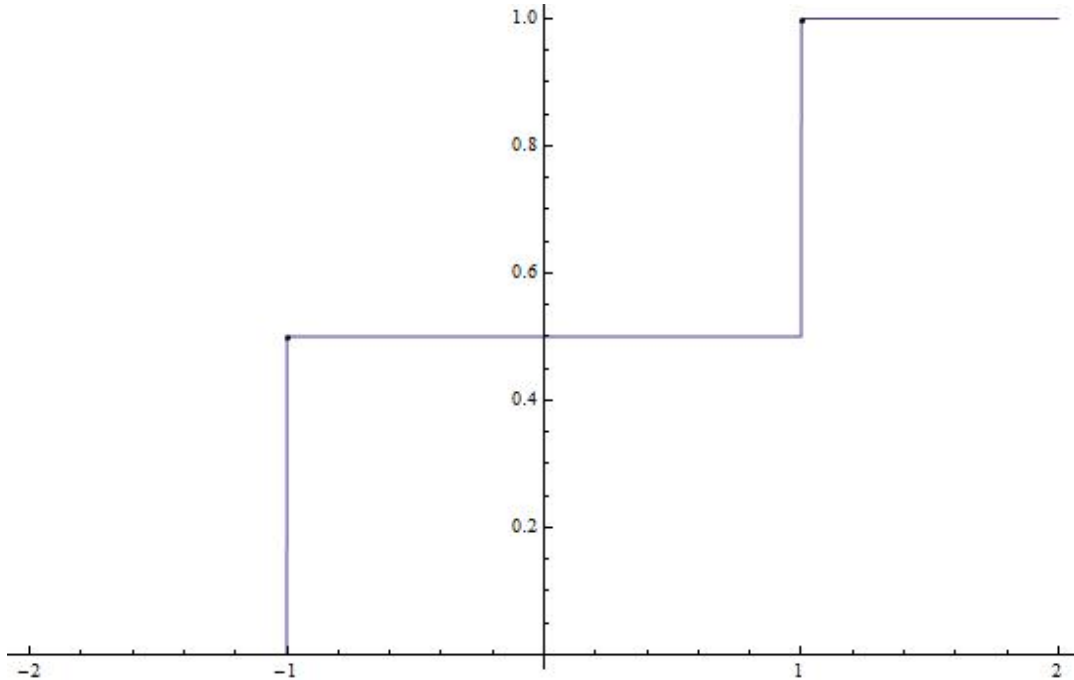
UWAGA:

X może przyjąć inne wartości, niż x_i z prawdopodobieństwem 1, ale to się "raczej" nie zdarza.

Przykład 1. *Dzielimy płaszczyznę prostą na dwie części: A i B . Jeśli wylosujemy punkt z A , to X przyjmuje wartość -1 , gdy wylosujemy punkt z B – $X = 1$, a gdy wylosujemy punkt z prostej, to $X = 0$. Wtedy $P(X = -1) = P(X = 1) = \frac{1}{2}$, a $P(X = 0) = 0$. Dystrybuanta (prawostronnie ciągła) jest wtedy określona wzorem:*

$$F(x) = \begin{cases} 0 & , \text{ gdy } x < -1; \\ \frac{1}{2} & , \text{ gdy } -1 \leq x < 1; \\ 1 & , \text{ gdy } x \geq 1; \end{cases}$$

Jak widać, w zerze nie ma wtedy atomu (skoku), a mimo wszystko da się wylosować 0.



Rysunek 4: Wykres dystrybuanty z Przykładu 1

Fakt 5. *W przypadku, gdy dystrybuanta zmiennej X ma skończenie wiele skoków, to zachodzi analogon równości (3):*

$$F(x) = \sum_{x_i \leq x} p_i ,$$

i (4):

$$1 = \sum_{i=1}^n p_i ,$$

gdzie n to liczba atomów dystrybuanty.

Definicja 1.3. *Podobnie jak poprzednio, określamy również wartości oczekiwane funkcji zmiennych losowych:*

$$\mathbb{E}\varphi(X) = \sum_{i=1}^n \varphi(x_i) p_i ,$$

w szczególności:

$$\mathbb{E}X = \sum_{i=1}^n x_i p_i .$$

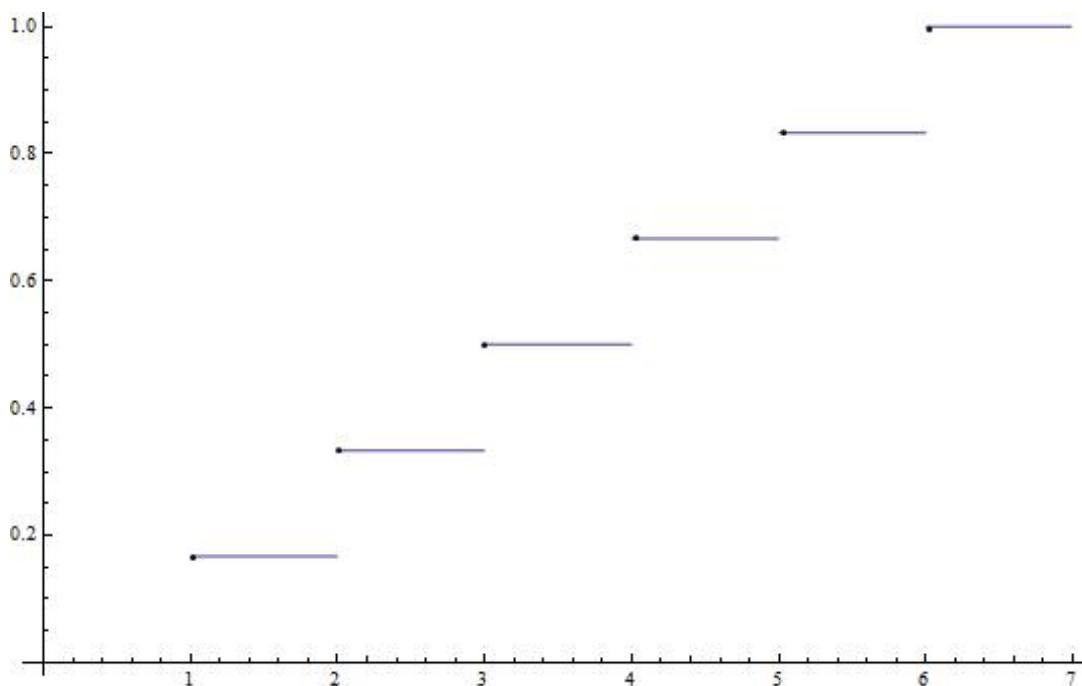
Przykład 2. Niech X przyjmuje n wartości. Każde z prawdopodobieństwem $\frac{1}{n}$ (np. $n = 6$ dla rzutu kostką sześcienną).

Wtedy średnia (wartość oczekiwana zmiennej X) jest dana wzorem:

$$\mathbb{E}X = \frac{\sum_{i=1}^n x_i}{n}.$$

W przypadku kostki sześcienniej:

$$\mathbb{E}X = \frac{\sum_{i=1}^6 i}{6}.$$



Rysunek 5: Wykres dystrybuanty dla kostki sześcienniej z Przykładu 2

BARDZO WAŻNA UWAGA:

Gdy mamy do czynienia z rozkładem, który ma nieskończenie wiele skoków, to jedyne zmiany we wzorach to zamiany n na ∞ .

Przykład 3. (Rozkład geometryczny) Student inżynierii biomedycznej zdaje kurs prawdopodobieństwa z prawdopodobieństwem 0.75, niezależnie od tego, czy podchodził to takiego kursu wcześniej, czy nie. Jakie jest prawdopodobieństwo, że zda w co najwyżej 8 podejściach?

