



Statystyczna analiza danych SAD-2020/21

Wykłady 3 i 4



Rozkład prawdopodobieństwa dyskretnej zmiennej losowej

Podstawowe pojęcia:

Zmienna losowa

- rozkład prawdopodobieństwa dyskretnej zmiennej losowej (skrót: d.z.l.)
- dystrybuanta d.z.l.
- parametry (charakterystyki) liczbowe d.z.l.
 - wartość oczekiwana (średnia) d.z.l.
 - wariancja i odchylenie standardowe d.z.l.



Rozkład prawdopodobieństwa ciągłej zmiennej losowej

Ciągła zmienna losowa (c.z.l.)

- gęstość prawdopodobieństwa i dystrybuanta c.z.l.
- parametry c.z.l. – wartość oczekiwana, wariancja, odchylenie standardowe

Własności wartości oczekiwanej i wariancji

Przykłady dyskretnych i ciągłych rozkładów prawdopodobieństwa

- dwupunktowy (Bernoulli'ego), dwumianowy, Poissona, jednostajny
- jednostajny, wykładniczy, normalny

Zmienne losowe

Zmienna losowa X

$$X : S \rightarrow (-\infty, \infty)$$

Przykłady.

■ rzut parą kostek sześciennych:

$$S = \{(i, j) : i, j \in \{1, 2, \dots, 6\}\}$$

$$X(s) : s = (i, j) \rightarrow i + j$$

■ rzut monetą: $S = \{0, 1\}$, gdzie 0 = orzeł, 1 = reszka

$$X : s \rightarrow X(s) = 1 - s \quad (= \text{liczba orłów})$$



Zmienne losowe

■ n - krotne powtórzenie doświadczenia Bernoulli'ego z prawdopodobieństwem sukcesu p , (sukces = 1, porażka = 0):

$$S = \{s = (x_1, x_2, \dots, x_n) : x_i \in \{0, 1\}\}$$

$$X : s = (x_1, x_2, \dots, x_n) \rightarrow X(s) = \sum_{i=1}^n x_i \text{ (liczba sukcesów).}$$

■ czas obsługi klienta, $S = \{x : 0 \leq x \leq T\}$

$$X : x \rightarrow X(x) = x.$$

Zmienne losowe

Definicja. Zmienną losową nazywamy funkcję rzeczywistą, określoną na przestrzeni zdarzeń elementarnych S , taką że dla dowolnego $x \in (-\infty, \infty)$ $\{s \in S : X(s) \leq x\}$ jest zdarzeniem.

- Zmienna losowa jest **dyskretna**, jeśli jej zbiór wartości jest przeliczalny (dyskretny): np. $\{0, 1, 2, \dots\}$, $\{0, 1, 2, 3\}$.
- Zmienna losowa jest **ciągła**, jeśli zakres (zbiór) jej wartości jest nieskończony i nieprzeliczalny („ciągły”), np. $(-\infty, \infty)$, $[0, \infty)$, $[-2, 2]$.

Zmienne losowe

Dyskretne zmienne losowe

Przykład. Niech zmienna losowa X będzie liczbą orłów w trzykrotnym rzucie monetą. Wówczas:

$$S = \{OOO, OOR, ORO, ROO, RRO, ROR, ORR, RRR\}$$

$$X = \quad 3 \quad \quad 2 \quad \quad 2 \quad \quad 2 \quad \quad 1 \quad \quad 1 \quad \quad 1 \quad \quad 0$$

- Zdarzenia elementarne są **jednakowo prawdopodobne**: moneta symetryczna i rzuty niezależne
- Możemy wyznaczyć prawdopodobieństwa tego, że zmienna losowa przyjmie wartości: 0, 1, 2, 3:

Dyskretne zmienne losowe

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

Notacja: $P(\{s \in S : X(s) = x\}) = P(X = x)$

Definicja.

- **Rozkładem prawdopodobieństwa** dyskretnej zmiennej losowej X nazywamy zbiór par uporządkowanych

$(x, P(X = x))$, gdzie x przebiega zakres wartości X

- **Funkcją prawdopodobieństwa (rozkładu)** dyskretnej zmiennej losowej X nazywamy funkcję:

$p(x) = P(X = x)$, gdzie x przebiega zakres wartości X .

Rozkład prawdopodobieństwa d.z.l.

Rozkład prawdopodobieństwa d.z.l. wygodnie jest przedstawić w postaci tabeli

x		x_1	x_2	...	x_n	...
$p(x)$		$p(x_1)$	$p(x_2)$		$p(x_n)$...

lub oznaczając $p(x_i) = p_i, i = 1, 2, \dots, n$, jako

x_i		x_1	x_2	...	x_n	...
p_i		p_1	p_2		p_n	...

Dyskretne zmienne losowe

Stwierdzenie. Niech $X : S \rightarrow \{x_1, x_2, \dots\}$. Wówczas

$$\sum_{i=1}^{\infty} p(x_i) = 1.$$

D. Z definicji funkcji prawdopodobieństwa i aksjomatów prawdopodobieństwa:

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P(X = x_i) = P\left(\bigcup_{i=1}^{\infty} \{s \in S : X(s) = x_i\}\right) = P(S) = 1$$

Dystrybuanta

Definicja.

Dystrybuantą zmiennej losowej X nazywamy funkcję:

$$F(x) = P(X \leq x), \quad x \in (-\infty, \infty).$$

- $F : (-\infty, \infty) \rightarrow [0, 1]$ (wartościami dystrybuanty są prawdopodobieństwami)
- Dla dyskretnej zmiennej losowej

$$F(x) = \sum_{i: x_i \leq x} p(x_i).$$

Wyznaczanie dystrybuanty

Przykład. Trzykrotny rzut monetą:

x	0	1	2	3
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$P(X \leq 0) = P(X = 0) = 1/8$$

$$P(X \leq 1) = P(X = 0) + P(X = 1) = 1/8 + 3/8 = 4/8$$

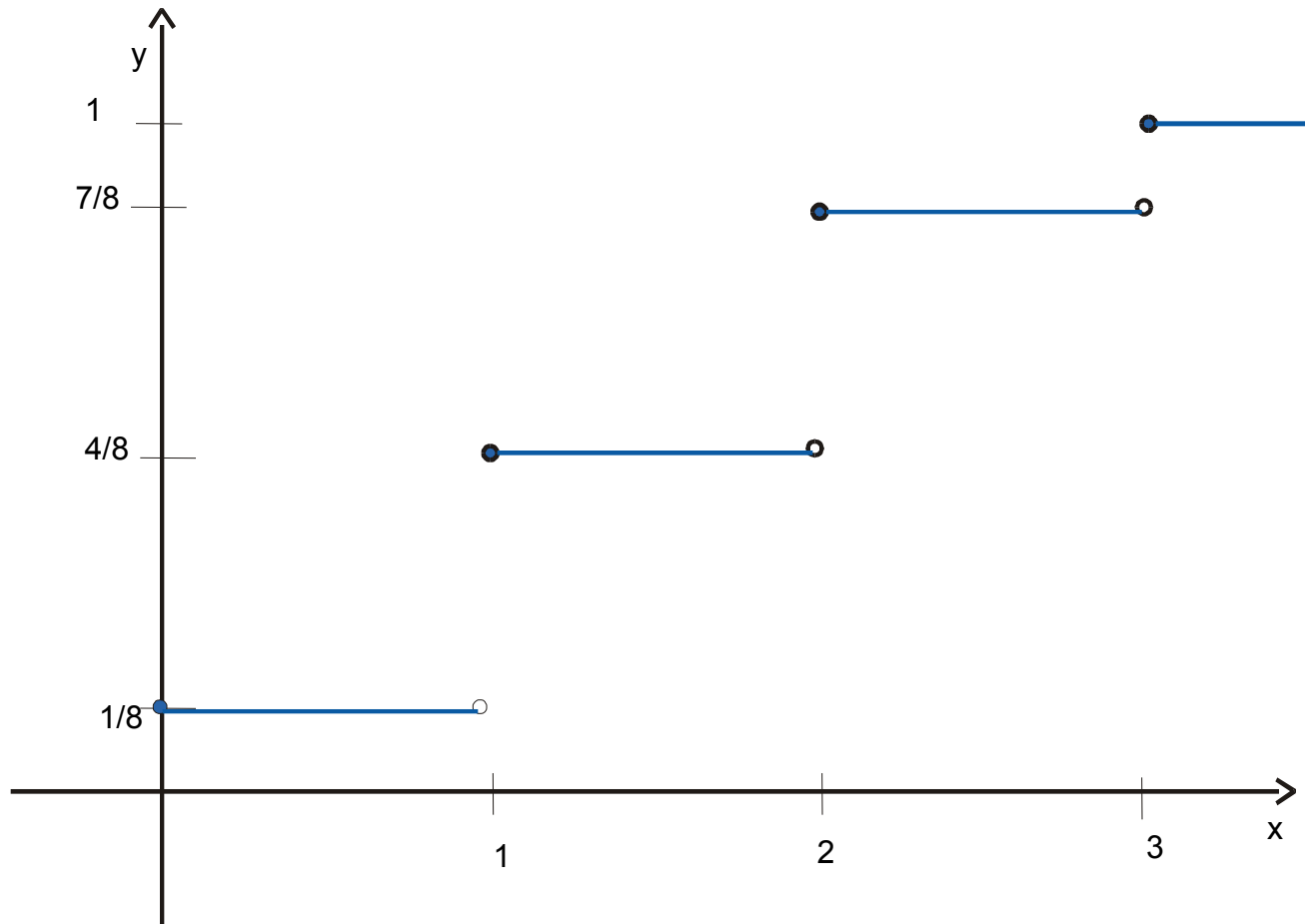
$$P(X \leq 2) = P(X \leq 1) + P(X = 2) = 4/8 + 3/8 = 7/8$$

Wyznaczanie dystrybuanty

- ◆ Dla $x < 0$ $F(x) = P(X \leq x) = P(\emptyset) = 0$
- ◆ Dla $0 \leq x < 1$ $F(x) = p(0) = 1/8$
- ◆ Dla $1 \leq x < 2$ $F(x) = p(0) + p(1) = 4/8$
- ◆ Dla $2 \leq x < 3$ $F(x) = p(0) + p(1) + p(2) = 7/8$
- ◆ Dla $x \geq 3$ $F(x) = p(0) + p(1) + p(2) + p(3) = 1.$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1/8 & 0 \leq x < 1 \\ 4/8 & \text{dla } 1 \leq x < 2 \\ 7/8 & 2 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

Wykres dystrybuanaty F





Wyznaczanie dystrybuanty d.z.l.

