Poprzedni wykład

- Dystrybuanta $F_X(x) = P(X \le x)$
- funkcja rozkładu prawdopodobieństwa i funkcja gęstości prawdopodobieństwa

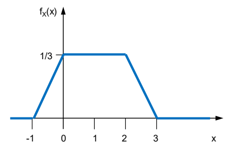
$$\forall k:\ P_X(x_k) = P(X = x_k) \qquad f_X(x) = \frac{d}{dx} F_X(x)$$

$$P(a < X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(t)dt$$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Przykład: FGP (trapez) i odwracanie dystrybuanty

Syt: Zmienna losowa X opisywana jest funkcją gęstości prawdopodobieństwa w kształcie trapezu



$$f_X(x) = \begin{cases} 0 & dla\ x \in (-\infty, -1] \\ \frac{1}{3}x + \frac{1}{3} & dla\ x \in (-1, 0] \\ \frac{1}{3} & dla\ x \in (0, 2] \\ \frac{-1}{3}x + 1 & dla\ x \in (2, 3] \\ 0 & dla\ x \in (3, +\infty] \end{cases}$$

- Wysokość trapezu obliczamy z warunku normalizacji

$$Pole\ trapezu = \frac{(a+b)\cdot h}{2} = \frac{(4+2)\cdot h}{2} = 3h \quad \to \quad 1 = 3h \quad \to \quad h = \frac{1}{3}$$

- Funkcję gęstości prawdopodobieństwa obliczamy wiedząc przez jakie punkty przechodzą poszczególne odcinki

Przykład: FGP (trapez) i odwracanie dystrybuanty

- Obliczamy dystrybuantę $F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$

$$f_X(x) = \begin{cases} 0 & dla\ x \in (-\infty, -1] \\ \frac{1}{3}x + \frac{1}{3} & dla\ x \in (-1, 0] \\ \frac{1}{3} & dla\ x \in (0, 2] \\ \frac{-1}{3}x + 1 & dla\ x \in (2, 3] \\ 0 & dla\ x \in (3, +\infty] \end{cases}$$

$$dla\ x \in (-\infty, -1]: \quad F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{x} 0\, dt = 0$$

$$dla\ x \in (-1, 0]: \quad F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{-1} 0\, dt + \int_{-1}^{x} \tfrac{1}{3}t + \tfrac{1}{3}\, dt = 0 + \left(\tfrac{1}{6}t^2 + \tfrac{1}{3}t\right)\Big|_{-1}^{x} =$$

$$= \tfrac{1}{6}x^2 + \tfrac{1}{3}x - \tfrac{1}{6}(-1)^2 - \tfrac{1}{3}(-1) = \tfrac{1}{6}x^2 + \tfrac{1}{3}x + \tfrac{1}{6}$$

$$dla\ x \in (0, 2]: \quad F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{0} f_X(t)\, dt + \int_{0}^{x} \tfrac{1}{3}\, dt = F_X(0) + \int_{0}^{x} \tfrac{1}{3}\, dt =$$

$$= \left(\tfrac{1}{6}\cdot 0^2 + \tfrac{1}{3}\cdot 0 + \tfrac{1}{6}\right) + \left(\tfrac{1}{3}t\right)\Big|_{0}^{x} = \tfrac{1}{6} + \tfrac{1}{3}x - \tfrac{1}{3}\cdot 0 = \tfrac{1}{3}x + \tfrac{1}{6}$$

$$dla\ x \in (2, 3]: \quad F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{2} f_X(t)\, dt + \int_{2}^{x} \tfrac{-1}{3}t + 1\, dt = F_X(2) + \int_{2}^{x} \tfrac{-1}{3}t + 1\, dt =$$

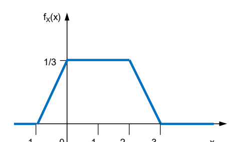
$$= \left(\tfrac{1}{3}\cdot 2 + \tfrac{1}{6}\right) + \left(\tfrac{-1}{6}t^2 + t\right)\Big|_{2}^{x} = \tfrac{5}{6} - \tfrac{1}{6}x^2 + x - \left(\tfrac{-1}{6}\cdot 2^2\right) - 2 = -\tfrac{1}{6}x^2 + x - \tfrac{1}{2}$$

$$dla\ x \in (3, +\infty]: \quad F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt = \int_{-\infty}^{3} f_X(t)\, dt + \int_{3}^{x} 0\, dt = F_X(3) + 0 = -\tfrac{1}{6}\cdot 3^2 + 3 - \tfrac{1}{2} = 1$$

