Poprzedni wykład

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

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

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

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

i ich charakterystyki

- kwantyl rzędu p (dla zmiennej losowej X) x_p: $F_X(x_p)=p$ $(0 \le p \le 1)$
- wartość oczekiwana X

$$E(X) = \sum_k P(x_k) x_k \qquad \text{lub} \qquad E(X) = \int_{-\infty}^{+\infty} f_X(x) x\, dx$$

- odchylenie standardowe

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

Poprzednim razem: 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

Niepewności pomiarowe - dygresja

- Niepewności systematyczne (dokładność przyrządu, wydajność przyrządu, przybliżenia teoretyczne, …)
- Niepewności statystyczne (pochodzące od losowej natury zjawiska i/lub czynników poza naszą kontrolą). Uwidacznia się przy powtarzaniu eksperymentu.

- Zał: zmierzona wartość X, oszacowana niepewność systematyczna Δ, oszacowane odchylenie standardowe σ(X)

- Zapis wyniku: X(σ(X))± Δ lub (mniej precyzyjne) X±σ(X)±Δ

  Np. X=2.34 Δ=0.01 σ(X) =0.12: 2.34(12) ±0.01 lub 2.34 ±0.12 ±0.01

  W tym przykładzie Δ<< σ(X) i wtedy Δ pomijamy: 2.34 ±0.12

Ale czasami chcemy podać jedną niepewność (=połączyć Δ oraz σ(X))

Nie ma dobrej teorii, która pokazuje jak to zrobić.

Istnieje teoria jak łączyć niepewności statystyczne z różnych źródeł

$$\sigma = \sqrt{(\sigma_1)^2 + (\sigma_2)^2 + \ldots}$$

Niektórzy zakładają, że niepewność systematyczna ma rozkład jednorodny co pozwala w miejsce Δ używać $\sigma = \frac{1}{\sqrt{3}}\Delta$ co prowadzi do wzoru $\sigma = \sqrt{(\sigma_{stat})^2 + \frac{1}{3}\Delta^2}$

Inne parametry - momenty

- **Momentem rzędu k względem początku układu współrzędnych** dla zmiennej losowej X nazywamy

$$\boxed{\tilde{\mu}_k \equiv E(X^k) \qquad k = 0,1,2,\ldots}$$

czyli

$$\tilde{\mu}_0 = 1 \qquad \tilde{\mu}_1 = E(X)$$

- **Momentem centralnym rzędu k** dla zmiennej losowej X nazywamy

$$\boxed{\mu_k \equiv E((X - E(X))^k) \qquad k = 0,1,2,\ldots}$$

- czyli

$$\mu_0 = 1 \qquad \mu_1 = 0 \qquad \mu_2 = var(X)$$

Inne parametry – skośność i kurtoza

- **Skośność** - miara asymetrii

$$\boxed{\beta_1 \equiv \frac{\mu_3}{(\sigma(X))^3}}$$

- **Kurtoza** - miara skupienia (spłaszczenia)

$$\boxed{\beta_2 \equiv \frac{\mu_4}{(\sigma(X))^4} - 3}$$

Przykład: rozkład prostokątny i rozkłady trójkątne

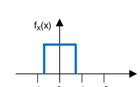
Przykład: rozkłady trójkątne

- Rozkład jednorodny na $x \in \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$

spr. (normalizacja)

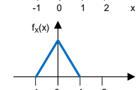
$$N = 2 \cdot \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}} = 1$$

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2}} & dla\ x \in \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \\ 0 & dla\ x \notin \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \end{cases}$$



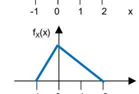
- Rozkład trójkątny I

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



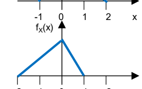
- Rozkład trójkątny II

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



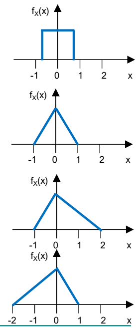
- Rozkład trójkątny III

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



Przykład: rozkłady trójkątne

	Prostokątny	Trójkątny I	Trójkątny II	Trójkątny III
E(X)	0	0	1/3	-1/3
$E(X^2)$	1/6	1/6	1/2	1/2
var(X)	1/6=0.16(6)	0.16(6)	7/18=0.38(8)	7/18
σ(X)	0.41	0.41	0.62	0.62
μ_3	0	0	2/27=0.07	-2/27
μ_4	0.05	0.06(6)	49/135=0.36	49/135
β_1	0	0	0.305	-0.305
β_2	-1.2	-0.6	-0.6	-0.6