Dystrybuanta: $F(x) = P(X \leq x)$, $x \in (-\infty, \infty)$

Niech $x_1 < x_2 < x_3 < \dots < x_{k-1} < x_k$ będą wartościami zmiennej losowej X oraz $P(X = x_j) := p(x_j)$,

$$p(x_1) + p(x_2) + \dots + p(x_k) = 1.$$

$$x < x_1 \quad \Rightarrow \quad F(x) = P(X \leq x) = P(\emptyset) = 0$$

$$x_1 \leq x < x_2 \quad \Rightarrow \quad F(x) = p(x_1)$$

$$x_2 \leq x < x_3 \quad \Rightarrow \quad F(x) = p(x_1) + p(x_2) = F(x_1) + p(x_2)$$

.....

$$x_j \leq x < x_{j+1} \quad \Rightarrow \quad F(x) = p(x_1) + \dots + p(x_{j-1}) + p(x_j) = F(x_{j-1}) + p(x_j), \quad j = 2, \dots, k-1$$



Własności dystrybuanty

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

- ◆ $0 \leq F(x) \leq 1, \quad x \in (-\infty, \infty)$
- ◆ funkcja niemalejąca
- ◆ funkcja prawostronnie ciągła
- ◆ $F(x) - F(x^-) = P(X = x)$
- ◆ $\lim_{x \rightarrow -\infty} F(x) = 0$
- ◆ $\lim_{x \rightarrow \infty} F(x) = 1$

Prawdopodobieństwo a dystrybuanta

- $P(a < X \leq b) = F(b) - F(a)$
- $P(a \leq X \leq b) = F(b) - F(a) + p(a)$
- $P(a < X < b) = F(b) - F(a) - p(b)$
- $P(a \leq X < b) = F(b) - F(a) + p(a) - p(b)$

D.

$$(-\infty, b] = (-\infty, a] \cup (a, b]$$

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

$$F(b) = F(a) + P(a < X \leq b)$$

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

$$[a, b] = (a, b] \cup \{a\}$$

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) + P(X = a) = \\ &= F(b) - F(a) + p(a). \end{aligned}$$



Wartość oczekiwana (średnia)

Definicja.

Wartością średnią (oczekiwaną) dyskretnej zmiennej losowej X o funkcji prawdopodobieństwa $p(\cdot)$ nazywamy liczbę

$$\mu_X = \sum_{i=1}^{\infty} x_i p(x_i)$$

gdzie x_1, x_2, \dots oznaczają wszystkie wartości X .

Notacja: μ_X lub $E(X)$.

Obliczanie wartości oczekiwanej

Przykłady.

■ $f(x) = ax + b, \quad Y = f(X) = aX + b,$

$$\mu_{aX+b} = \sum_{i=1}^{\infty} (ax_i + b)p(x_i) = a\mu_X + b.$$

■ Wykonujemy niezależne rzuty monetą symetryczną aż do momentu wyrzucenia orła. Niech X oznacza liczbę wykonanych rzutów, $Y = 2^{X-1}$.

$$\sum_{i=1}^{\infty} 2^{i-1} P(X = i) = \sum_{i=1}^{\infty} 2^{i-1} \times \left(\frac{1}{2}\right)^i = \sum_{i=1}^{\infty} \frac{1}{2} = \infty.$$

Wartość średnia nie istnieje.

Obliczanie wartości oczekiwanej

■ Wygrana na loterii jest zmienną losową X o dystrybuancie:

$$F(x) = \begin{cases} 0 & x < 0, \\ 0,5 & 0 \leq x < 100, \\ 0,75 & 100 \leq x < 200, \\ 1 & x \geq 200. \end{cases} \text{ dla}$$

- ◆ $P(X = 0) = P(X \leq 0) - P(X < 0) = F(0) - F(0^-) = 0,5$
- ◆ $P(X = 100) = P(X \leq 100) - P(X < 100) = F(100) - F(100^-) = 0,75 - 0,5 = 0,25$

Wartość oczekiwana

$$\begin{aligned} \diamond P(X = 200) &= P(X \leq 200) - P(X < 200) = \\ &F(200) - F(200^-) = 1 - 0,75 = 0,25. \end{aligned}$$

$$\mu_X = 0 \times 0,5 + 100 \times 0,25 + 200 \times 0,25 = 75.$$

Twierdzenie.

$$\mu_{f(X)} = \sum_{i=1}^{\infty} f(x_i) p(x_i).$$

Wariancja

Definicja. Wariancją dyskretnej zmiennej losowej o funkcji prawdopodobieństwa $p(\cdot)$ nazywamy wielkość

$$\sigma_X^2 = \sum_{i=1}^{\infty} (x_i - \mu_X)^2 p(x_i).$$

Odchylenie standardowe: $\sigma_X = \sqrt{\sigma_X^2}$

Uwaga. $\sigma_X^2 = E(X - \mu_X)^2$

Interpretacja: wariancja - miara rozproszenia wartości zmiennej losowej względem wartości średniej.

Wariancja

Zadanie. Zmienne losowe X i Y mają rozkłady jednostajne na zbiorach punktów $\{-1, 0, 1\}$ oraz $\{-2, 0, 2\}$. Obliczyć wartości średnie i wariancje zmiennych X i Y .

$$\mu_X = -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0, \quad \mu_Y = 0.$$


$$\sigma_X^2 = (-1-0)^2 \times \frac{1}{3} + (0-0)^2 \times \frac{1}{3} + (1-0)^2 \times \frac{1}{3} = \frac{2}{3}$$


$$\sigma_Y^2 = (-2-0)^2 \times \frac{1}{3} + (0-0)^2 \times \frac{1}{3} + (2-0)^2 \times \frac{1}{3} = \frac{8}{3}$$

$$\sigma_Y^2 > \sigma_X^2$$

Własności wariancji i średniej

Twierdzenie.


$$\sigma_X^2 = \mu_{X^2} - (\mu_X)^2$$


$$\sigma_{aX+b}^2 = a^2 \sigma_X^2$$


$$\mu_{aX+b} = a\mu_X + b$$

■ Rozkład dwupunktowy

Zmienna losowa X ma rozkład dwupunktowy, jeśli

$$P(X = x_1) = p, \quad P(X = x_2) = q, \quad q = 1 - p, \quad 0 < p < 1.$$

Funkcja prawdopodobieństwa:

x	x_1	x_2
$p(x)$	p	q

Rozkład Bernoulli'ego

■ Rozkład zero – jedynkowy (rozkład Bernoulli'ego z prawdopodobieństwem sukcesu p)

$$P(X = 1) = p, \quad P(X = 0) = 1 - p = q$$

$$\mu_X = 0 \times (1 - p) + 1 \times p = p$$

$$\sigma_X^2 = 1^2 \times p + 0^2 \times (1 - p) - p^2 = p - p^2 = p \cdot q$$

Dyskretny rozkład jednostajny

- **Rozkład jednostajny na k punktach:** rozkład zmiennej losowej X o funkcji prawdopodobieństwa:

$$P(X = x_1) = P(X = x_2) = \dots = P(X = x_k) = 1/k.$$

$$\mu_X = \sum_{i=1}^k x_i \times \frac{1}{k} = \frac{1}{k} \sum_{i=1}^k x_i,$$

$$\sigma_X^2 = \sum_{i=1}^k (x_i - \mu_X)^2 \times \frac{1}{k} = \frac{1}{k} \sum_{i=1}^k (x_i - \mu_X)^2.$$

Przykład: X = liczba oczek w rzucie kostką sześcienną.

Rozkład dwumianowy

Rozkład dwumianowy

Wykonujemy n niezależnych jednakowych doświadczeń Bernoulli'ego z prawdopodobieństwem sukcesu p (w każdym doświadczeniu możliwy sukces z prawdopodobieństwem p lub porażka z prawdopodobieństwem $1 - p$). **Funkcja prawdopodobieństwa zmiennej losowej X będącej liczbą sukcesów:**

$$P(X = k) = b(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n.$$

$$\mu_X = np,$$

$$\sigma_X^2 = np(1 - p).$$



Rozkład dwumianowy

Uzasadnienie:

$$S = \{s = (x_1, \dots, x_n) : x_i \in \{0, 1\}\},$$

$$P(\{s\}) = p^{\sum_{i=1}^n x_i} \times (1-p)^{(n - \sum_{i=1}^n x_i)},$$

$$P(X = k) = P(\{s \in S : \sum_{i=1}^n x_i = k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Notacja: $X \sim \text{Bin}(n, p)$.

