Statystyczna analiza danych SAD-2020/21

Wykład 3

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

- $$\boxed{X : S \to (-\infty, \infty)}$$

Przykłady.

- rzut parą kostek sześciennych:

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

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

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

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

Zmienne losowe

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

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

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

- czas obsługi klienta, $S = \{x : 0 \le x \le T\}$

$$\boxed{X : x \rightarrow X(x) = x}.$$

Zmienne losowe

<u>Definicja.</u> **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,... }, {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 =	3	2	2	2	1	1	1	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: $P(\{s \in S : X(s) = x\})$ = $P(X = x)$

Dyskretne zmienne losowe

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

<u>**Stwierdzenie.**</u> Niech $X : S \to \{x_1, x_2, \ldots\}$. Wówczas

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

<u>**D.**</u> Z definicji funkcji prawdopodobieństwa i aksjomatów prawdopodobieństwa:

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

Dystrybuanta

Definicja.

Dystrybuantą zmiennej losowej X nazywamy funkcję:

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

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

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

Wyznaczanie dystrybuanty

Przykład. Trzykrotny rzut monetą:

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

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

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

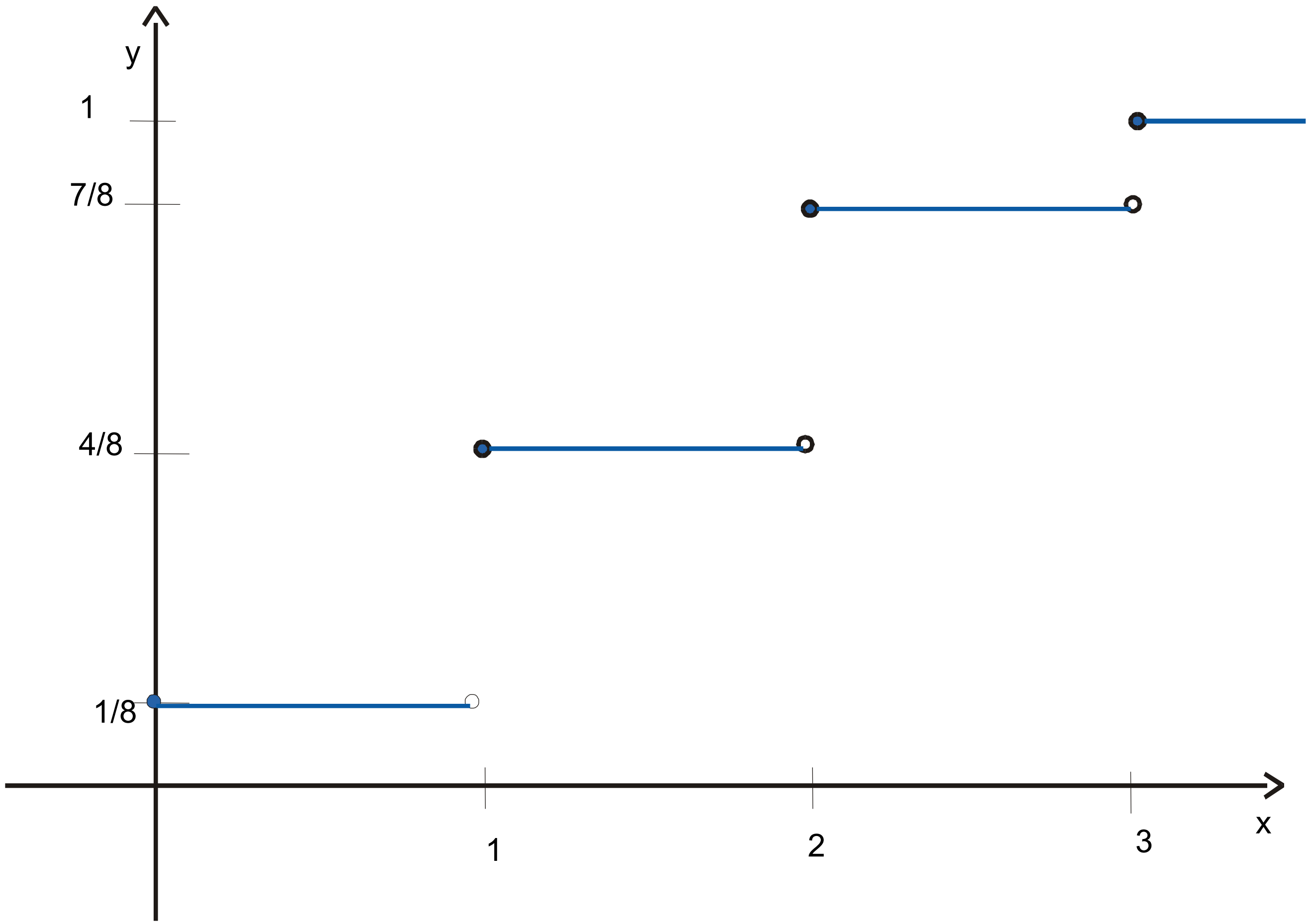
$$P(X \leq 2) = P(X \leq 1) + P(X = 2) = 4/8 + 3/8 = 7/8$$

Wyznaczanie dystrybuanty

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

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

Wykres dystrybuanty *F*



Wyznaczanie dystrybuanty d.z.l.

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

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

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

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

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

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

…….

$$x_j \leq x < x_{j+1} \quad \Longrightarrow F(x) = p(x_1) + \cdots + p\left(x_{j-1}\right) + p\left(x_j\right) =$$
$$F\left(x_{j-1}\right) + p\left(x_j\right),\ \ j = 2, .., k-1$$

Własności dystrybuanty

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

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

Prawdopodobieństwo a dystrybuanta

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

Prawdopodobieństwo a dystrybuanta

D.