Transformacje zmiennych losowych

Syt. X – zmienna losowa
g(X) – funkcja o wartościach rzeczywistych
Y=g(X) też jest zmienną losową

Pytanie: jak wygląda rozkład prawdopodobieństwa (funkcja gęstości prawdopodobieństwa) zmiennej losowej Y

- Zmienna X jest dyskretna
  Wtedy również Y jest dyskretne. Niech x_i dla *i=1,…,j* są takie, że $g(x_i)=y_k$. Wtedy

$$P_Y(y_k)=\sum_{i=1}^{j}P_X(x_i)$$

Przykład:
S_X={-2,0,2} P(X=-2)=1/4 P(X=0)=1/4 P(X=2)=1/2
$Y=X^2$
S_Y={0,4} P(Y=0)=P(X=0)=1/4
P(Y=4)=P(X=-2)+P(X=2)=1/4+1/2=3/4

Transformacje zmiennych losowych

- Zmienna X jest ciągła, g(x) jest ciągłe
  Wtedy również Y jest ciągłe.
- I sposób (przez dystrybuantę)
  Szukamy $F_Y(y)$ a z niej liczymy $f_Y(y)$
- II sposób

$$\boxed{f_Y(y)=f_X(x)\cdot\left|\frac{dx}{dy}\right|}$$

Ogólny przypadek – sumujemy po przedziałach monotoniczności g(X)

$$\boxed{f_Y(y)=\sum_k f_{X,(k)}(x)\cdot\left|\frac{dx_{(k)}}{dy}\right|}$$

Przykłady: $Y=X^2$, $Y=X^4$

Przykład $Y=X^2$

Zał: X ma rozkład jednorodny na przedziale $x\in(0,2)$

$$f_X(x)=\begin{cases}\frac{1}{2} & dla\ x\in(0,2)\\ 0 & dla\ x\notin(0,2)\end{cases}\quad\rightarrow\quad F_X(x)=\begin{cases}0 & dla\ x\in(-\infty,0)\\ \int_0^x\frac{1}{2}dt=\frac{1}{2}t\Big|_0^x=\frac{1}{2}x & dla\ x\in[0,2)\\ 1 & dla\ x\in[2,+\infty)\end{cases}$$

$Y=X^2$ $y\in(0,4)$

- I sposób (przez dystrybuantę)

$$F_Y(y)=P(Y\le y)=P(X^2\le y)=P\left(-\sqrt{y}\le X\le\sqrt{y}\right)=P\left(0\le X\le\sqrt{y}\right)=$$

$$=F_X\left(\sqrt{y}\right)-F_X(0)=\frac{1}{2}\sqrt{y}-0=\frac{1}{2}\sqrt{y}$$

$$f_Y(y)=\frac{d}{dy}F_Y(y)=\frac{d}{dy}\left(\frac{1}{2}\sqrt{y}\right)=\frac{1}{4\sqrt{y}}\qquad y\in(0,4)$$

- II sposób (ze wzoru $f_Y(y)=f_X(x)\left|\frac{dx}{dy}\right|$)

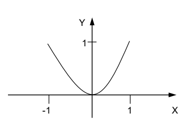
$$Y=X^2\quad\rightarrow\quad X=\sqrt{Y}\quad\rightarrow\quad\frac{dx}{dy}=\frac{1}{2\sqrt{Y}}$$

$$\rightarrow\quad f_Y(y)=\frac{1}{2}\left|\frac{1}{2\sqrt{Y}}\right|=\frac{1}{4\sqrt{Y}}$$

Spr: $N=\int_0^4\frac{1}{4\sqrt{y}}dy=\frac{1}{4}\int_0^4 y^{-1/2}dy=\frac{1}{4}\cdot 2\sqrt{y}\Big|_0^4=\frac{1}{4}\cdot 2\sqrt{4}=1$