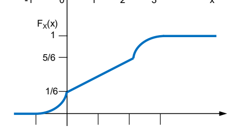
Przykład: FGP (trapez) i odwracanie dystrybuanty

- Łącznie funkcja gęstości prawdopodobieństwa i dystrybuanta

$$f_X(x) = \begin{cases} 0 & dla\ x \in (-\infty, -1] \\ \frac{1}{3}x + \frac{1}{3} & dla\ x \in (-1, 0] \\ \frac{1}{3} & dla\ x \in (0, 2] \\ \frac{-1}{3}x + 1 & dla\ x \in (2, 3] \\ 0 & dla\ x \in (3, +\infty] \end{cases}$$



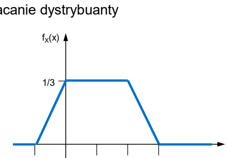
$$F_X(x) = \begin{cases} 0 & dla\ x \in (-\infty, -1] \\ \frac{1}{6}x^2 + \frac{1}{3}x + \frac{1}{6} & dla\ x \in (-1, 0] \\ \frac{1}{3}x + \frac{1}{6} & dla\ x \in (0, 2] \\ \frac{-1}{6}x^2 + x - \frac{1}{2} & dla\ x \in (2, 3] \\ 1 & dla\ x \in (3, +\infty] \end{cases}$$



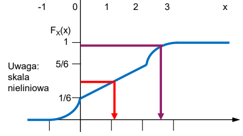
Przykład: FGP (trapez) i odwracanie dystrybuanty

- Generacja liczb pseudolosowych: odwracanie dystrybuanty

$$f_X(x) = \begin{cases} 0 & dla\ x \in (-\infty, -1] \\ \frac{1}{3}x + \frac{1}{3} & dla\ x \in (-1, 0] \\ \frac{1}{3} & dla\ x \in (0, 2] \\ \frac{-1}{3}x + 1 & dla\ x \in (2, 3] \\ 0 & dla\ x \in (3, +\infty] \end{cases}$$



$$F_X(x) = \begin{cases} 0 & dla\ x \in (-\infty, -1] \\ \frac{1}{6}x^2 + \frac{1}{3}x + \frac{1}{6} & dla\ x \in (-1, 0] \\ \frac{1}{3}x + \frac{1}{6} & dla\ x \in (0, 2] \\ \frac{-1}{6}x^2 + x - \frac{1}{2} & dla\ x \in (2, 3] \\ 1 & dla\ x \in (3, +\infty] \end{cases}$$



Przykład: FGP (trapez) i odwracanie dystrybuanty

- Obliczamy funkcję odwrotną do dystrybuanty $F_X^{-1}(x)$

$$F_X(x) = \begin{cases} 0 & dla\ x \in (-\infty, -1] \\ \frac{1}{6}x^2 + \frac{1}{3}x + \frac{1}{6} & dla\ x \in (-1, 0] \\ \frac{1}{3}x + \frac{1}{6} & dla\ x \in (0, 2] \\ \frac{-1}{6}x^2 + x - \frac{1}{2} & dla\ x \in (2, 3] \\ 1 & dla\ x \in (3, +\infty] \end{cases}$$

$$\begin{array}{ll} dla\ x \in (-1, 0]: & F_X(x) = \frac{1}{6}x^2 + \frac{1}{3}x + \frac{1}{6} \\ \quad y \in (0, \frac{1}{6}] & y = \frac{1}{6}x^2 + \frac{1}{3}x + \frac{1}{6} \\ zmiana\ oznacze\acute{n} & x = \frac{1}{6}y^2 + \frac{1}{3}y + \frac{1}{6} \\ & x = \frac{1}{6}\cdot\left(y^2 + 2y + 1\right) \\ & 6x = (y+1)^2 \\ & y = \sqrt{6x} - 1 \quad x \in (0, \frac{1}{6}] \end{array}$$

$$\begin{array}{ll} dla\ x \in (0, 2]: & F_X(x) = \frac{1}{3}x + \frac{1}{6} \\ \quad y \in (\frac{1}{6}, \frac{5}{6}] & y = \frac{1}{3}x + \frac{1}{6} \\ zmiana\ oznacze\acute{n} & x = \frac{1}{3}y + \frac{1}{6} \\ & y = 3x - \frac{1}{2} \quad x \in (\frac{1}{6}, \frac{5}{6}] \end{array}$$

$$\begin{array}{ll} dla\ x \in (2, 3]: & F_X(x) = \frac{-1}{6}x^2 + x - \frac{1}{2} \\ \quad y \in (\frac{5}{6}, 1] & y = \frac{-1}{6}x^2 + x - \frac{3}{6} \\ zmiana\ oznacze\acute{n} & x = \frac{-1}{6}y^2 + y - \frac{3}{6} \\ & x = \frac{-1}{6}\cdot\left(y^2 - 6y + 3\right) \\ & x = \frac{-1}{6}\cdot\left((y-3)^2 - 6\right) \\ & -6x + 6 = (y-3)^2 \\ & -\sqrt{-6x+6} = y - 3 \\ & y = 3 - \sqrt{6 - 6x} \quad x \in (\frac{5}{6}, 1] \end{array}$$