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

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

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

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

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

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

$$= \; F(b) - F(a) + p(a).$$

Wartość oczekiwana (średnia)

Definicja.

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

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

gdzie $x_1, x_2, \ldots$ 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:

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

- $P(X=0) = P(X \le 0) - P(X < 0) = F(0) - F(0^-) = 0{,}5$
- $P(X=100) = P(X \le 100) - P(X < 100) =$ $F(100) - F(100^-) = 0{,}75 - 0{,}5 = 0{,}25$

Wartość oczekiwana

- $P(X = 200) = P(X \leq 200) - P(X < 200) =$

  $F(200) - F(200^{-}) = 1 - 0{,}75 = 0{,}25.$

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

<u>Twierdzenie.</u>

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

Wariancja

<u>Definicja.</u> 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: $\sigma_X = \sqrt{\sigma_X^2}$

<u>Uwaga.</u> $\sigma_X^2 = E(X - \mu_X)^2$

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

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

Funkcja prawdopodobieństwa:

x	x_1	x_2
$p(x)$	p	q

Rozkład Bernoulli'ego

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

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

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

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

Dyskretny rozkład jednostajny

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

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

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

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

Przykład: *X* = liczba oczek w rzucie kostką sześcienną.

Rozkład dwumianowy

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

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

$$\boxed{\mu_X = np}, \qquad \boxed{\sigma_X^2 = np(1-p)}.$$

Rozkład dwumianowy

Uzasadnienie:

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

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

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

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

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.

$X \sim$ 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

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

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

<u>Notacja:</u> $X \sim P(\lambda)$

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

Rozkład Poissona

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

Własności rozkładu Poissona;

- Niech $n \to \infty, \;\; p = p_n \to 0, \; np = \lambda > 0.$

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

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

Ciągłe zmienne losowe

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

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

Dystrybuanta c.z.l.

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

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

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

czyli dystrybuantę znajdujemy na podstawie gęstości

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

Dystrybuanta i gęstość c.z.l.

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

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

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

$$P(X = c) = \int_{c}^{c} f(t)\,dt = 0$$

- Dowolność $a \le b \;\; \Longrightarrow \qquad f(x) \ge 0,\;\; x \in (-\infty, \infty)$

Dystrybuanta

Definicja. Funkcję $\boxed{F(x) = \int_{-\infty}^{x} f(s)ds}, \quad x \in (-\infty, \infty),$

nazywamy **dystrybuantą** zmiennej losowej ***X***.

- $F(x) = P(X \le x),$ dla każdego *x*
- $P(-\infty \le X \le \infty) = P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(s)ds = 1$
- $f(x) \ge 0,$ dla każdego *x*
- $P(X = c) = 0$, dla każdej stałej *c*

Dystrybuanta

Stwierdzenie. Dla ciągłej zmiennej losowej o dystrybuancie F zachodzi

$$P(a < X < b) = P(a < X \le b) = P(a \le X < b) =$$

$$= P(a \le X \le b) = F(b) - F(a).$$

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

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

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

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:

- $f(x) \geq 0,\ x \in (-\infty, \infty),$
- $\int\limits_{-\infty}^{\infty} f(x)dx = 1,$

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.

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

Charakterystyki liczbowe c.z.l.

Wskaźniki położenia i rozproszenia dla ciągłych zmiennych losowych

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

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

Charakterystyki liczbowe c.z.l.

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

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

Np.

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

Charakterystyki liczbowe c.z.l.

<u>Definicja</u>. **Wariancją** ciągłej zmiennej losowej *X* o gęstości *f* nazywamy liczbę

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

Odchylenie standardowe:

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

<u>Uwaga.</u> Z definicji wariancji oraz wartości oczekiwanej funkcji zmiennej losowej

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

Własności wartości średniej i wariancji

<u>Twierdzenie</u>. Jeśli ciągła zmienna losowa ma wariancję, to dla dowolnych liczb *a*, *b* zachodzą wzory

- $$\mu_{aX+b} = a\mu_X + b$$
- $$\sigma^2_{aX+b} = a^2\sigma^2_X$$
- $$\sigma^2_X = \mu_{X^2} - (\mu_X)^2.$$

Powyższe wzory wynikają z własności (liniowości) całki.

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

Standaryzacja zmiennej losowej

<u>Stwierdzenie.</u> (standaryzacja)

Jeśli zmienna losowa X ma wartość średnią $\boxed{\mu_X}$ oraz wariancję $\boxed{\sigma_X^2}$, to <u>standaryzowana zmienna losowa</u>