Przykłady: liczba elementów wadliwych spośród n wylosowanych z dużej partii towaru o wadliwości p ,
liczba trafień do celu na zawodach sportowych w n próbach

Rozkład dwumianowy

Przykład. Urządzenie składa się z 14 identycznych pracujących niezależnie podzespołów. Ulegnie ono awarii, jeśli co najmniej 3 podzespoły będą niesprawne. Prawdopodobieństwo awarii podzespołu wynosi 0,1. Znaleźć prawdopodobieństwo awarii urządzenia.

$$X \sim \text{Bin}(14, 0.1), \quad P(X \geq 3) = 1 - P(X < 3).$$

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \\ b(0; 14, 0.1) + b(1; 14, 0.1) + b(2; 14, 0.1) =$$

$$0,229 + 0,356 + 0,257 = 0,842$$

$$P(X \geq 3) = 1 - 0,842 = 0,178$$

Rozkład Poissona

■ Rozkład Poissona

Definicja. Zmienna losowa X ma **rozkład Poissona** z parametrem $\lambda, \lambda > 0$, jeśli

$$P(X = k) = p(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Notacja: $X \sim P(\lambda)$

Przykłady: liczba klientów w systemie masowej obsługi,
liczba cząstek emitowanych przez substancję radioaktywną,
liczba awarii sieci informatycznej w określonym przedziale czasu,

Rozkład Poissona

Twierdzenie. $\boxed{\mu_X = \lambda}, \quad \boxed{\sigma_X^2 = \lambda}.$

Własności rozkładu Poissona;

■ Niech $n \rightarrow \infty, p = p_n \rightarrow 0, np = \lambda > 0.$

Wówczas dla ustalonego k , przy $n \rightarrow \infty$

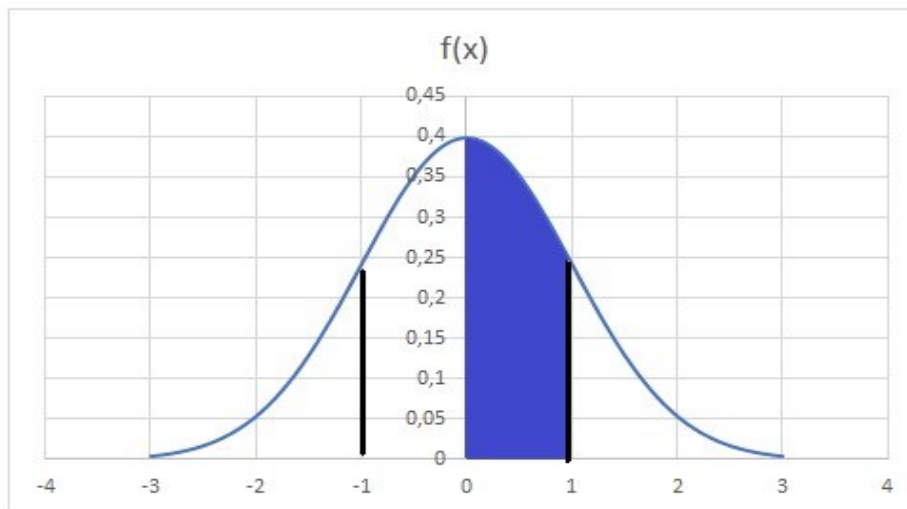
$$b(k; n, p) \rightarrow p(k, \lambda).$$

Ciągłe zmienne losowe

Definicja. Zmienną losową X nazywamy **ciągłą** zmienną losową, jeśli istnieje nieujemna funkcja f , zwana **gęstością**, taka że dla dowolnych a, b , $-\infty \leq a \leq b \leq \infty$,

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Gęstość prawdopodobieństwa c.z.l.



$$P(0 < X < 1) = \int_0^1 f(x) dx$$

Dystrybuanta c.z.l.

$$P(a \leq X \leq b) = \int_a^b f(t)dt$$

- Przyjmując $a = -\infty$, $b = x$ otrzymujemy

$$P(-\infty \leq X \leq x) = P(X \leq x) = \int_{-\infty}^x f(t)dt,$$

czyli dystrybuantę znajdujemy na podstawie gęstości

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

Dystrybuanta i gęstość c.z.l.

- Przyjmując $a = -\infty$, $b = \infty$ otrzymujemy

$$P(-\infty \leq X \leq \infty) = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(t) dt = 1$$

- Przyjmując $a = b = c$, c – dowolna stała, otrzymujemy

$$P(X = c) = \int_c^c f(t) dt = 0$$

- Dowolność $a \leq b \implies f(x) \geq 0, x \in (-\infty, \infty)$

Definicja. Funkcję

$$F(x) := P(X \leq x), x \in (-\infty, \infty)$$

nazywamy dystrybuantą zmiennej losowej X .

- $F(x) = \int_{-\infty}^x f(t)dt, \quad x \in (-\infty, \infty)$
- $P(-\infty \leq X \leq \infty) = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(t)dt = 1$
- $f(x) \geq 0$, dla każdego x
- $P(X = c) = 0$, dla każdej stałej c

Dystrybuanta

Stwierdzenie. Dla ciągłej zmiennej losowej o dystrybuancie F zachodzi

$$\begin{aligned} P(a < X < b) &= P(a < X \leq b) = P(a \leq X < b) = \\ &= P(a \leq X \leq b) = F(b) - F(a). \end{aligned}$$

D. $P(X = b) = P(X = a) = 0$. Zatem dołączenie lub usunięcie brzegu przedziału nie wpływa na wartość prawdopodobieństwa, np.

$$[a, b] = (a, b) \cup \{a\} \cup \{b\},$$

$$P(a \leq X \leq b) = P(a < X < b) + P(X = a) + P(X = b),$$

Dystrybuanta a gęstość

$$F(b) - F(a) = P(a < X < b).$$

Twierdzenie. Jeśli gęstość zmiennej losowej X jest funkcją ciągłą, to dla każdego x zachodzi

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

D.
$$F'(x) = \frac{d}{dx} \int_{-\infty}^x f(s) ds = f(x).$$



Gęstość prawdopodobieństwa

Definicja. Funkcja $f(x), x \in (-\infty, \infty)$, spełniająca warunki:

■ $f(x) \geq 0, x \in (-\infty, \infty),$

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

nazywana jest gęstością.



Gęstość prawdopodobieństwa

Definicja. Funkcja $f(x), x \in (-\infty, \infty)$, spełniająca warunki:

■ $f(x) \geq 0, x \in (-\infty, \infty),$

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

nazywana jest gęstością.

$$P(X \in [a, b]) = \int_a^b f(x) dx$$

Wskaźniki położenia i rozproszenia dla ciągłych zmiennych losowych

Definicja. Wartością średnią (oczekiwaną) ciągłej zmiennej losowej X mającej gęstość f nazywamy liczbę

$$\mu_X = \int_{-\infty}^{\infty} xf(x)dx$$



Charakterystyki liczbowe c.z.l.

Wartością oczekiwaną zmiennej losowej $g(X)$ nazywamy liczbę

$$\mu_{g(X)} = E(g(X)) := \int_{-\infty}^{\infty} g(x)f(x)dx$$

Np.

$$\mu_{X^2} = E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$

Charakterystyki liczbowe c.z.l.

Definicja. **Wariancją** ciągłej zmiennej losowej X o gęstości f nazywamy liczbę

$$\sigma_X^2 = \int_{-\infty}^{\infty} (s - \mu_X)^2 f(s) ds$$

Odchylenie standardowe:

$$\sigma_X = \sqrt{\sigma_X^2}.$$

Uwaga. Z definicji wariancji oraz wartości oczekiwanej funkcji zmiennej losowej

$$\sigma_X^2 = E(X - \mu_X)^2$$

Własności wartości średniej i wariancji

Twierdzenie. Jeśli ciągła zmienna losowa ma wariancję, to dla dowolnych liczb a, b zachodzą wzory

$$\mu_{aX+b} = a\mu_X + b$$

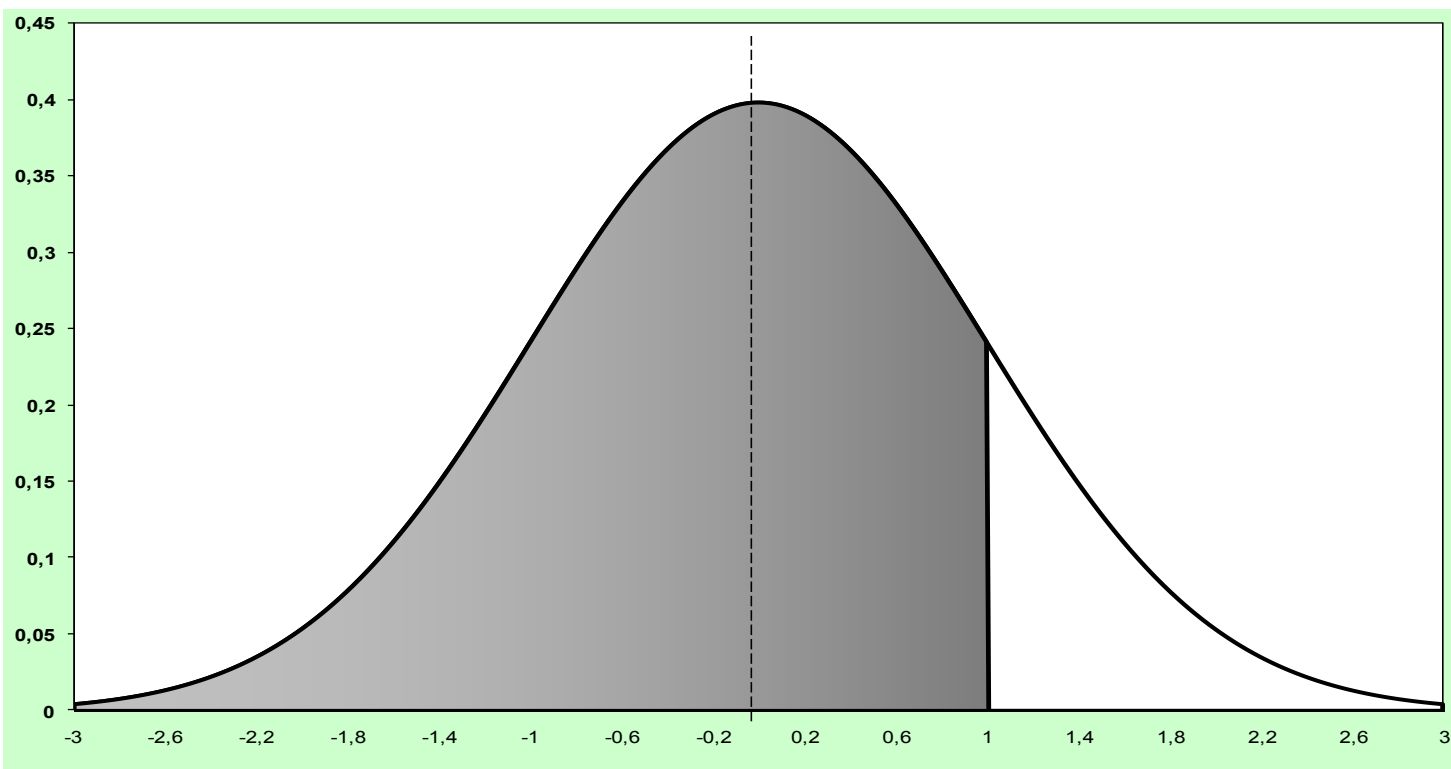
$$\sigma_{aX+b}^2 = a^2 \sigma_X^2$$

$$\sigma_X^2 = \mu_{X^2} - (\mu_X)^2.$$

Powyższe wzory wynikają z własności (liniowości) całki.