Niech X oznacza liczbę podejść studenta. Wtedy

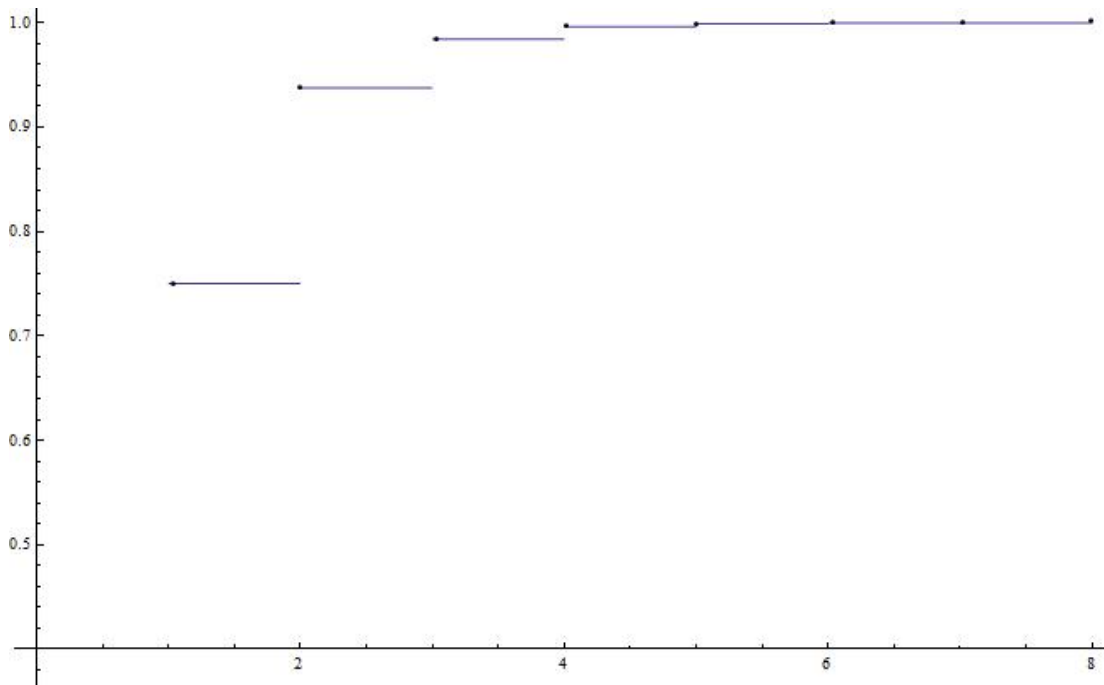
$$P(X = k) = 0.25^{k-1} \cdot 0.75, \quad k = 1, 2, \dots$$

Można wtedy wyznaczyć dystrybuantę:

$$F(x) = \sum_{k \leq x} k \cdot \frac{3}{4} \left(\frac{1}{4} \right)^{k-1}.$$

Chcemy obliczyć $P(X \leq 8)$, czyli $F(8)$:

$$F(8) = \sum_{k=1}^8 k \cdot \frac{3}{4} \left(\frac{1}{4} \right)^{k-1} = \frac{3}{4} \frac{1 - \left(\frac{1}{4} \right)^8}{1 - \frac{1}{4}} = 1 - \left(\frac{1}{4} \right)^8.$$



Rysunek 6: Wykres dystrybuanty dla rozkładu $\mathcal{Geo}(\frac{1}{4})$ z Przykładu 3

1.3 Dystrybuanta ciąгло-skokowa

W trzecim rozważanym przypadku, dystrybuanta to kombinacja dwóch poprzednich. Wszelkie wzory zachodzą w bardzo analogiczny sposób:

$$F(x) = \int_{-\infty}^x f(t)dt + \sum_{x_i \leq x} p_i ,$$

$$1 = \int_{-\infty}^{\infty} f(t)dt + \sum_i p_i ,$$

$$\mathbb{E}\varphi(X) = \int_{-\infty}^{\infty} \varphi(x)f(x)dx + \sum_i \varphi(x_i)p_i ,$$

$$\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)dx + \sum_i x_i p_i .$$

Fakt 6. Można wyliczać prawdopodobieństwo dla konkretnych wartości x_0 , przy użyciu dystrybuanty:

$$P(X = x_0) = F(x_0^+) - F(x_0^-) ,$$

gdzie $F(x_0^\pm) := \lim_{x \rightarrow x_0^\pm} F(x)$ to granice jednostronne („+” – prawo-, a „lewo-”).

ISTOTNA UWAGA:

Wzór ten ma sens stosować dla rozkładów skokowych lub ogólniej — ciąгло-skokowych.

Warto zauważyć, że ciągłość dystrybuant powoduje, że: dla prawostronnych — $F(x^+) = F(x)$, a dla lewostronnych — $F(x^-) = F(x)$.

Przykłady dystrybuant ciąгло-skokowych są przedstawione na Rysunkach 1,2,9,11 i 13.

2 Gęstość

Fakt 7. *Dystrybuanta ciągła (ciągło-skokowa) posiada gęstość prawdopodobieństwa (odpowiednio: niepełną gęstość pr.), daną wzorem:*

$$f(x) = F'(x) .$$

UWAGA:

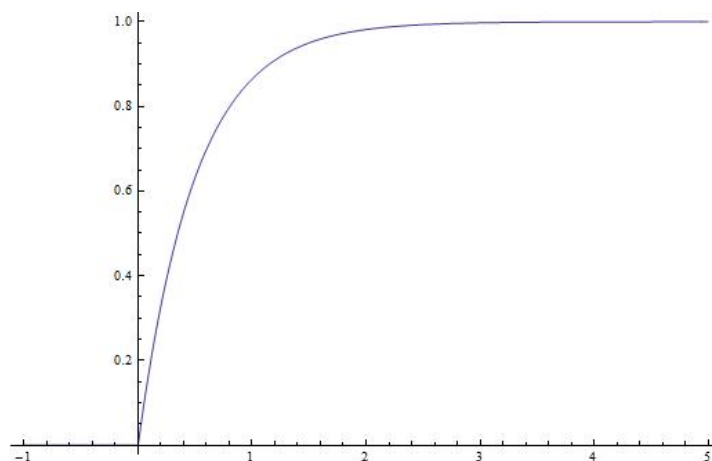
Funkcja f , która pojawiała się w poprzedniej sekcji jako funkcja podcałkowa jest gęstością prawdopodobieństwa.