Przykład: FGP (trapez) i odwracanie dystrybuanty

- Łącznie $F_x^{-1}(x)$

$$F_x^{-1}(x)=\begin{cases}\sqrt{6x}-1 & dla\ x\in(0,\frac{1}{6}]\\ 3x-\frac{1}{2} & dla\ x\in(\frac{1}{6},\frac{5}{6}]\\ 3-\sqrt{6-6x} & dla\ x\in(\frac{5}{6},1]\end{cases}$$

Generator liczb pseudolosowych o badanej $f_X(x)$:

- Losujemy jednorodnie liczbę *x* z przedziału (0,1)
- Obliczamy $y:=F_x^{-1}(x)$ zgodnie z powyższym wzorem
- Zwracamy *y*
- Liczby y będą miały zadaną funkcję gęstości prawdopodobieństwa w przedziale (-1,3)

Kiedy metoda odwracania dystrybuanty nie zadziała (od razu) ?

Odwracanie dystrybuanty jako metoda generowania liczb pseudolosowych

- Syt: dana funkcja gęstości prawdopodobieństwa $f_X(x)$, chcemy uzyskać liczby pseudolosowe przez nią opisywane.

- Obliczamy dystrybuantę $F_X(x)$.
- Znajdujemy funkcję odwrotną do dystrybuanty $F^{-1}_X(x)$.
- Losujemy liczbę *y* z przedziału <0,1> używając „standardowego" (zaimplementowanego) generatora liczb pseudolosowych.
- Obliczmy $z=F^{-1}_X(y)$.
- Tak otrzymane liczby opisane są funkcją gęstości prawdopodobieństwa $f_X(x)$

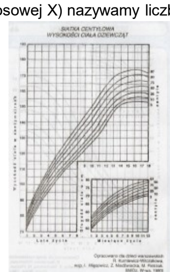
Globalny opis dystrybuanty i funkcji gęstości prawdopodobieństwa

- **Kwantylem rzędu p** (dla zmiennej losowej X) nazywamy liczbę x_p: $F_X(x_p)=p \quad (0\le p\le 1)$

W szczególności:

- **Mediana** to kwantyl rzędu ½
- **Kwartyle** to kwantyle rzędu ¼ (pierwszy kwartyl), rzędu ½ (drugi kwartyl), rzędu ¾ (trzeci kwartyl)
- **Percentyle** to kwantyle rzędu 0.01, 0.02, …, 0.99

- **Moda** – wartość najbardziej prawdopodobna czyli x: max{ $f_X(x)$ }



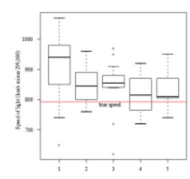
Globalny opis dystrybuanty i funkcji gęstości prawdopodobieństwa

- **Kwantylem rzędu p** (dla zmiennej losowej X) nazywamy liczbę x_p: $F_X(x_p)=p \quad (0\le p\le 1)$

W szczególności:

- **Mediana** to kwantyl rzędu ½
- **Kwartyle** to kwantyle rzędu ¼ (pierwszy kwartyl), rzędu ½ (drugi kwartyl), rzędu ¾ (trzeci kwartyl)
- **Percentyle** to kwantyle rzędu 0.01, 0.02, …, 0.99

- **Moda** – wartość najbardziej prawdopodobna czyli x: max{ $f_X(x)$ }



Źródło: Wikipedia

Wartość oczekiwana

- **Wartością oczekiwaną** zmienną losowej X nazywamy liczbę

$$E(X)=\sum_k P_X(x_k)\,x_k \quad \text{dla zmiennej dyskretnej}$$

$$E(X)=\int_{-\infty}^{+\infty} f_X(x)\,x\,dx \quad \text{dla zmiennej ciągłej}$$

- Uwaga 1:
  E(X) jest średnią ważoną możliwych wartości zmiennej X z wagą daną przez $P_X(x)$ lub $f_X(x)$.
- Uwaga 2:
  Dla zmiennej typu mieszanego łączymy obie definicje.
- Uwaga 3:
  Spotyka się też warunkową wartość oczekiwaną, wtedy $P_X(x)\to P_X(x|A)$, $f_X(x)\to f_X(x|A)$.