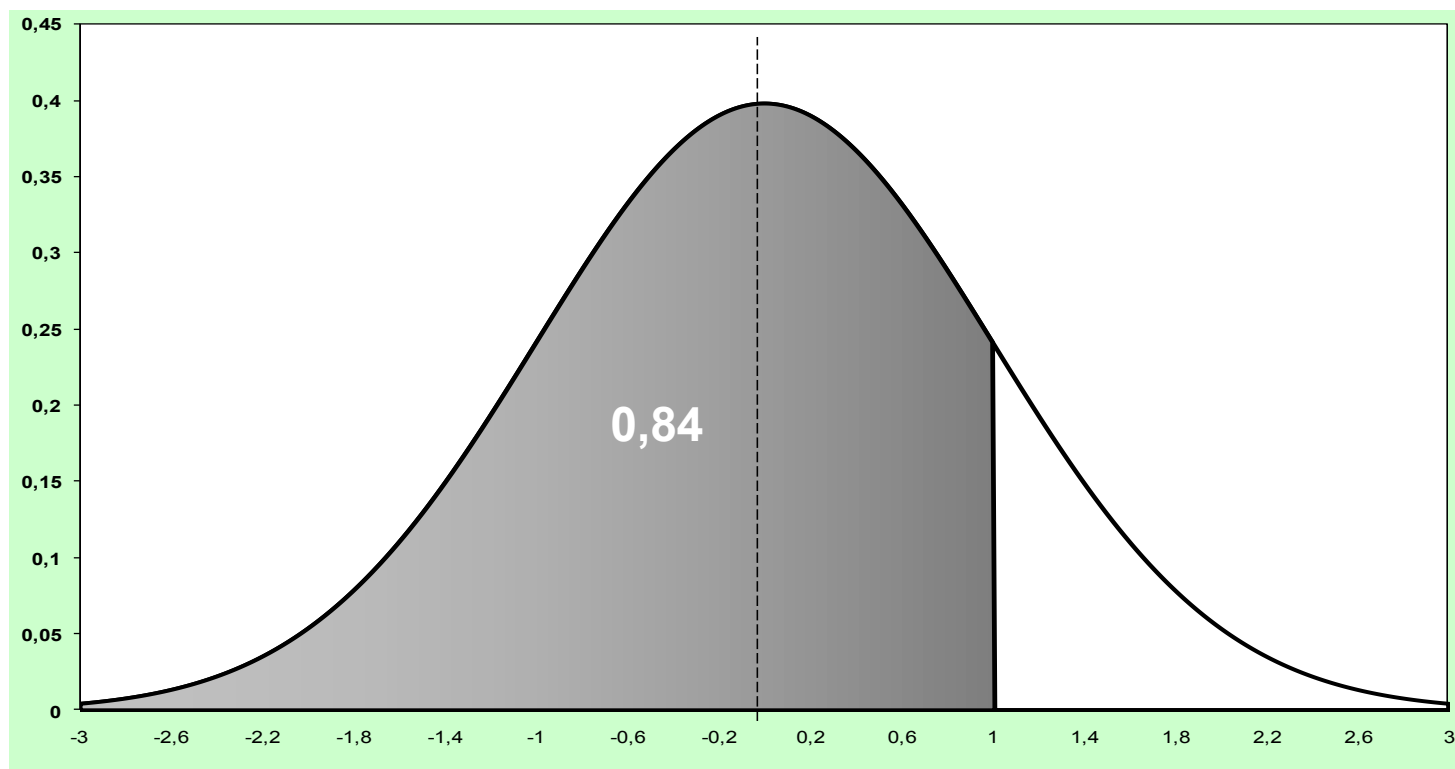
Notacja: $\mu_X = E(X)$, $\sigma_X^2 = Var(X)$

Kwantyle zmiennej losowej



Definicja. Niech $0 < p < 1$.

Kwantylem rzędu p nazywamy punkt q_p na osi poziomej, taki że pole pod gęstością na lewo od niego wynosi p



Pole zacięniowane = 0,84. Zatem kwantyl rzędu 0,84 = 1.



Kwartyle zmiennej losowej

- Dolny (pierwszy kwartył) = kwantyl rzędu 0,25:
 $P(X \leq q_{0,25}) = 0,25$
- Górny kwartył (trzeci kwartył) = kwantyl rzędu 0,75: $P(X \leq q_{0,75}) = 0,75$
- Mediana (drugi kwartył) = kwantyl rzędu 0,5:
 $P(X \leq q_{0,5}) = 0,5$

Mediana

Liczba $q_{0,5}$, taka że pole pod wykresem gęstości na lewo od mediany wynosi 0,5. Zatem

$$\int_{-\infty}^{q_{0,5}} f(x)dx = 0,5 = \int_{q_{0,5}}^{\infty} f(x)dx.$$

Stwierdzenie. (standaryzacja)

Jeśli zmienna losowa X ma wartość średnią μ_X oraz wariancję σ_X^2 , to standaryzowana zmienna losowa

$$Z := \frac{X - \mu_X}{\sigma_X}$$

ma wartość średnią 0 i wariancję 1.

Standaryzacja zmiennej losowej

D.
$$\mu_Z = E\left(\frac{X - \mu_X}{\sigma_X}\right) = E\left(\frac{1}{\sigma_X} \cdot X - \frac{\mu_X}{\sigma_X}\right) =$$

$$= \frac{1}{\sigma_X} E(X) - \frac{\mu_X}{\sigma_X} = 0$$

$$\sigma_Z^2 = E\left(\frac{X - \mu_X}{\sigma_X}\right)^2 = \left(\frac{1}{\sigma_X}\right)^2 \times E(X - \mu_X)^2 = 1.$$



Ciągłe zmienne losowe - przykłady

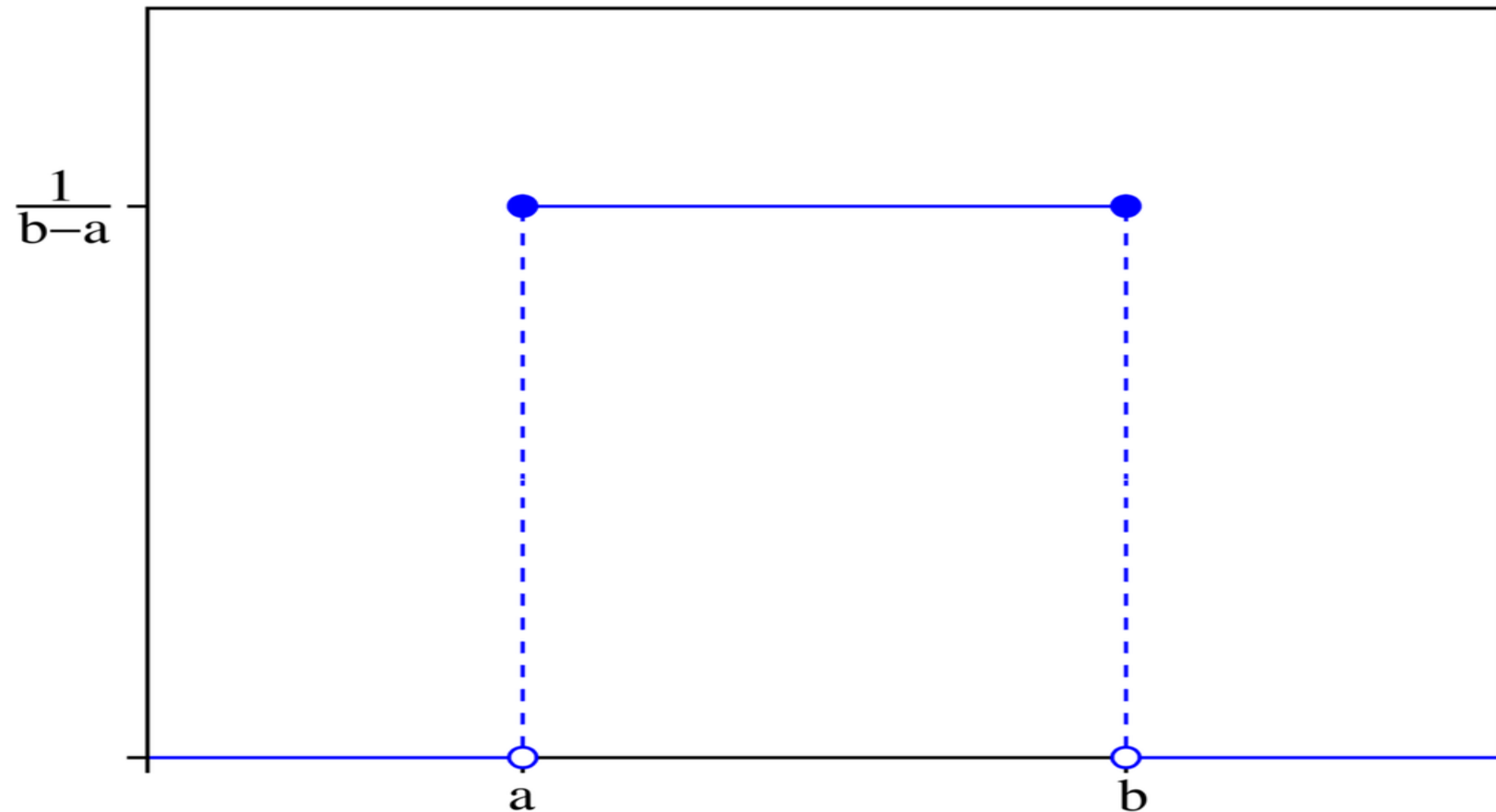
- Zmienna losowa X ma rozkład jednostajny na przedziale $[a,b]$, jeśli ma gęstość:

$$f(x) = \begin{cases} 1/(b-a) & \text{dla } x \in [a, b] \\ 0 & \text{dla } x \notin [a, b] \end{cases}$$

$$\mu_X = \frac{a+b}{2}, \quad \sigma_X^2 = \frac{(b-a)^2}{12}$$

Notacja: $X \sim U(a, b)$.