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

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

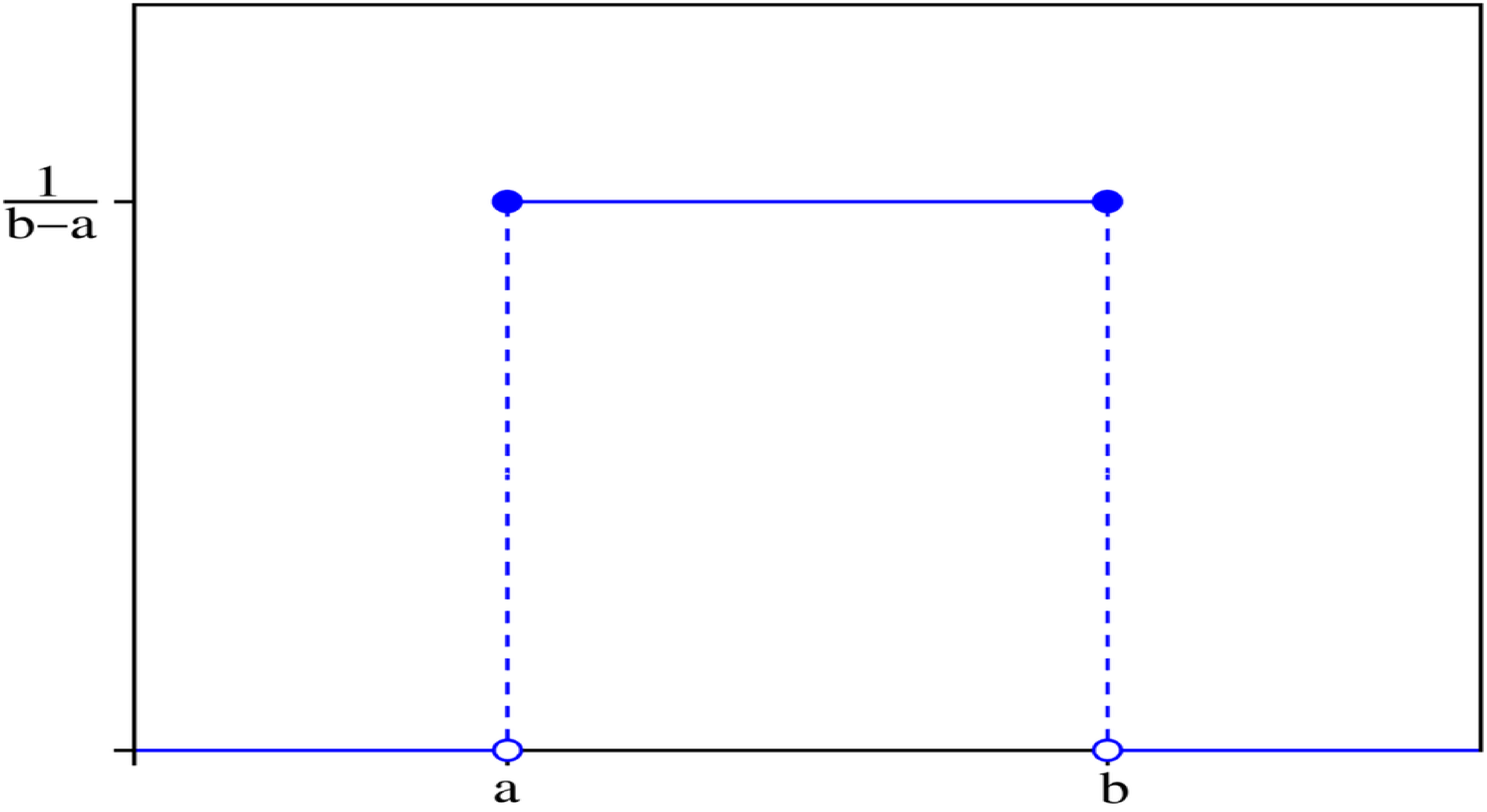
- Zmienna losowa ***X*** ma rozkład <u>jednostajny na przedziale [*a*,*b*],</u> jeśli ma gęstość:

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

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

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

Wykres gęstości rozkładu U(a,b)



Dystrybuanta rozkładu U(a,b)

1) Niech $x < a$.

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

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

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

$$= \frac{1}{b-a}[t]_a^x = \frac{1}{b-a}(x-a) = \frac{1}{b-a} \cdot x - \frac{a}{b-a}$$

3) Niech $x > b$.

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

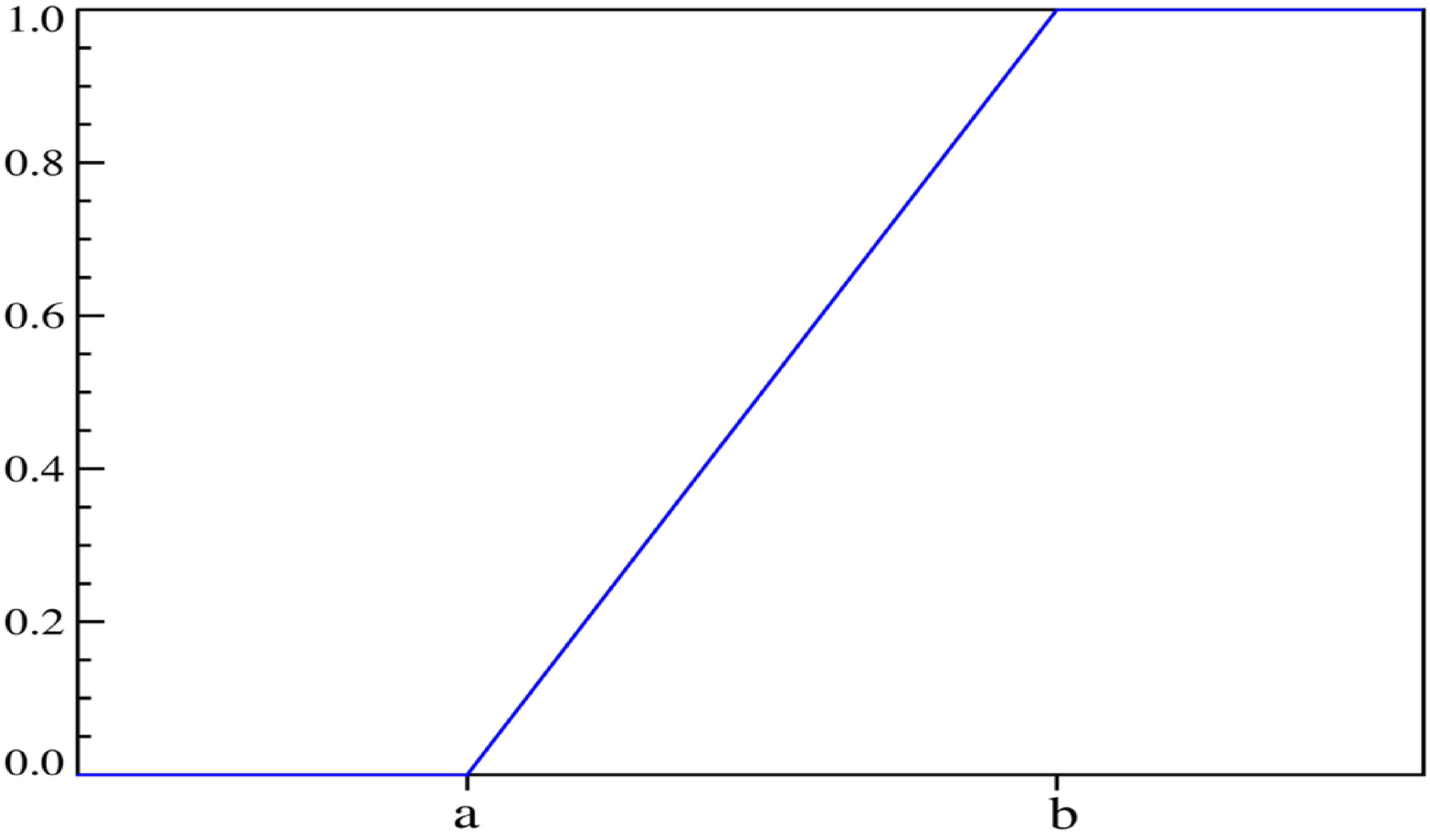
Dystrybuanta rozkładu U(a,b)

