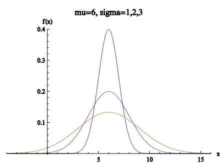
Poprzedni wykład

- Rozkład jednorodny
- Rozkład wykładniczy
- Rozkład Weibulla
- Rozkład normalny N(μ,σ^2)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad x \in R$$

- problem z dystrybuantą
- ważność wynika z CTG
- standaryzacja
- specjalne metody generowania



RPiS 2024/2025 1

1

Rozkład normalny -własności

- Wartość oczekiwana i wariancja: $E(X) = \mu \qquad var(X) = \sigma^2$
- **Standaryzacja dowolnego rozkładu** pozwala przechodzić pomiędzy rozkładami normalnymi:
  Jeżeli zmienna losowa X ma rozkład N(μ,σ^2) to zmienna losowa Y=(X-μ)/σ ma rozkład N(0,1) (dowód na ćwiczeniach).
- **„Reguła trzech sigma”** :

k	P(μ−kσ<X<μ+kσ)
1	0.6826895
2	0.9544997
3	**0.9973002**
4	0.9999367
5	0.9999994

P(μ−kσ<X<μ+kσ)	k
0.50	0.6745
0.80	1.2816
0.90	1.6449
0.95	1.96
0.99	2.5758

k	Surprisal #bits S=log$_2$(1/P) dla P=0.5 (rzut monetą)
1	1.66
3	**8.53**
5	20.67

RPiS 2024/2025 2

2

Rozkład normalny -własności

- Szerokość połówkowa FWHM (full width at half maximum)

$$f_X(x_{1/2}) = 0.5 \cdot f_X^{\max} = 0.5 \cdot f_X(\mu) = \frac{1}{2\sigma\sqrt{2\pi}}$$

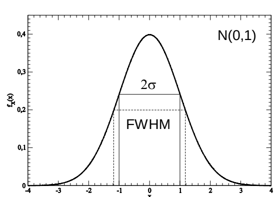
$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\mu - x_{1/2})^2}{2\sigma^2}} = \frac{1}{2\sigma\sqrt{2\pi}}$$

$$x_{1/2}^{(\pm)} = \mu \pm \sigma\sqrt{2\ln(2)}$$

$$FWHM \equiv \Gamma = x_{1/2}^{(+)} - x_{1/2}^{(-)} =$$

$$= 2\sigma\sqrt{2\ln(2)} \approx 2.355\sigma$$

$$\rightarrow \sigma = \frac{\Gamma}{2.355} = 0.425\Gamma$$



RPiS 2024/2025 3

3

Rozkład Pareto

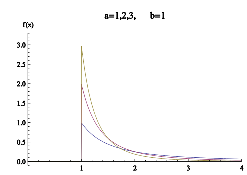
$$f_X(x) = \frac{ab^a}{x^{a+1}} \qquad x > b, a > 0, b > 0$$

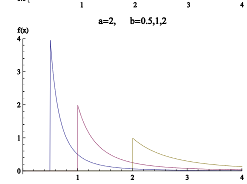
- Parametr b opisuje położenie, a-kształt
- Wartość oczekiwana i wariancja

$$E(X) = \frac{ab}{a-1} \qquad dla\ a > 1$$

$$var(X) = \frac{ab^2}{(a-1)^2(a-2)} \qquad dla\ a > 2$$

- Zastosowanie: rozkład dochodów, rozkład długości plików, wielkość cząstek piasku, czasopisma w bibliotece (rozkład Bradforda)
- Zasada Pareto: np. 20% populacji ma 80% bogactw (wartości zależą od parametrów)
- Pojawia się przy grupowaniu w rangi





RPiS 2024/2025 4

4

Rozkład Gamma

$$f_X(x) = \frac{(\lambda x)^{\alpha-1}\lambda e^{-\lambda x}}{\Gamma(\alpha)} \qquad x \geq 0, \lambda > 0, \alpha > 0 \quad \Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

- Parametr λ opisuje położenie i wielkość, parametr α opisuje kształt.
- Wartość oczekiwana i wariancja

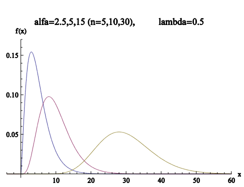
$$E(X) = \frac{\alpha}{\lambda} \qquad var(X) = \frac{\alpha}{\lambda^2}$$

- Z tego rozkładu wyprowadza się inne rozkłady, np:

dla α=1 $\quad f_X(x) = \frac{(\lambda x)^{1-1}\lambda e^{-\lambda x}}{\Gamma(1)} = \lambda e^{-\lambda x}$

dla α=n/2, λ=1/2 $\quad f_X(x) = \frac{\left(\frac{x}{2}\right)^{n/2-1} e^{-\frac{x}{2}}}{2\Gamma(n/2)}$

Jest to rozkład $\chi_n^2(x)$ „Chi-kwadrat o *n* stopniach swobody”