Przykład 4. *Znaleźć gęstość rozkładu zmiennej X o dystrybuancie zadanej wzorem:*

$$F(x) = \begin{cases} 1 - e^{-2x}, & \text{dla } x \geq 0 \\ 0, & \text{dla } x < 0 \end{cases} ,$$

a następnie obliczyć wartość oczekiwaną z e^X .

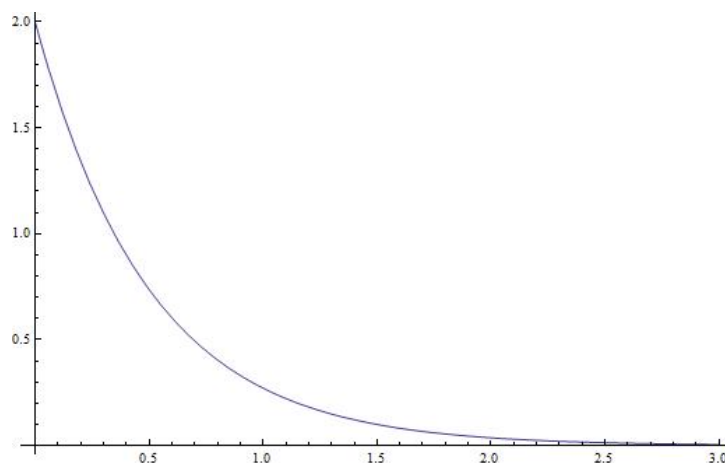
Zaczynamy od policzenia:



Rysunek 7: Wykres dystrybuanty z Przykładu 4

$$F'(x) = f(x) = \begin{cases} 2e^{-2x}, & \text{dla } x \geq 0 \\ 0, & \text{dla } x < 0 \end{cases} .$$

Następnie sprawdzamy, czy $F(x)$ jest ciągła, żeby sprawdzić, czy nie ma przypadkiem skoków. Jedyne



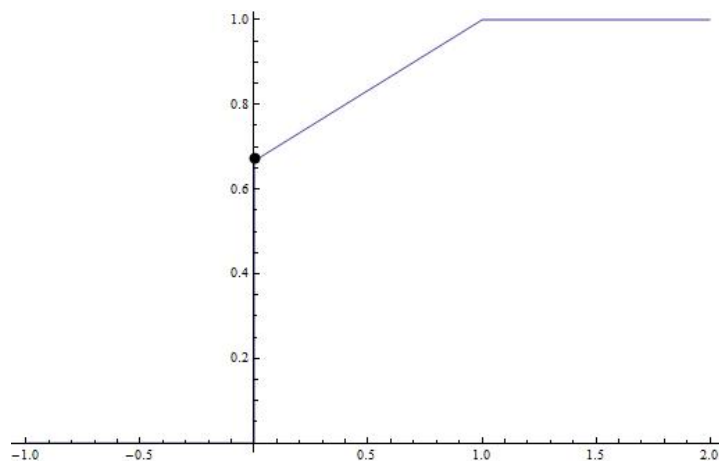
Rysunek 8: Wykres gęstości z Przykładu 4

podejrzany punkt to 0, więc liczymy $F(0) = 0$ oraz $F(0^-) = 0$. Zatem F — ciągła. Następne zadanie to policzenie:

$$\mathbb{E}e^X = \int_0^{\infty} e^x f(x) dx = 2 \int_0^{\infty} e^{-x} dx = 2(-e^{-x}) \Big|_0^{\infty} = 0 + 2 = 2 .$$

Przykład 5. Dana jest dystrybuanta:

$$F(x) = \begin{cases} 0, & \text{dla } x < 0 \\ \frac{2+x}{3}, & \text{dla } 0 \leq x < 1 \\ 1, & \text{dla } x \geq 1 \end{cases} .$$



Rysunek 9: Wykres dystrybuanty $F(x)$ z Przykładu 5

Jaki to rodzaj zmiennej? Znaleźć gęstość zmiennej X oraz dystrybuantę, gęstość i typ zmiennej warunkowej $X|X > 0$. Obliczyć $\mathbb{E}X$ oraz $\text{Var}X$.