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

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

Jej wykres jest przedstawiony na kolejnym slajdzie.

Dystrybuanta rozkładu jednostajnego



Ciągłe zmienne losowe - PRZYKŁADY

Zmienna losowa o rozkładzie wykładniczym

Niech $\boxed{X_t}$ ma rozkład Poissona $\boxed{P(\lambda t)}$ (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**

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

Ciągłe zmienne losowe

Zmienna losowa ma rozkład <u>wykładniczy z parametrem </u>λ,

$\lambda > 0$, jeśli ma gęstość $f(t) = F'(t)$:

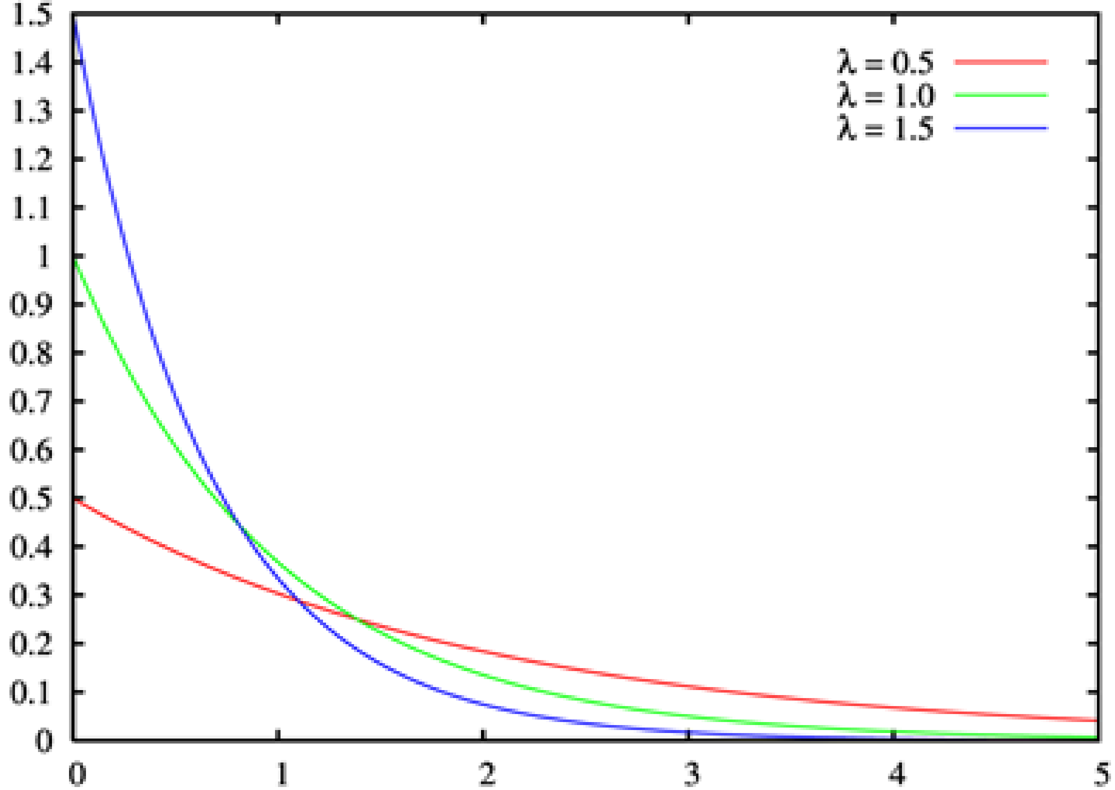
$$f(t) = \begin{cases} 0 & \text{dla } t < 0 \\ \lambda e^{-\lambda t} & \phantom{\text{dla }} t \geq 0 \end{cases}.$$