Przykład $Y=X^4$

Zał: X ma rozkład $f_X(x)=\begin{cases}\frac{-1}{2}x+\frac{1}{2} & dla\ x\in[-1,1]\\ 0 & dla\ x\notin[-1,1]\end{cases}$



$Y=X^4$ $y\in[0,1]$

- Tylko II sposób ($f_Y(y)=\sum_k f_{X,(k)}(x)\left|\frac{dx_{(k)}}{dy}\right|$)

a) $x\in[-1,0)$

$$Y=X^4\quad\rightarrow\quad X=-Y^{1/4}\quad y\in(0,1]$$

$$\frac{dx}{dy}=\frac{-1}{4}y^{-3/4}$$

$$f_{Y,(I)}(y)=f_X(x)\cdot\left|\frac{dx}{dy}\right|=\left(\frac{-1}{2}x+\frac{1}{2}\right)\left|\frac{-1}{4}y^{-3/4}\right|=\left(\frac{-1}{2}\left(-y^{1/4}\right)+\frac{1}{2}\right)\cdot\frac{1}{4}y^{-3/4}=\frac{1}{8}\left(\sqrt[4]{y}+1\right)y^{-3/4}$$

b) $x\in(0,1]$

$$Y=X^4\quad\rightarrow\quad X=+Y^{1/4}\quad y\in(0,1)$$

$$\frac{dx}{dy}=\frac{1}{4}y^{-3/4}$$

$$f_{Y,(II)}(y)=f_X(x)\cdot\left|\frac{dx}{dy}\right|=\left(\frac{-1}{2}x+\frac{1}{2}\right)\left|\frac{1}{4}y^{-3/4}\right|=\left(\frac{-1}{2}\left(y^{1/4}\right)+\frac{1}{2}\right)\cdot\frac{1}{4}y^{-3/4}=\frac{1}{8}\left(-\sqrt[4]{y}+1\right)y^{-3/4}$$

c) $x=0$

Łącznie $f_Y(y)=f_{Y,(I)}(y)+f_{Y,(II)}(y)=\frac{1}{8}\left(\sqrt[4]{y}+1\right)y^{-3/4}+\frac{1}{8}\left(-\sqrt[4]{y}+1\right)y^{-3/4}=\frac{1}{8}\cdot 2\cdot y^{-3/4}=\frac{1}{4}y^{-3/4}$

Dyskretne rozkłady prawdopodobieństwa

- **Rozkład zero-jedynkowy (rozkład Bernoulliego)** o parametrze p

  Zmienna losowa to ilość sukcesów w jednokrotnym powtórzeniu eksperymentu w którym możliwe są tylko dwa wyniki
  S_X={0,1}
  Rozkład prawdopodobieństwa $P(x)=\begin{cases}1-p & dla\ x=0\\ p & dla\ x=1\end{cases}$

  zatem łącznie $P(X=x)=p^x(1-p)^{1-x}$ dla x=0,1
  Dystrybuanta

$$F_X(x)=\begin{cases}0 & dla\ x<0\\ 1-p & dla\ 0\le x<1\\ 1 & dla\ x\ge 1\end{cases}$$

Dyskretne rozkłady prawdopodobieństwa

- **Rozkład dwupunktowy** o parametrze p
  Zmienna losowa przyjmuje dwie dowolne wartości, jedną z prawdopodobieństwem p, drugą z prawdopdobieństwem 1-p.

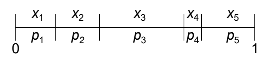
- **Rozkład wielopunktowy** o parametrach p_i, i=1,2,...,n
  przy czym $p_1+p_2+\ldots+p_n=1$.
  Zmienna losowa przyjmuje skończoną (n) ilość dyskretnych wartości.