Policzmy gęstość:

$$f_X(x) = \begin{cases} 0, & \text{dla } x < 0 \\ \frac{1}{3}, & \text{dla } 0 \leq x < 1 \\ 0, & \text{dla } x \geq 1 \end{cases} .$$

Niestety

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^1 \frac{1}{3} dx = \frac{1}{3} ,$$

więc zmienna X jest typu ciągło-skokowego, a suma długości skoków wynosi $1 - \frac{1}{3} = \frac{2}{3}$.

W jakich punktach dochodzi do skoków? Na trzech zdefiniowanych fragmentach dystrybuanta jest ciągła, więc pozostają dwa punkty podejrzane: 0 i 1. $F(0) = \frac{2}{3}$, a $F(0^-) = 0$, co oznacza, że jedyny skok jest w zerze, bo $P(X = 0) = \frac{2}{3}$, a to suma długości wszystkich skoków.

Aby wyznaczyć gęstość zmiennej warunkowej, należy skorzystać ze wzoru:

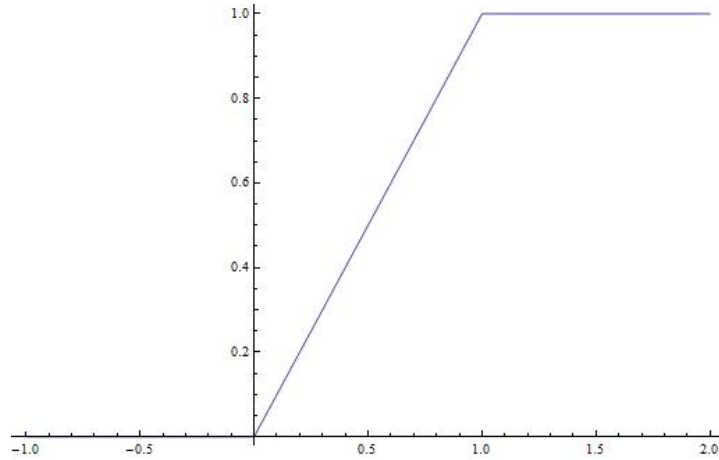
$$f_{X|X>0}(x) = \frac{f_X(x)}{P(X > 0)} = \frac{f_X(x)}{1 - P(X \leq 0)} = \frac{f_X(x)}{1 - F(0)} = 3f_X(x) = \begin{cases} 0, & \text{dla } x < 0 \\ 1, & \text{dla } 0 \leq x < 1 \\ 0, & \text{dla } x \geq 1 \end{cases} ,$$

co jest gęstością rozkładu jednostajnego $\mathcal{U}(0, 1)$.

Aby wyliczyć dystrybuantę, korzystamy ze wzoru:

$$F_{X|X>0}(x) = \int_{-\infty}^x 3f_X(x)dx = \begin{cases} 0, & \text{dla } x < 0 \\ \int_0^x 1dx, & \text{dla } 0 \leq x < 1 \\ \int_0^1 1dx, & \text{dla } x \geq 1 \end{cases} = \begin{cases} 0, & \text{dla } x < 0 \\ x, & \text{dla } 0 \leq x < 1 \\ 1, & \text{dla } x \geq 1 \end{cases}$$

Łatwo zauważyć, że to zmienna typu ciągłego.



Rysunek 10: Wykres dystrybuanty $F_{X|X>0}(x)$ z Przykładu 5

Przejdźmy do wartości oczekiwanej:

$$\mathbb{E}X = \int_0^1 x \frac{1}{3} dx + 0 \cdot 23 = \frac{1}{6} (x^2) \Big|_0^1 = \frac{1}{6}$$

oraz wariancji:

$$\text{Var}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - 2\mathbb{E}(X\mathbb{E}X) + (\mathbb{E}X)^2 = \mathbb{E}X^2 - 2\mathbb{E}X \cdot \mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X^2 - \frac{1}{36}.$$

Policzmy zatem drugi moment:

$$\mathbb{E}X^2 = \int_0^1 x^2 \frac{1}{3} dx + 0^2 \cdot 23 = \frac{1}{9} (x^3) \Big|_0^1 = \frac{1}{9},$$

stąd:

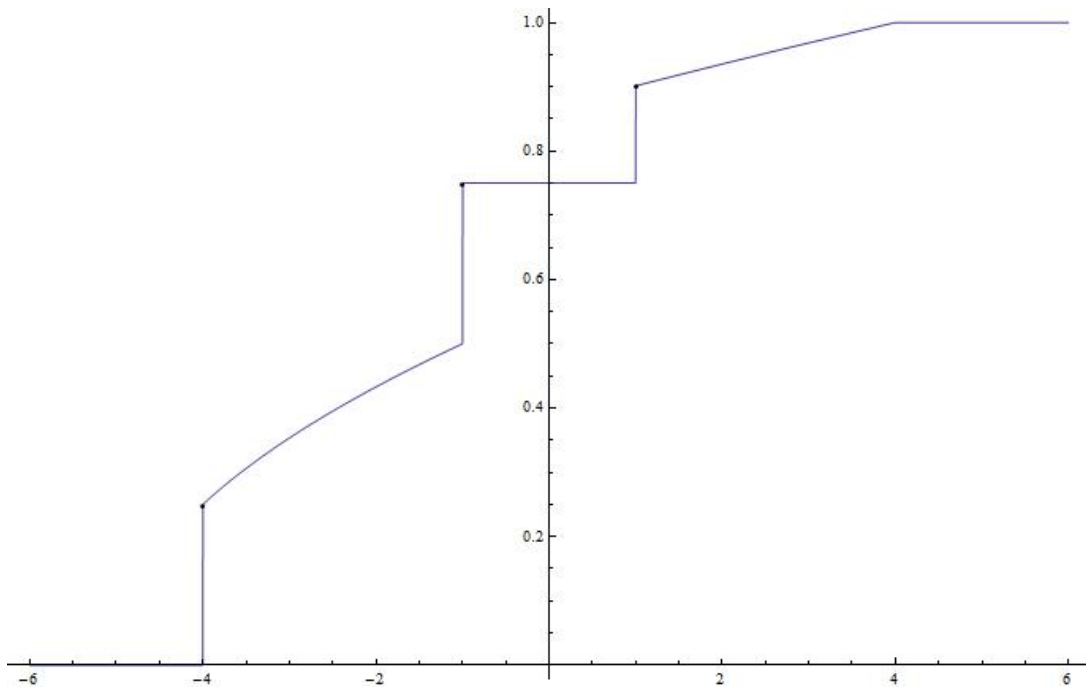
$$\text{Var}X = \frac{1}{9} - \frac{1}{36} = \frac{3}{36} = \frac{1}{12}.$$

Przykład 6. Niech dystrybuanta zmiennej X będzie zadana wzorem:

$$F(x) = \begin{cases} 0, & \text{dla } x < -4 \\ c\sqrt{x+5}, & \text{dla } -4 \leq x < -1 \\ \frac{3}{4}, & \text{dla } -1 \leq x < 1 \\ c\sqrt{x+12}, & \text{dla } 1 \leq x < 4 \\ 1, & \text{dla } x \geq 4 \end{cases}.$$

Wyznaczyć parametr c , jeśli wiadomo, że $P(X = -1) = \frac{1}{4}$. Następnie wyznaczyć:

1. $P(X \in \{-4, 1\})$,
2. $P(X = 4)$,
3. $P(X < 1)$,
4. $P(X \leq 1)$,
5. $P(-5 \leq 2X + 3 < 3)$,
6. $P(\ln X \geq 1)$,
7. $P(|X| \geq 1)$,
8. gęstość zmiennej X ,
9. wartość oczekiwaną zmiennej $Y = \sqrt{X + 3.5 \frac{X}{|X|}} + 8.5$, przy założeniu, że Y jest dobrze określona oraz dystrybuantę tej zmiennej i jej gęstość.



Rysunek 11: Wykres dystrybuanty $F_X(x)$ z Przykładu 6

Zauważmy, że

$$P(X = -1) = F(-1) - F(-1^-) = \frac{3}{4} - \lim_{x \rightarrow -1^-} c\sqrt{x+5} = \frac{3}{4} - c\sqrt{4} = \frac{3}{4} - 2c = \frac{1}{4},$$

więc $c = \frac{1}{4}$.