Wartość oczekiwana - własności

- Operator E( ) jest liniowy: $E\left(\sum_{k=1}^{n} a_k g_k(x)\right)=\sum_{k=1}^{n} a_k E\left(g_k(x)\right)$

a_k – stałe,
$g_k(x)$ – funkcje o wartościach rzeczywistych

Stąd (a,b – stałe):
E(a)=a
E(aX+b)=aE(X)+b

- Niech *Y* to nowa zmienna losowa i niech *Y=g(x)* wtedy

$$E(Y)=\sum_k P_X(x_k)\,g(x_k) \quad \text{dla zmiennej dyskretnej}$$

$$E(Y)=\int_{-\infty}^{+\infty} f_X(x)\,g(x)\,dx \quad \text{dla zmiennej ciągłej}$$

Przykład: wartość oczekiwana funkcji zmiennej losowej

Niech zmienna losowa X ma rozkład jednorodny na (0,1):

$$f_X(x) = \begin{cases} 1 & dla\ x \in (0,1) \\ 0 & dla\ x \notin (0,1) \end{cases}$$

Niech $Y=X^2$.

Znaleźć wartości oczekiwane zmiennych X i Y.

$$E(X) = \int_{-\infty}^{+\infty} f_X(x) x\, dx = \int_0^1 x\, dx = \tfrac{1}{2}x^2\Big|_0^1 = \tfrac{1}{2}$$

$$E(Y) = E(X^2) = \int_{-\infty}^{+\infty} f_X(x) x^2\, dx = \int_0^1 x^2\, dx = \tfrac{1}{3}x^3\Big|_0^1 = \tfrac{1}{3}$$

czyli aby wyliczyć E(Y) nie musimy znać $f_Y(y)$.

gdybyśmy jednak ją znali to można liczyć $E(Y) = \int_{-\infty}^{+\infty} f_Y(y) y\, dy$

Wariancja zmiennej losowej

- **Wariancją zmiennej losowej** nazywamy liczbę:

$$\boxed{var(X) \equiv \sigma_X^2 \equiv \sigma^2(X) = E\left((X - E(X))^2\right)}$$

- Wariancja jest miarą rozrzutu zmiennej losowej wokół wartości średniej i jest nieujemna
- var(a)=0 dla a=const
- E(X-E(X)) nie jest przydatną wielkością, gdyż

$$E(X - E(X)) = E(X) - E(E(X)) = E(X) - E(X) = 0$$

- **Odchylenie standardowe zmiennej losowej**

$$\sigma_X \equiv \sigma(X) = \sqrt{var(X)}$$

Wariancja zmiennej losowej - własności

- $\boxed{var(aX + b) = a^2\, var(X)}$

Dowód:

$$var(aX + b) = E\left((aX + b - E(aX + b))^2\right) =$$
$$= E\left((aX + b - aE(X) - b)^2\right) = E\left((aX - aE(X))^2\right) =$$
$$= E\left(a^2(X - E(X))^2\right) = a^2 E\left((X - E(X))^2\right) = a^2 var(X)$$

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

Dowód:

$$var(X) = E\left((X - E(X))^2\right) = E\left(X^2 - 2XE(X) + (E(X))^2\right) =$$
$$= E(X^2) - 2E(XE(X)) + E\left((E(X))^2\right) =$$
$$= E(X^2) - 2(E(X))^2 + (E(X))^2 = E(X^2) - (E(X))^2$$

Przykład: E(X), var(X), σ(X) dla rozkładu jednorodnego

Niech zmienna losowa X ma rozkład jednorodny na [a,b]:

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

$$E(X) = \int_{-\infty}^{+\infty} f_X(x) x\, dx = \int_a^b \tfrac{1}{b-a} x\, dx = \tfrac{1}{b-a}\int_a^b x\, dx = \tfrac{1}{b-a}\left(\tfrac{1}{2}x^2\right)\Big|_a^b = \tfrac{1}{2(b-a)}(b^2 - a^2) =$$
$$= \tfrac{(b-a)(b+a)}{2(b-a)} = \tfrac{a+b}{2}$$

$$\mathrm{var}(X) = E(X^2) - \left(E(X)\right)^2$$

$$E(X^2) = \int_{-\infty}^{+\infty} f_X(x) x^2\, dx = \tfrac{1}{b-a}\int_a^b x^2\, dx = \tfrac{1}{3(b-a)}(b^3 - a^3) = \tfrac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \tfrac{b^2+ab+a^2}{3}$$