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

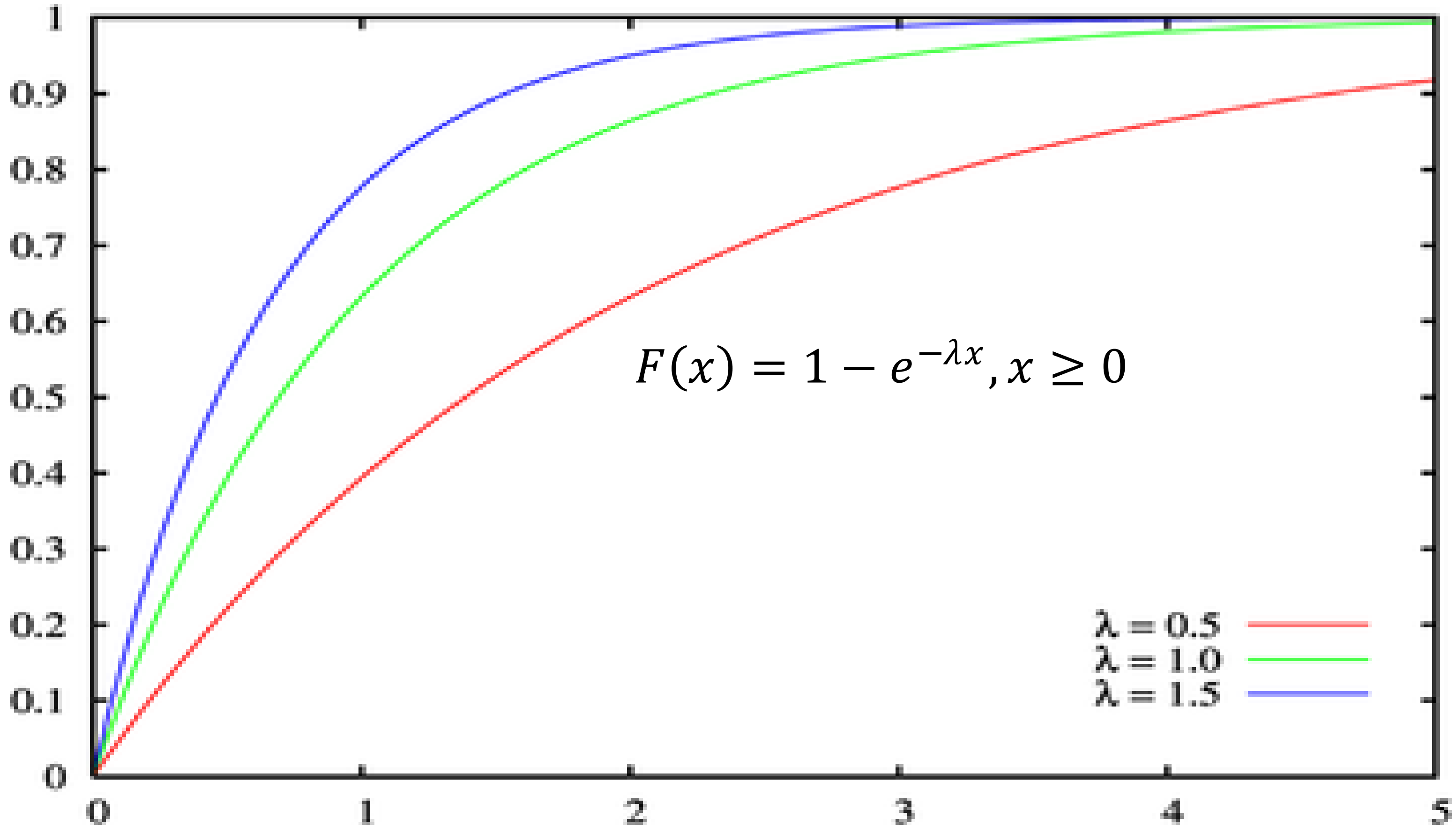
$$\sigma_T^2 = \mu_{T^2} - (\mu_T)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}.$$