1. W -4 oraz 1 są skoki, bo:

$$P(X \in \{-4, 1\}) = (F(-4) - F(-4^-)) + (F(1) - F(1^-)) = \left(\frac{1}{4}\sqrt{1} - 0\right) + \left(\frac{1}{4}\sqrt{13} - \frac{3}{4}\right) = \frac{\sqrt{13} - 2}{4}.$$

2. Można zauważyć, że w 4 nie ma skoku, gdyż:

$$P(X = 4) = (F(4) - F(4^-)) = 1 - \frac{1}{4}\sqrt{16} = 0 .$$

3.

$$P(X < 1) = \lim_{x \rightarrow 1^-} P(X \leq x) = F(1^-) = \frac{3}{4} .$$

4.

$$P(X \leq 1) = F(1) = \frac{1}{4}\sqrt{13} .$$

5.

$$\begin{aligned} P(-5 \leq 2X + 3 < 3) &= P(-8 \leq 2X < 0) = P(-4 \leq X < 0) = P(X < 0) - P(X < -4) = \\ &= F(0^-) - F(4^-) = \frac{3}{4} - 0 = \frac{3}{4} . \end{aligned}$$

6.

$$P(\ln X \geq 1) = P(X \geq e) = 1 - P(X < e) = 1 - F(e^-) = 1 - \frac{\sqrt{12+e}}{4} .$$

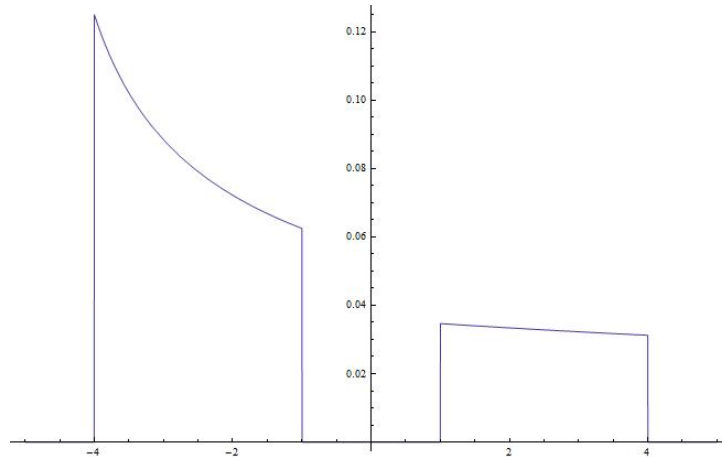
7.

$$\begin{aligned} P(|X| \geq 1) &= 1 - P(|X| < 1) = 1 - P(-1 < X < 1) = 1 - (P(X < 1) - P(X \leq -1)) = \\ &= 1 - F(1^-) + F(-1) = 1 - \frac{3}{4} + \frac{3}{4} = 1 . \end{aligned}$$

8.

$$f(x) = \begin{cases} 0, & \text{dla } x < -4 \\ \frac{1}{4} \frac{1}{2\sqrt{x+5}}, & \text{dla } -4 \leq x < -1 \\ 0, & \text{dla } -1 \leq x < 1 \\ \frac{1}{4} \frac{1}{2\sqrt{x+12}}, & \text{dla } 1 \leq x < 4 \\ 0, & \text{dla } x \geq 4 \end{cases} ,$$

a ponadto posiada skoki, które zostały wyliczone powyżej, $P(X = -4) = \frac{1}{4}$, $P(X = -1) = \frac{\sqrt{13}-3}{4}$ oraz $P(X = 1) = \frac{1}{4}$.



Rysunek 12: Wykres gęstości $f_X(x)$ z Przykładu 6

9. Po pierwsze, należy wyznaczyć dziedzinę Y , czyli spostrzec, że $X \neq 0$ oraz $X \geq -5$. Prawdopodobieństwa takich zdarzeń wynoszą 0. Wystarczy zauważyć, że w przypadku, gdy $-5 \leq X < 0$, to $Y = \sqrt{X+5}$, a gdy $X > 0$, to $Y = \sqrt{X+12}$, więc:

$$\begin{aligned}\mathbb{E}Y &= 0 + \sqrt{-4+5} \cdot \frac{1}{4} + \int_{-4}^{-1} \sqrt{x+5} \cdot \frac{1}{4} \frac{1}{2\sqrt{x+5}} dx + \sqrt{-1+5} \cdot \frac{\sqrt{13}-3}{4} + \\ &+ 0 + \sqrt{1+12} \cdot \frac{1}{4} + \int_1^4 \sqrt{x+12} \cdot \frac{1}{4} \frac{1}{2\sqrt{x+12}} dx + 0 = \\ &= \frac{1}{4} + \int_{-4}^{-1} \frac{1}{8} dx + \frac{2\sqrt{13}-6}{4} + \frac{\sqrt{13}}{4} + \int_1^4 \frac{1}{8} dx = \frac{3\sqrt{13}-5}{4} + \frac{6}{8} = \frac{3\sqrt{13}-2}{4} .\end{aligned}$$