$$\mathrm{var}(X) = \tfrac{b^2+ab+a^2}{3} - \left(\tfrac{a+b}{2}\right)^2 = \tfrac{1}{12}\left(4b^2 + 4ab + 4a^2 - 3a^2 - 3b^2 - 6ab\right) =$$
$$= \tfrac{1}{12}\left(a^2 - 2ab + b^2\right) = \tfrac{(b-a)^2}{12}$$

$$\sigma(X) = \sqrt{\mathrm{var}(X)} = \sqrt{\tfrac{(b-a)^2}{12}} = \tfrac{b-a}{2\sqrt{3}} = \tfrac{1}{\sqrt{3}} \cdot \tfrac{b-a}{2}$$

czynnik $1/\sqrt{3}$ pojawia się w teorii błędów pomiarowych

Przykład: E(X), var(X), σ(X) dla rozkładu geometrycznego

Niech zmienna losowa X ma rozkład geometryczny:

$$P(X = k) = (1-p)^{k-1} p \qquad k = 1, 2, \ldots$$

Rozkład ten jest unormowany, zatem: $\sum_{k=1}^{\infty} (1-p)^{k-1} p = 1$

$$1 = \sum_k (1-p)^{k-1} p$$
$$1 = p \sum_k (1-p)^{k-1}$$
$$\tfrac{1}{p} = \sum_k (1-p)^{k-1} \quad / \tfrac{d}{dp}$$
$$\tfrac{-1}{p^2} = \sum_k (k-1)(1-p)^{k-2}(-1)$$
$$\tfrac{1}{p^2} = \sum_k k(1-p)^{k-2} - \sum_k (1-p)^{k-2}$$
$$\tfrac{1}{p^2} = \tfrac{1}{1-p}\sum_k k(1-p)^{k-1} - \tfrac{1}{1-p}\sum_k (1-p)^{k-1} \quad / \cdot p$$

$$\tfrac{1}{p} = \tfrac{1}{1-p}\underbrace{\sum k(1-p)^{k-1} p}_{E(X)} - \tfrac{1}{1-p}\underbrace{\sum (1-p)^{k-1} p}_{1}$$
$$\tfrac{1}{p} = \tfrac{1}{1-p} E(X) - \tfrac{1}{1-p} \cdot 1 \quad / \cdot (1-p)$$
$$\tfrac{1-p}{p} = E(X) - 1$$
$$E(X) = \tfrac{1-p}{p} + 1 = \tfrac{1-p+p}{p} = \tfrac{1}{p}$$

Przykład: E(X), var(X), σ(X) dla rozkładu geometrycznego

$$P(X = k) = (1-p)^{k-1} p \qquad k = 1, 2, \ldots$$

Aby policzyć $E(X^2)$ korzystamy z E(X)

$$E(X) = \sum_k k(1-p)^{k-1} p = \tfrac{1}{p}$$
$$\sum_k k(1-p)^{k-1} = \tfrac{1}{p^2} \quad / \tfrac{d}{dp}$$
$$\sum_k k(k-1)(1-p)^{k-2}(-1) = \tfrac{-2}{p^3}$$
$$\sum_k k^2(1-p)^{k-2} - \sum_k k(1-p)^{k-2} = \tfrac{2}{p^3}$$
$$\tfrac{1}{1-p}\sum_k k^2(1-p)^{k-1} - \tfrac{1}{1-p}\sum_k k(1-p)^{k-1} = \tfrac{2}{p^3} \quad / \cdot p(1-p)$$
$$\underbrace{\sum k^2(1-p)^{k-1} p}_{E(X^2)} - \underbrace{\sum k(1-p)^{k-1} p}_{E(X)} = \tfrac{2(1-p)}{p^2}$$
$$E(X^2) = \tfrac{2-2p}{p^2} + \tfrac{1}{p} = \tfrac{2-2p+p}{p^2} = \tfrac{2-p}{p^2}$$

$$\mathrm{var}(X) = E(X^2) - \left(E(X)\right)^2 =$$
$$= \tfrac{2-p}{p^2} - \left(\tfrac{1}{p}\right)^2 = \tfrac{2-p-1}{p^2} = \tfrac{1-p}{p^2}$$
$$\sigma(X) = \sqrt{\mathrm{var}(X)} = \tfrac{1}{p}\sqrt{1-p}$$

Dla rzutu kostką i wypadnięcia „6”:
E(X)=6, var(X)=30, σ(X)=5.48