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

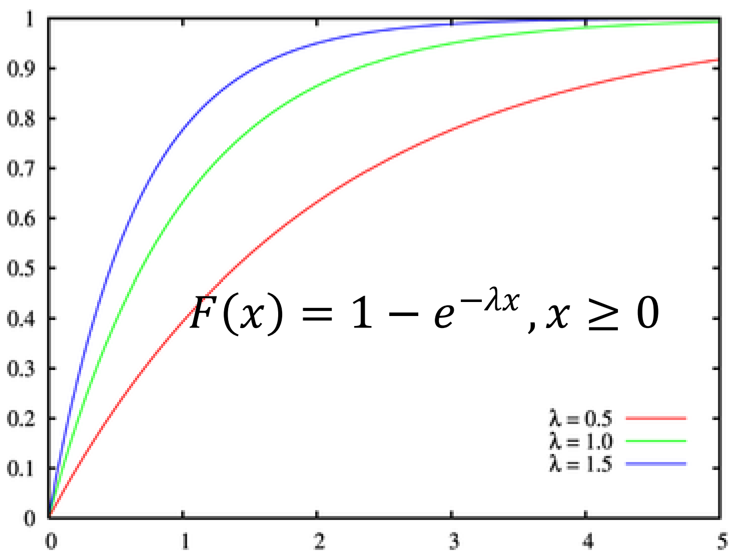
Gęstości rozkładu wykładniczego

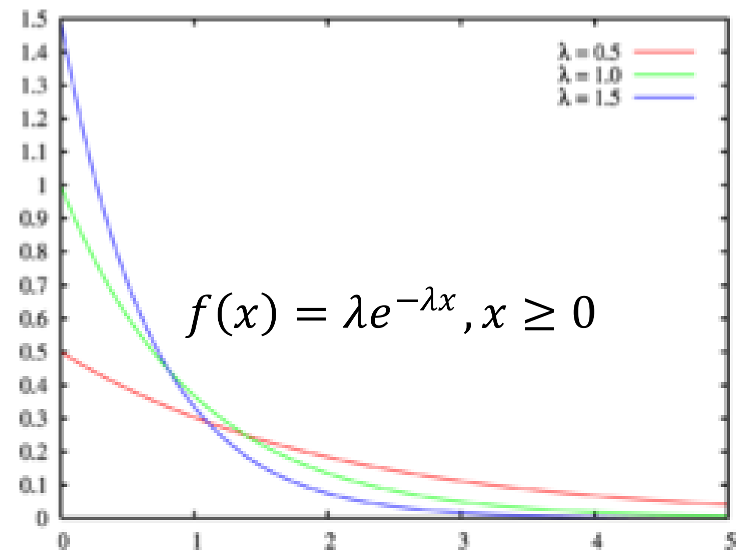


Wykresy dystrybuant r. Exp(λ)



Dystrybuanta i gęstość r. wykł.