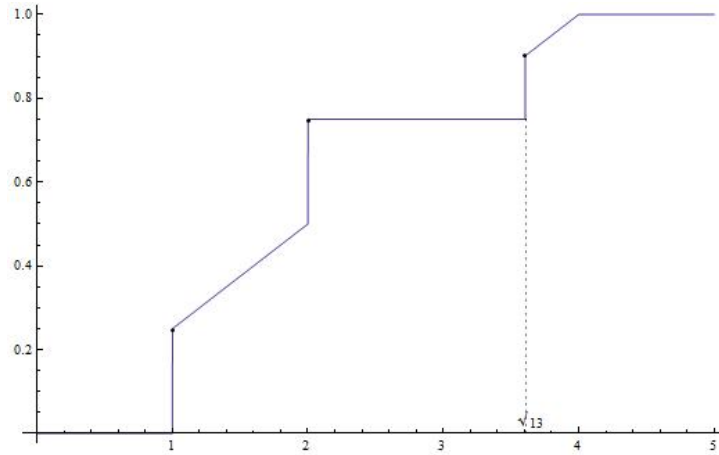
Aby wyliczyć dystrybuantę tej zmiennej, trzeba rozważyć dwa przypadki: gdy $X > 0$ i gdy $-5 \leq X < 0$. W pierwszym — $Y > \sqrt{12}$, a w drugim — $0 \leq Y < \sqrt{5}$. Niech najpierw $y > \sqrt{12}$. Wtedy

$$\begin{aligned}F_Y(y) &= P(Y \leq y) = 1 - P(Y > y) = 1 - P(\sqrt{X+12} > y) = 1 - P(X+12 > y^2) = \\ &= 1 - P(X > y^2 - 12) = P(X \leq y^2 - 12) = F_X(y^2 - 12) .\end{aligned}$$

Gdy natomiast $0 \leq y < \sqrt{5}$, to

$$F_Y(y) = P(Y \leq y) = P(\sqrt{X+5} \leq y) = P(X+5 \leq y^2) = P(X \leq y^2 - 5) = F_X(y^2 - 5) .$$

Stąd, po dłuższym rozważeniu możliwych przypadków, otrzymujemy:



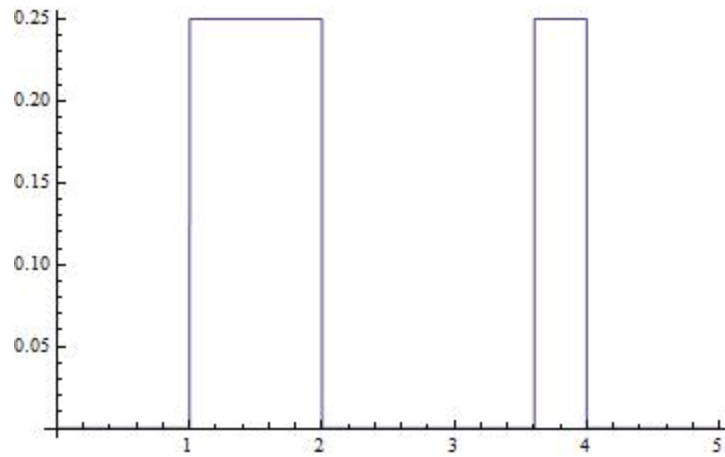
Rysunek 13: Wykres dystrybuanty $F_Y(y)$ z Przykładu 6

$$F_Y(y) = \begin{cases} 0, & \text{dla } y < 1 \\ \frac{1}{4}y, & \text{dla } 1 \leq y < 2 \\ \frac{3}{4}, & \text{dla } 2 \leq y < \sqrt{13} \\ \frac{1}{4}y, & \text{dla } \sqrt{13} \leq y < 4 \\ 1, & \text{dla } y \geq 4 \end{cases} ,$$

a zatem również funkcję gęstości:

$$f_Y(y) = \begin{cases} 0, & \text{dla } y < 1 \\ \frac{1}{4}, & \text{dla } 1 \leq y < 2 \\ 0, & \text{dla } 2 \leq y < \sqrt{13} \\ \frac{1}{4}, & \text{dla } \sqrt{13} \leq y < 4 \\ 0, & \text{dla } y \geq 4 \end{cases} ,$$

ze skokami (zostawiam jako ćwiczenie) $P(Y = 1) = P(Y = 2) = \frac{1}{4}$ oraz $P(Y = \sqrt{13}) = \frac{\sqrt{13}-3}{4}$.
 Porównać długości skoków dla rozkładu Y , z długościami skoków dla X . Dlaczego tak się dzieje?



Rysunek 14: Wykres gęstości $f_Y(y)$ z Przykładu 6