Wykres gęstości rozkładu $U(a,b)$



Parametry rozkładu jednostajnego

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{\infty} x \frac{1}{b-a} dx = \int_{-\infty}^a 0 dx + \frac{1}{b-a} \int_a^b x dx + \int_b^{\infty} 0 dx = \\
 &= 0 + \frac{1}{b-a} \left| \frac{x^2}{2} \right|_a^b + 0 = \frac{1}{b-a} \left(\frac{b^2}{2} - \frac{a^2}{2} \right) = \\
 &= \frac{1}{2} \frac{(b-a)(b+a)}{b-a} = \frac{a+b}{2}
 \end{aligned}$$

Parametry rozkładu jednostajnego

$$\begin{aligned}
 \text{Var}(X) &= E(X^2) - (E(X))^2 = \int_a^b x^2 \frac{1}{b-a} dx - \frac{(a+b)^2}{4} = \\
 &= \frac{1}{b-a} \left| \frac{x^3}{3} \right|_a^b - \frac{(a+b)^2}{4} = \frac{1}{b-a} \left(\frac{b^3}{3} - \frac{a^3}{3} \right) - \frac{(a+b)^2}{4} = \\
 &= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}
 \end{aligned}$$

Dystrybuanta rozkładu $U(a,b)$

1) Niech $x < a$.

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x 0 dt = 0$$

2) Niech $a \leq x \leq b$.

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt = \int_{-\infty}^a 0 dt + \int_a^x \frac{1}{b-a} dt = \\ &= \frac{1}{b-a} [t]_a^x = \frac{1}{b-a} (x - a) = \frac{1}{b-a} \cdot x - \frac{a}{b-a} \end{aligned}$$

3) Niech $x > b$.

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^a 0 dt + \int_a^b \frac{1}{b-a} dt + \int_b^x 0 dt = 1$$

Dystrybuanta rozkładu $U(a,b)$

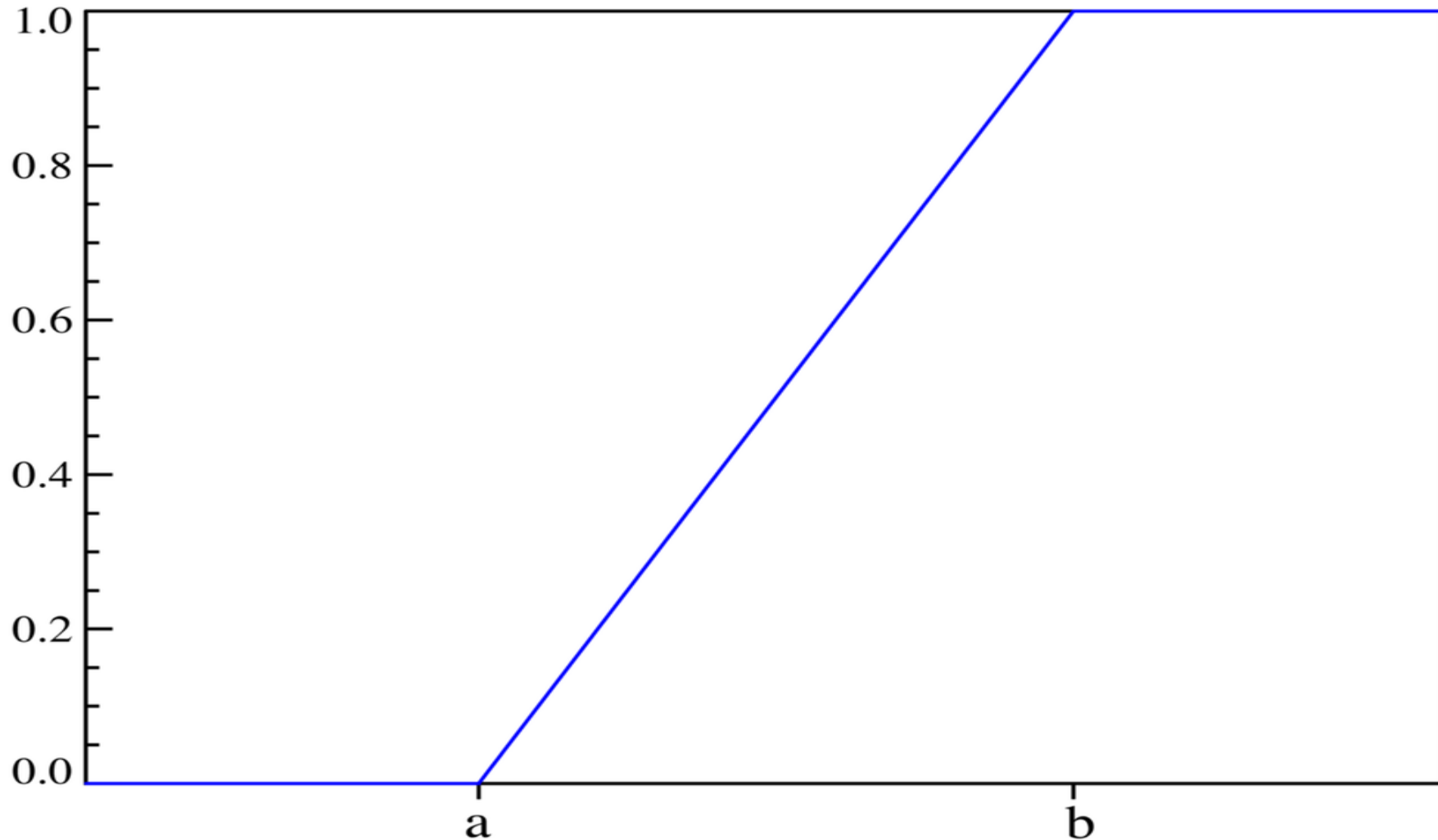
Ostatecznie możemy zapisać dystrybuantę zmiennej losowej o rozkładzie jednostajnym na przedziale $[a,b]$ w postaci

$$F(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a} \cdot x - \frac{a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Jej wykres jest przedstawiony na kolejnym slajdzie.



Dystrybuanta rozkładu jednostajnego



■ Zmienna losowa o rozkładzie wykładniczym

Niech X_t ma rozkład Poissona $P(\lambda t)$ (liczba zdarzeń w przedziale czasu $[0, t]$). Wówczas **czas oczekiwania** na zdarzenie jest zmienną losową T , taką że

$$P(T > t) = P(X_t = 0) = \frac{e^{-\lambda t} (\lambda t)^0}{0!} = e^{-\lambda t}, \text{ dla } t \geq 0.$$

Zmienna losowa T ma **dystrybuantę**

$$F(t) = 1 - P(T > t) = \begin{cases} 0 & \text{dla } t < 0 \\ 1 - e^{-\lambda t} & \text{dla } t \geq 0 \end{cases}$$

Ciągłe zmienne losowe

Zmienna losowa ma rozkład wykładniczy z parametrem λ ,
 $\lambda > 0$, jeśli ma gęstość $f(t) = F'(t)$:

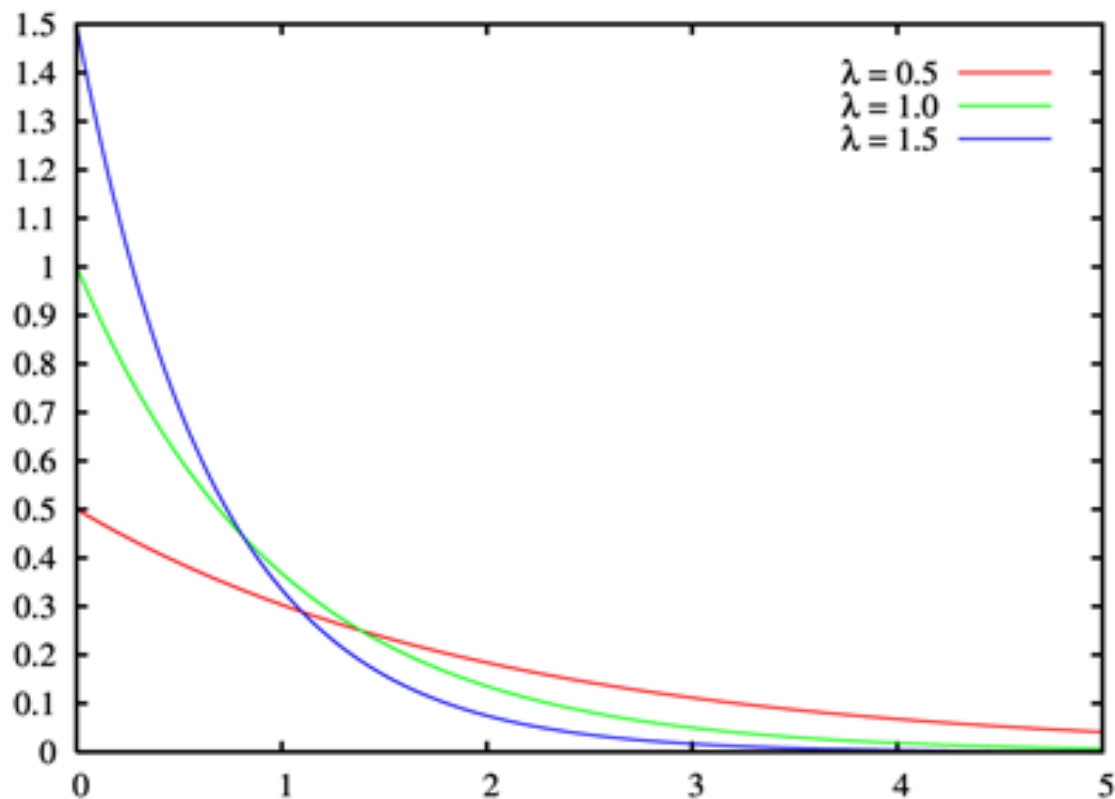
$$f(t) = \begin{cases} 0 & \text{dla } t < 0 \\ \lambda e^{-\lambda t} & \text{dla } t \geq 0 \end{cases} .$$