Ciągłe zmienne losowe

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

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

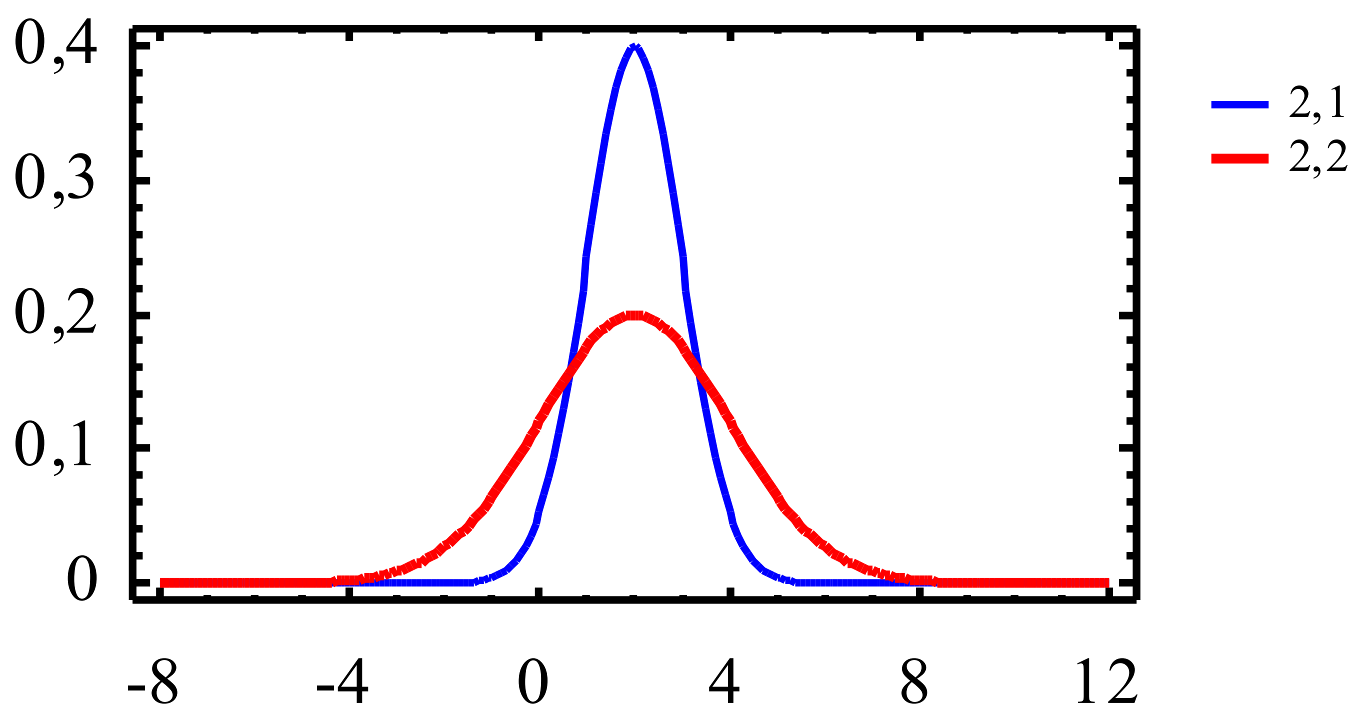
$$-\infty < x < \infty,$$

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

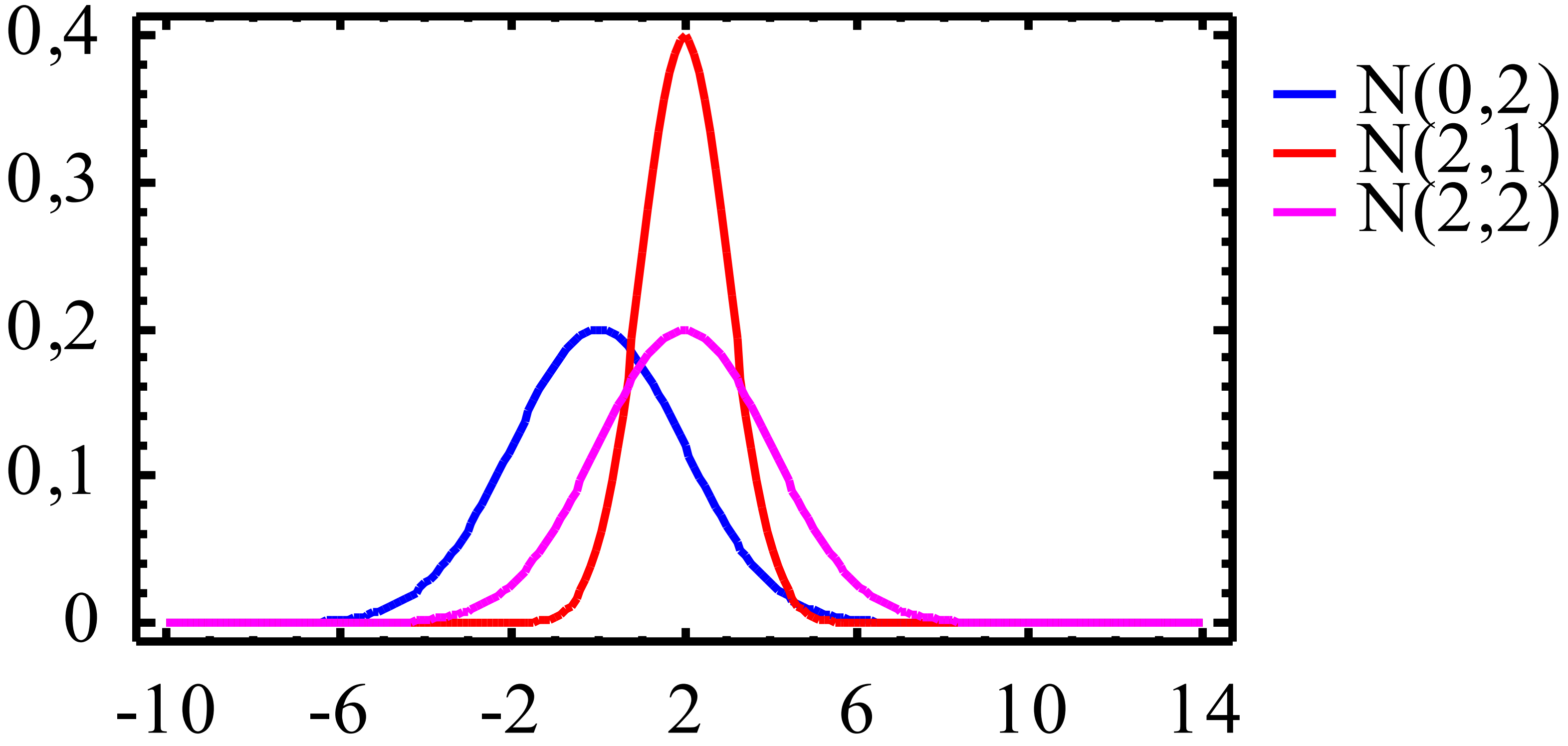
Notacja: $X \sim N(\mu, \sigma)$.

Wykresy gęstości normalnych

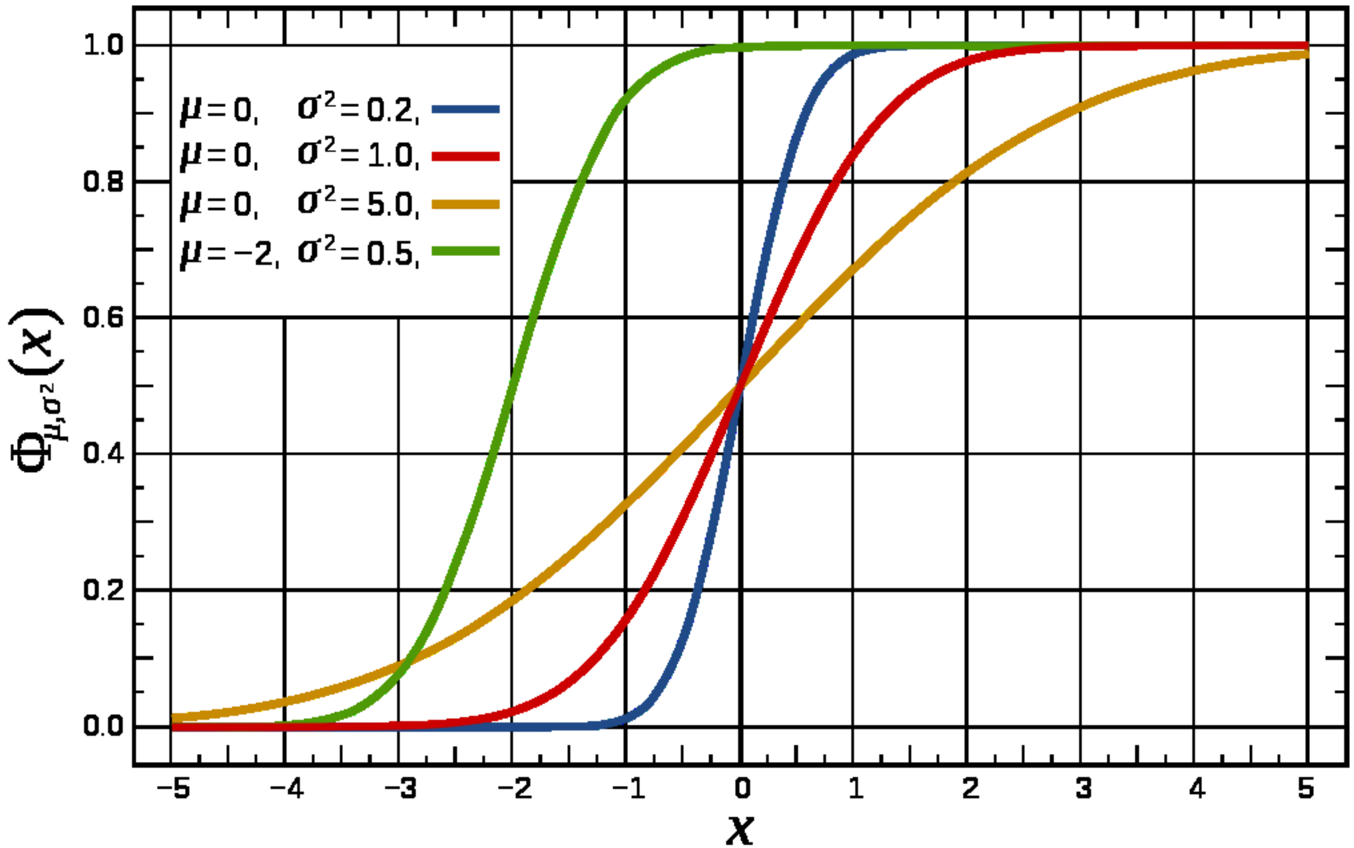
$F(2) = F(2) = 0{,}5$, bo wykres gęstości symetryczny względem prostej x=2, a pole pod wykresem gęstości po całej prostej wynosi 1.