- **Rozkład wielopunktowy** o parametrach p_i, i=1,2,...
  przy czym $p_1+p_2+\ldots=1$.
  Zmienna losowa przyjmuje nieskończoną ilość dyskretnych wartości

Dyskretne rozkłady prawdopodobieństwa – realizacja numeryczna

- 1. Odwracanie dystrybuanty (przykład w przyszłości)
- 2. Syt. Dysponujemy generatorem liczb pseudolosowych z przedziału (0,1)
- Dla rozkładu ze skończoną (=n) liczbą wartości zmiennej losowej X. Przedział (0,1) dzielimy na n przedziałów o długości p_i :

$$(0,p_1]\cup(p_1,p_1+p_2]\cup\ldots\cup\left(\sum_{k=1}^{n-1}p_k,1\right)$$



Losujemy liczbę Y $\in$(0,1) i znajdujemy dla niej przedział: $Y\in\left(\sum_{k=1}^{j-1}p_k,p_j\right]$

Jako wylosowaną wartość zmiennej losowej X przyjmujemy x_j.

Dyskretne rozkłady prawdopodobieństwa – realizacja numeryczna

- Syt. Dysponujemy generatorem liczb pseudolosowych z przedziału (0,1)
- Dla rozkładu z nieskończoną liczbą wartości zmiennej losowej X. Wybieramy ε (bardzo małe), p_i (jak największe) i n_{max}:

$$\sum_{k=1}^{n_{max}}p_k=1-\varepsilon \qquad (\varepsilon>0,\varepsilon\approx 0)$$

Przedział (0,1-ε) dzielimy na n_{max} przedziałów o długości p_i :

$$(0,p_1]\cup(p_1,p_1+p_2]\cup\ldots\cup\left(\sum_{k=1}^{n_{max}-1}p_k,1-\varepsilon\right)$$

Losujemy liczbę Y $\in$(0,1).
Jeżeli Y $\in$(0,1-ε) to znajdujemy dla niej przedział: $Y\in\left(\sum_{k=1}^{j-1}p_k,p_j\right]$

i jako wylosowaną wartość zmiennej losowej X przyjmujemy x_j.
Jeżeli Y $\in$[1-ε,1) to dodajemy przedziały powyżej n_{max} tak długo, aż znajdziemy przedział do którego należy Y.

Próba Bernoulliego i rozkład dwumienny

- **Próba Bernoulliego** to sekwencja powtórzeń n razy **tego samego** eksperymentu losowego, w wyniku którego możemy uzyskać jeden spośród dwóch wyników, zwanych jako „sukces" i „porażka".
  Muszą być spełnione dwa warunki:
  1. Powtórzenia są niezależne
  2. Prawdopodobieństwo sukcesu jest takie samo we wszystkich powtórzeniach.

  Przykład: n-krotny rzut monetą.
- **Rozkład dwumienny** (**dwumianowy**) (zwany w Polsce **rozkładem Bernoulliego**)
  Zmienna losowa X oznacza liczbę osiągniętych sukcesów w n-elementowej próbie Bernoulliego.
  S_X={0,1,2,…,n}

  Rozkład prawdopodobieństwa zmiennej X, przyjmującej wartości k$\in S_X$ jest postaci

$$\boxed{P_X(k)=\binom{n}{k}p^k(1-p)^{n-k}}$$

Rozkład dwumienny

Uzasadnienie:
Niech zdarzenie S oznacza osiągnięcie sukcesu w pojedynczym eksperymencie.
Niech zdarzenie P oznacza osiągnięcie porażki w pojedynczym eksperymencie.
Wynik próby (zdarzenie A) opisujemy jako n-elementowy ciąg wyników pojedynczych eksperymentów
np. A={S,P,S,S,…,P,P,S}
Prawdopodobieństwo uzyskania pojedynczego ciągu w którym wystąpiło k sukcesów to $P(A)=p^k(1-p)^{n-k}$ (niezależność prób)

Ilość ciągów sprzyjających wystąpieniu k sukcesów to $\binom{n}{k}$ (kombinacja bez powtórzeń)
Zatem (rozłączność ciągów)

$$P_X(k)=\binom{n}{k}P(A)=\binom{n}{k}p^k(1-p)^{n-k}$$

Rozkład dwumienny

- Rozkład ten jest poprawnie unormowany:

$$\sum_{k=0}^{n}P_X(k)=\sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k}=(p+(1-p))^n=1^n=1$$