$$\mu_T = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

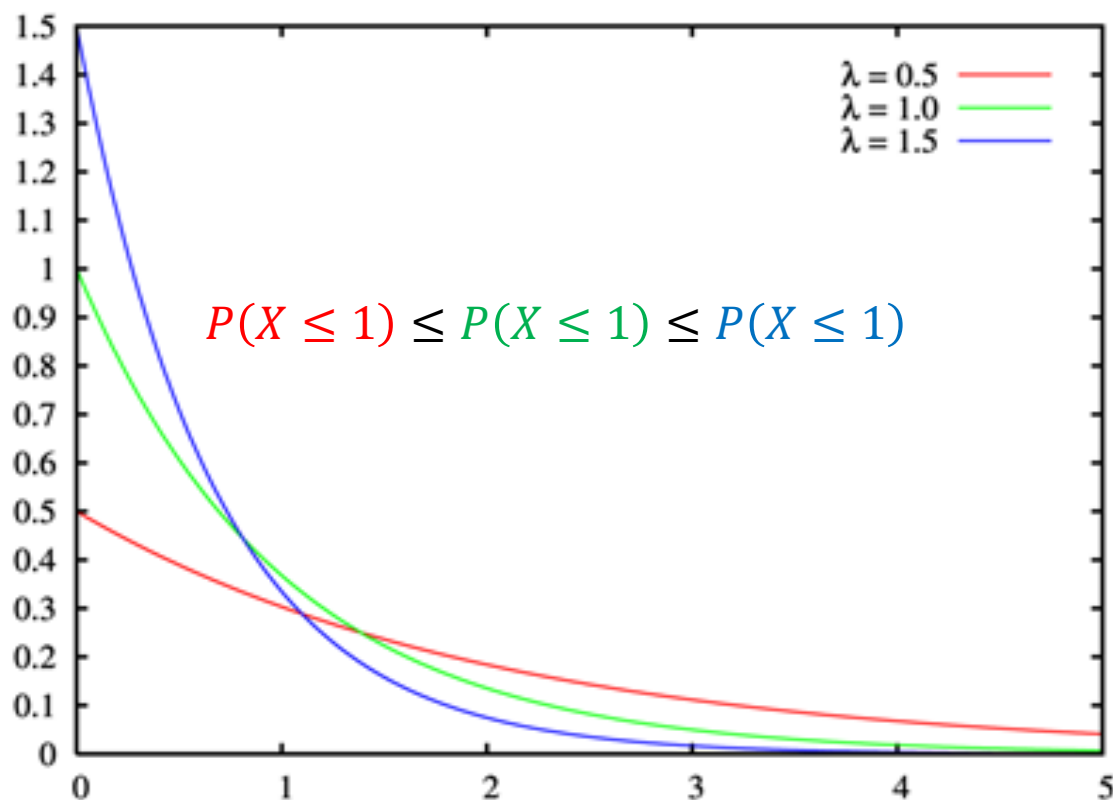
$$\sigma_T^2 = \mu_{T^2} - (\mu_T)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} .$$

Notacja: $X \sim \text{exp}(\lambda)$, zastosowanie – teoria niezawodności, masowej obsługi (czas życia elementu, czas obsługi, oczekiwania na obsługę, etc)

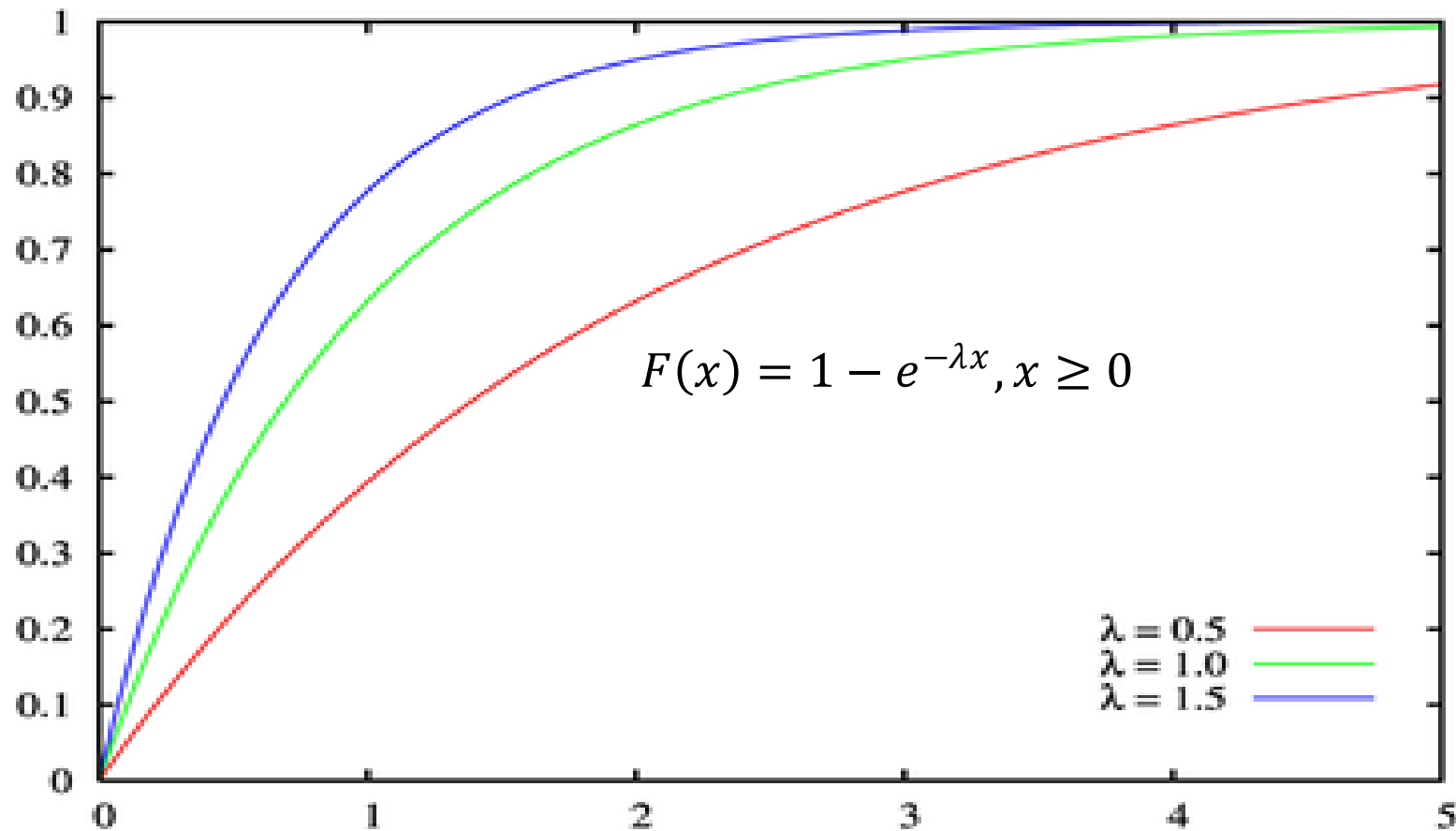
Wykresy gęstości rozkładu wykładniczego



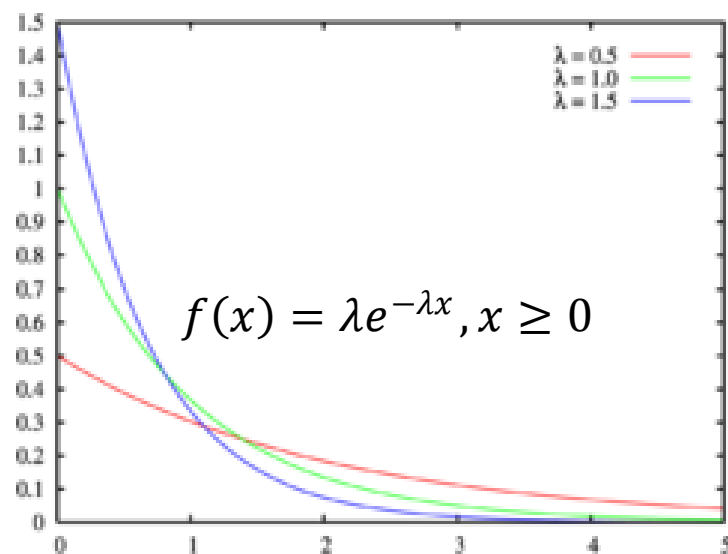
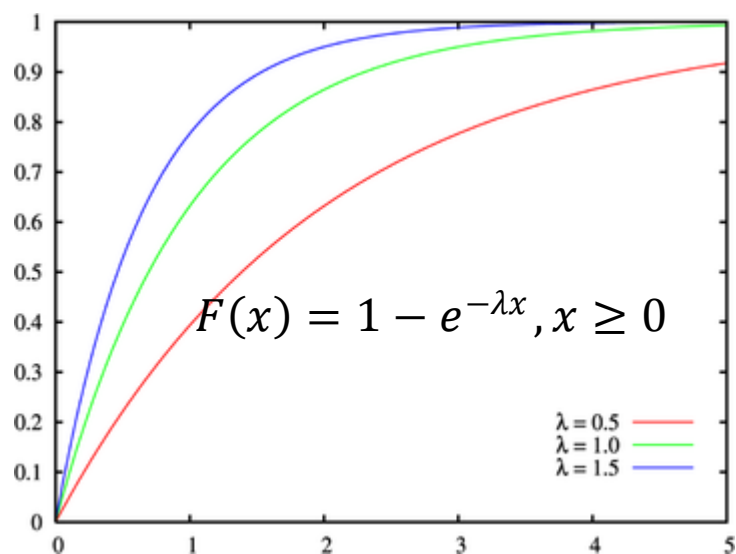
Wykresy gęstości rozkładu wykładniczego



Wykresy dystrybuant r. $\text{Exp}(\lambda)$



Dystrybuanta i gęstość r. wykł.