Wykresy gęstości normalnych



Wykresy dystrybuant rozkładu normalnego



Własności zmiennej losowej o rozkładzie normalnym

Twierdzenie. Niech

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

Wówczas

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

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

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

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

Stąd

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

Rozkład Poissona a rozkład normalny

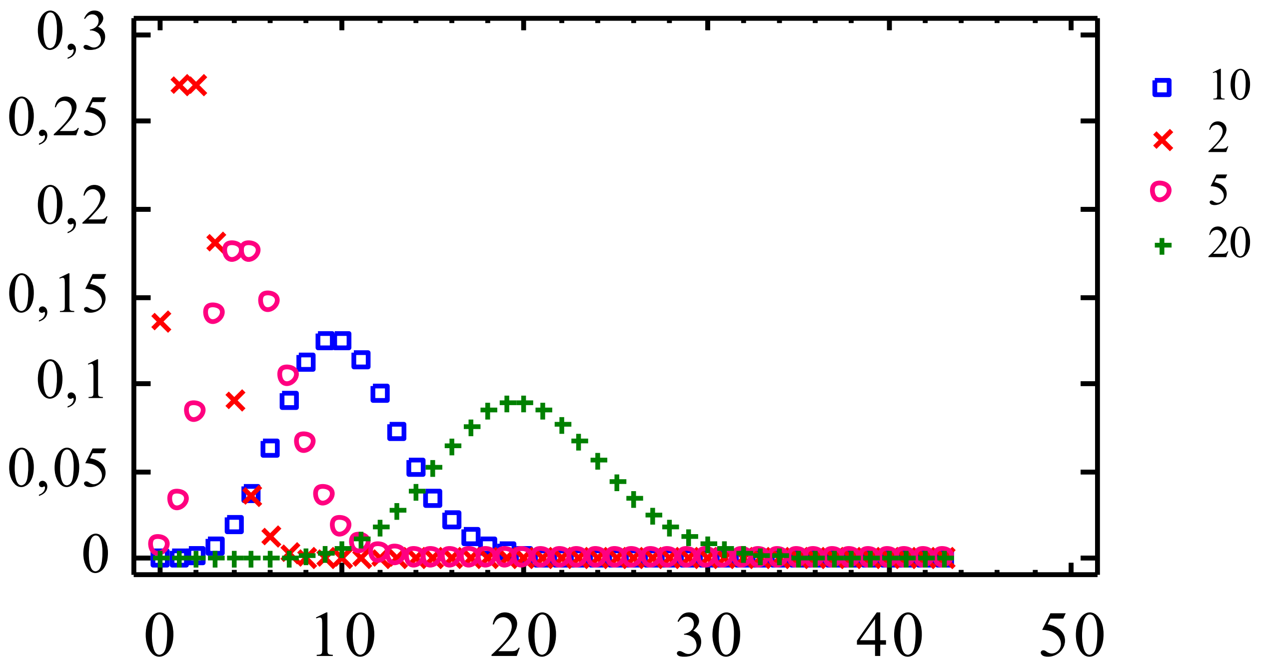
- Jeśli $\boxed{X \sim P(\lambda)}$ dla dużego λ, to rozkład standaryzowanej zmiennej $(X-\lambda)/\sqrt{\lambda}$ jest w przybliżeniu normalny, tzn. $P((X-\lambda)/\sqrt{\lambda} \le z) \approx \Phi(z)$,

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

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

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

Rozkład Poissona



Rozkład Poissona a rozkład normalny

<u>Przykład.</u> 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 ?

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

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

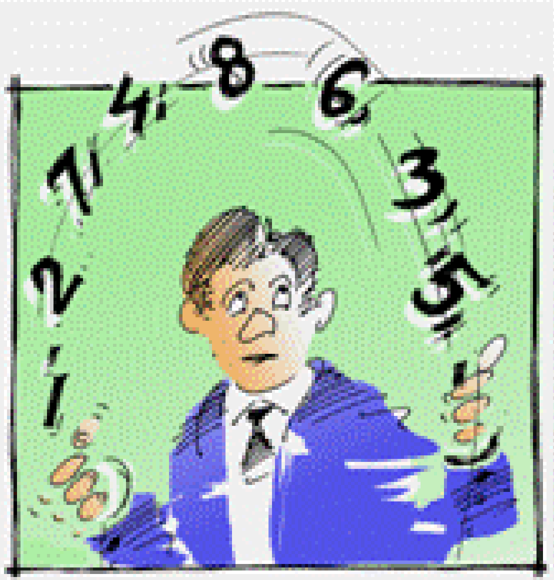
gdzie $\Phi(z), z \in (-\infty, \infty)$, jest dystrybuantą rozkładu $N(0{,}1)$.

Nierówność Czebyszewa

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

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

dla dowolnego $\varepsilon > 0$.



Dziękuję za uwagę