Dystrybuanta rozkładu wykładniczego

Przykład. Czas rozmowy telefonicznej jest zmienną losową o rozkładzie wykładniczym z parametrem $\lambda = 1/7$. Jakie jest prawdopodobieństwo, że telefonująca osoba będzie rozmawiała nie krócej niż 5 i nie dłużej niż 10 minut.

$$\begin{aligned} P(5 < T < 10) &= F(10) - F(5) == \left(1 - e^{-\frac{1}{7} \cdot 10}\right) - \left(1 - e^{-\frac{1}{7} \cdot 5}\right) \\ &= e^{-\frac{5}{7}} - e^{-\frac{10}{7}} \cong 0,4895 - 0,2965 \cong 0,25 \end{aligned}$$

Rozkład wykładniczy – brak pamięci

Przykład. Czas oczekiwania na połączenie z pewną siecią teleinformatyczną jest zmienną losową X o wartości średniej 0,5 (minut) i mającą rozkład wykładniczy. Znaleźć:

- (a) **medianę i dolny kwartył** czasu oczekiwania X .
(b) **prawdopodobieństwo**, że czas oczekiwania będzie dłuższy niż 3 minuty, jeśli wiadomo, że po 1 minucie jeszcze nie otrzymano połączenia.

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad E(X) = 1 / \lambda = 0,5.$$

$$\text{Stąd } \lambda = 2. \quad F(q_{0,5}) = 0,5, \quad F(q_{0,25}) = 0,25.$$

Rozkład wykładniczy

- $-2q_{0,5} = \ln(0,5) \equiv$

$$q_{0,5} = -\ln(0,5)/2 = \ln 2/2$$

- $1 - e^{-2 \times q_{0,25}} = 0,25, \quad e^{-2 \times q_{0,25}} = 0,75$

$$q_{0,25} = -\ln(0,75)/2 = [\ln(4/3)]/2.$$

Rozkład wykładniczy

$$(b) P(X > 3 | X \geq 1) = ?$$

$$P(X > t + h | X \geq t) = \frac{P(\{X > t + h\} \cap \{X \geq t\})}{P(X \geq t)} =$$

$$\frac{P(X > t + h)}{P(X \geq t)} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h}, \text{ dla } h > 0.$$

W zadaniu: $\lambda = 2$, $t + h = 3$, $t = 1$. Stąd $h = 2$.

$$P(X > 3 | X \geq 1) = e^{-4}$$



Rozkład wykładniczy – brak pamięci

Uwaga: Własność **braku pamięci rozkładu wykładniczego:**

$$\boxed{P(X > t + h | X \geq t)} = \underline{e^{-\lambda h}}, \text{ dla } h > 0$$

(c) W jakim zakresie czasu znajduje się 10% najdłużej trwających oczekiwań na połączenie z siecią ?

$$t \geq q_{0,9} = 0,5 \times \ln(10).$$

$$P(X \leq q_{0,9}) = 0,9 \implies P(X > q_{0,9}) = P(X \geq q_{0,9}) = 0,1$$

Ciągłe zmienne losowe

- Zmienna losowa X ma rozkład **normalny** z parametrami μ , $\sigma > 0$, jeśli ma gęstość

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2 / 2\sigma^2},$$

$$-\infty < x < \infty,$$

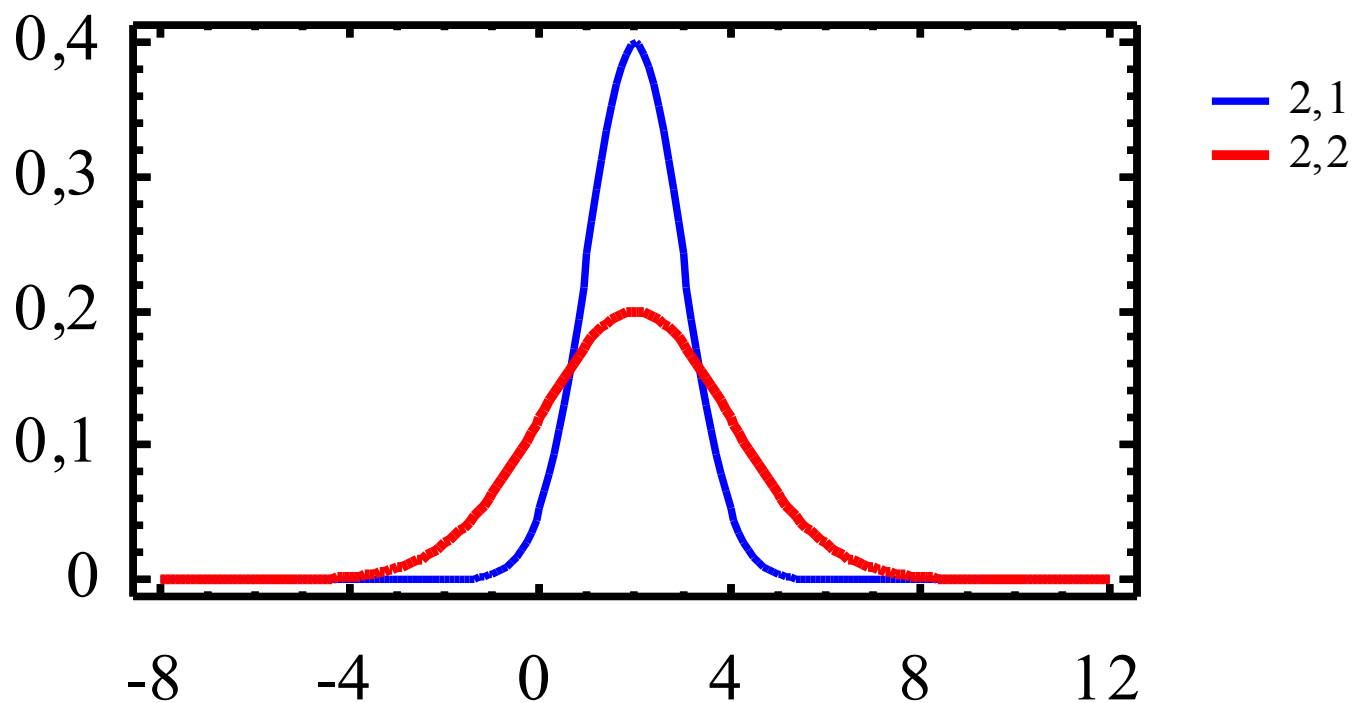
$$\mu_X = \mu, \quad \sigma_X = \sigma$$

Notacja: $X \sim N(\mu, \sigma)$.



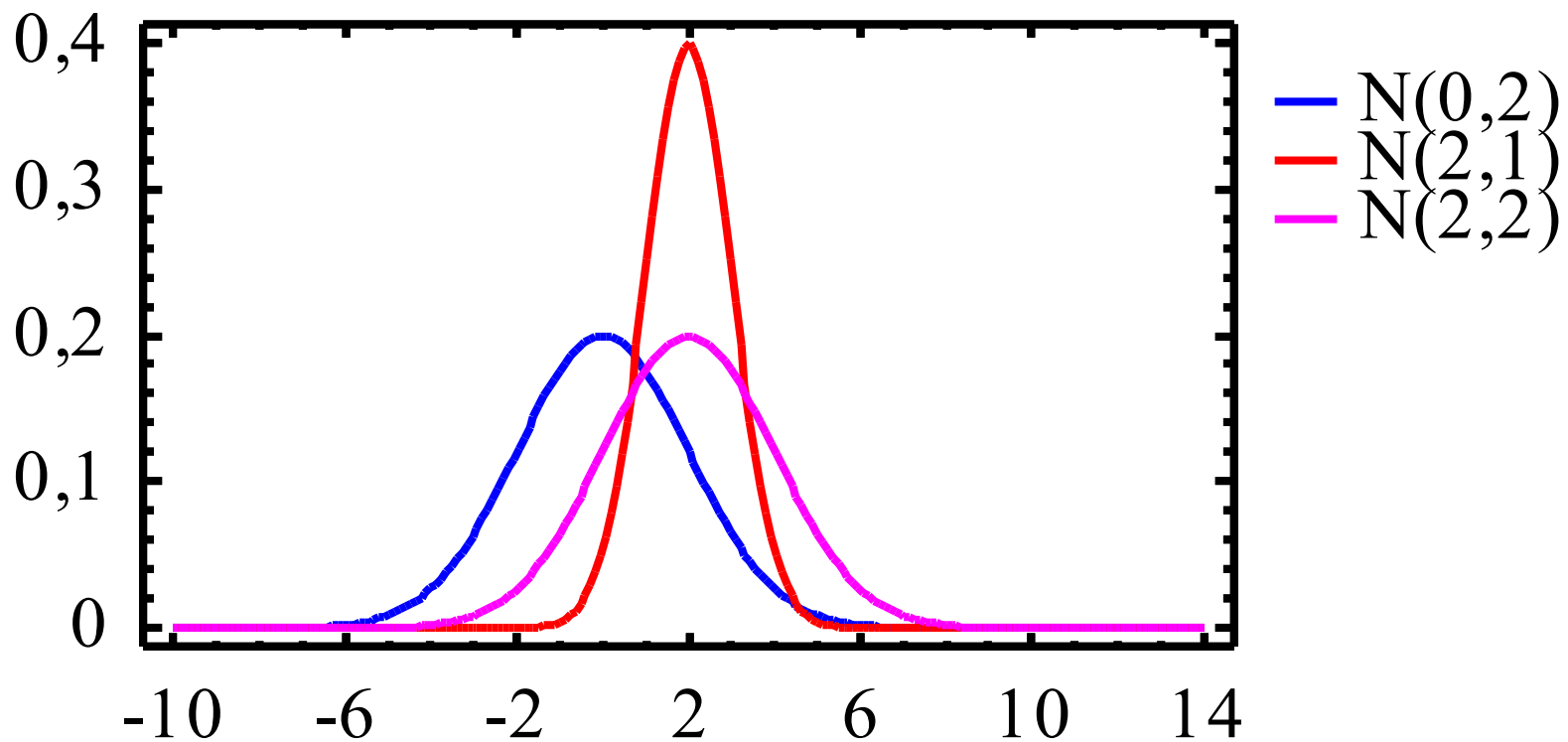
Wykresy gęstości normalnych

$F(2) = F(2) = 0,5$, bo wykres gęstości symetryczny względem prostej $x=2$, a pole pod wykresem gęstości po całej prostej wynosi 1.

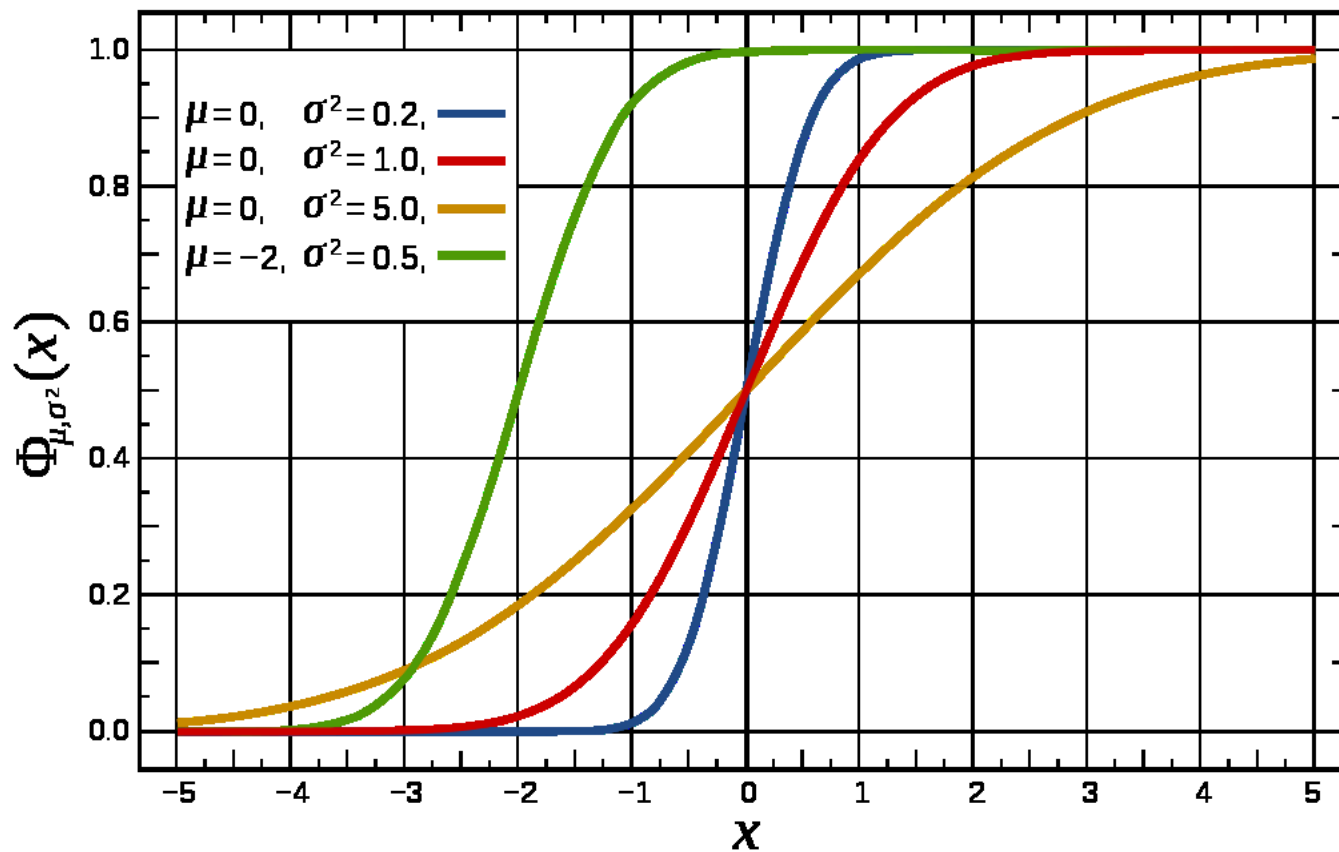




Wykresy gęstości normalnych



Wykresy dystrybuant rozkładu normalnego





Własności zmiennej losowej o rozkładzie normalnym

Twierdzenie. Niech

$$X \sim N(\mu, \sigma), \quad Z = \frac{X - \mu}{\sigma}.$$

Wówczas

- ♦ $Z \sim N(0,1)$
- ♦ $\mu_X = \mu, \quad \sigma_X^2 = \sigma^2$

Wyznaczanie dystrybuanty r. $N(\mu, \sigma)$

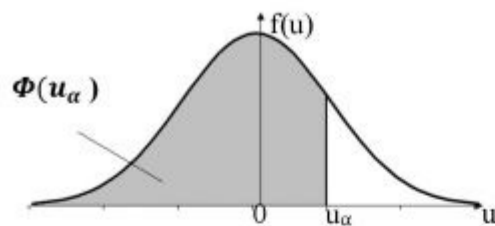
Wniosek. Niech $X \sim N(\mu, \sigma)$, niech $a < b$. Niech $Z \sim N(0,1)$, $\Phi(z) = P(Z \leq z)$, $z \in (-\infty, \infty)$ oznacza dystrybuantę standardowego rozkładu normalnego. Dystrybuantę zmiennej losowej X znajdujemy przy pomocy dystrybuanty zmiennej losowej Z .

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Stąd

$$P(a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

Tablica 1. Dystrybuanta standardowego rozkładu normalnego



$$\Phi(u) = P(U \leq u) \text{ dla } u \geq 0$$

u	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995

Rozkład Poissona a rozkład normalny

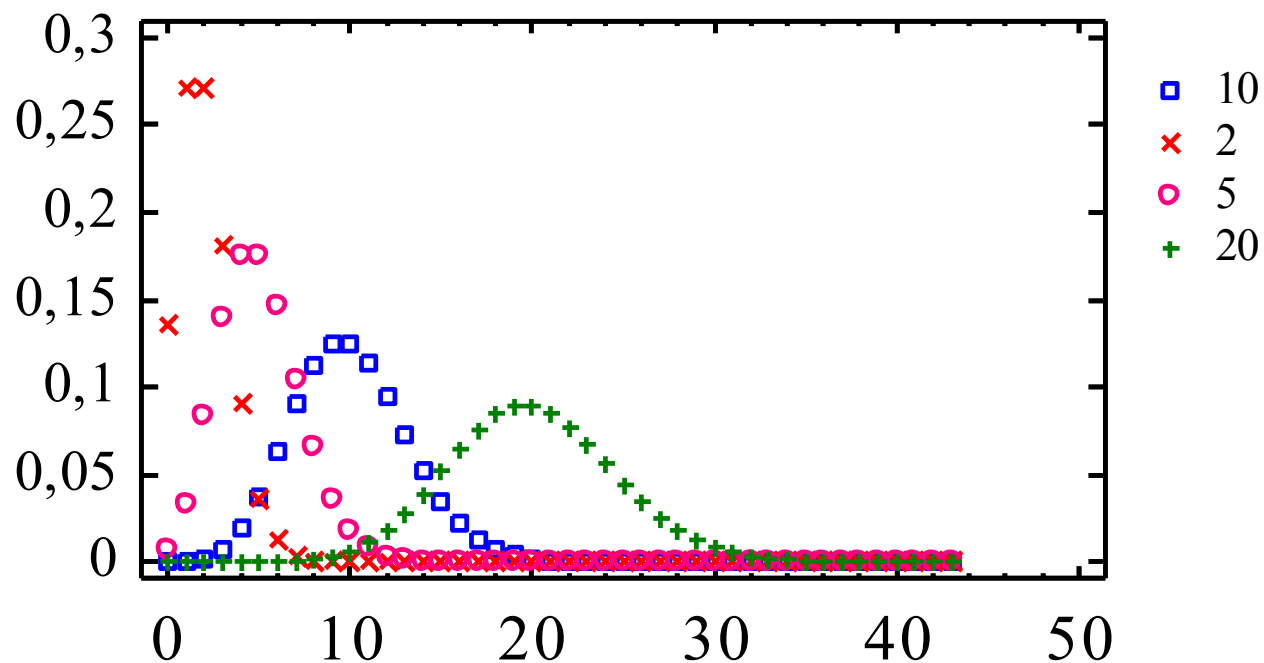
■ Jeśli $X \sim P(\lambda)$ dla dużego λ , to rozkład standaryzowanej zmiennej $(X - \lambda) / \sqrt{\lambda}$ jest w przybliżeniu normalny, tzn. $P((X - \lambda) / \sqrt{\lambda} \leq z) \approx \Phi(z)$,

dla dowolnego z . Zatem **dystrybuanta zmiennej losowej X jest bliska dystrybuancie** zmiennej losowej o rozkładzie

$$N(\lambda, \sqrt{\lambda}),$$

a funkcje prawdopodobieństwa są bliskie wartościom funkcji gęstości rozkładu normalnego $N(\lambda, \sqrt{\lambda})$, co ilustruje rysunek:

Rozkład Poissona



Rozkład Poissona a rozkład normalny

Przykład. Liczba awarii sprzętu komputerowego supermarketu w ciągu miesiąca jest zmienną losową X o rozkładzie Poissona o średniej 36. Jakie jest prawdopodobieństwo, że w ciągu miesiąca będzie co najwyżej 30 awarii ?

$$P(X \leq 30) = P\left(\frac{X - 36}{\sqrt{36}} \leq \frac{30 - 36}{\sqrt{36}}\right) = P(Z \leq -1) \cong$$

$$\Phi(-1) = 0,1587,$$

gdzie $\Phi(z)$, $z \in (-\infty, \infty)$, jest dystrybuantą rozkładu $N(0,1)$.

Nierówność Czebyszewa

Twierdzenie. Niech zmienna losowa X ma wartość średnią μ oraz wariancję σ^2 . Wówczas

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2},$$

dla dowolnego $\varepsilon > 0$.



Dziękuję za uwagę