RPiS 2024/2025 5

5

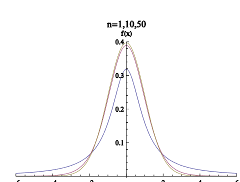
Rozkład t-Studenta

$$f_X(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{(n+1)}{2}\right)}{\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{(n+1)}{2}}$$

- Parametr n=1,2,… opisuje kształt.
- Wartość oczekiwana (dla n>1) i wariancja (dla n>2):

$$E(X) = 0 \qquad var(X) = \frac{n}{n-2}$$

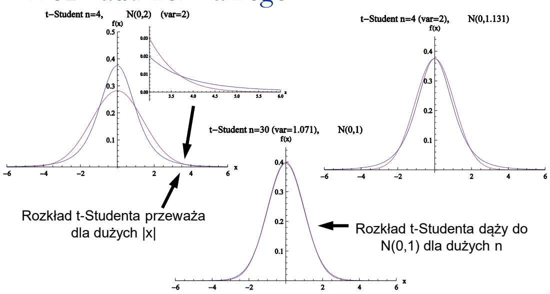
- Dla n=1 jest to rozkład Cauchy
- Dla n>30 dobrym przybliżeniem jest rozkład normalny N(0,n/(n-2))
- Generowanie:
  Jeżeli zmienna losowa X ma rozkład N(0,1), a zmienna losowa Y ma rozkład $\chi_n^2(y)$ to $Z = X\sqrt{n/Y}$ ma rozkład t-Studenta o n stopniach swobody.



RPiS 2024/2025 6

6

Porównanie rozkładu t-Studenta i rozkładu normalnego



7

Wielowymiarowe zmienne losowe (wektory losowe)

- Syt. Eksperyment losowy E, związana z nim przestrzeń próbek S
- **n-wymiarowy wektor losowy** $\vec{X}$ to funkcja, która każdemu elementowi s należącemu zbioru S przyporządkowuje wektor n liczb rzeczywistych (X_1(s), X_2(s),…,X_n(s)) (n zmiennych losowych),
- Np. Losujemy punkt na płaszczyźnie $X_1(s)=x, X_2(s)=y$
- Nie ma związku pomiędzy wymiarem elementu w S (czyli liczbą zmiennych potrzebnych do opisania wyniku eksperymentu) a wymiarem wektora losowego $\vec{X}$.

  np. $\vec{X} = (X_1, X_2, X_3, X_4)$:

  $$X_1(s) = x_a, X_2(s) = y_a, X_3(s) = x_a + y_a, X_4(s) = sign(x_a)$$
- W dalszym ciągu ograniczymy się do n=2

8

Łączny rozkład prawdopodobieństwa

Syt. $\vec{Z} = (X, Y)$ $\qquad S_{\vec{Z}} = \{(x_i, y_j)\}$ $\qquad i = 1,2,\ldots \quad j = 1,2,\ldots$

- **Łącznym rozkładem prawdopodobieństwa** nazywamy (dwuargumentową) funkcję, która

  $$\forall i, j: \quad P_{X,Y}(x_i, y_j) = P[\{X = x_i\} \cap \{Y = y_j\}] \equiv P[X = x_i, Y = y_j]$$
- Własności:

  $$\forall i, j \quad P_{X,Y}(x_i, y_j) \ge 0$$

  $$\sum_i \sum_j P_{X,Y}(x_i, y_j) = 1$$

  $$P(A) = \sum_{(x_i, y_j) \in A} P_{X,Y}(x_i, y_j)$$
- **Łączna dystrybuanta**

  $$F_{X,Y}(x, y) = P[X \le x, Y \le y] = \sum_{x_i \le x} \sum_{y_j \le y} P_{X,Y}(x_i, y_j)$$

9

Brzegowy rozkład prawdopodobieństwa

- **Brzegowym rozkładem prawdopodobieństwa zmiennej X** nazywamy (jednoargumentową) funkcję $P_X(x_i)$:

  $$P_X(x_i) \equiv P(X = x_i) = \sum_j P(\{X = x_i\} \cap \{Y = y_j\}) = \sum_j P_{X,Y}(x_i, y_j)$$
- Analogicznie: $P_Y(y_j) = \sum_i P_{X,Y}(x_i, y_j)$
- Przykład: dwukrotny rzut monetą.

  X – liczba orłów w pierwszym rzucie

  Y – liczba orłów w dwóch rzutach

x_i	0	1
$P_X(x_i)$	1/2	1/2

y_j	0	1	2
$P_Y(y_j)$	1/4	1/2	1/4

y_j \ x_i	0	1	$P_Y(y_j)$
0	**1/4**	**0**	=1/4
1	**1/4**	**1/4**	=1/2
2	**0**	**1/4**	=1/4
$P_X(x_i)$	=1/2	=1/2	=1

10

Łączna funkcja gęstości prawdopodobieństwa

Syt. $\vec{Z} = (X, Y)$, X,Y – ciągłe zmienne losowe

- **Łączną funkcją gęstości prawdopodobieństwa pary (X,Y)** nazywamy funkcję, która wiąże się z prawdopodobieństwem zdarzenia A poprzez:

  $$P(A) = \iint_A f_{X,Y}(x, y)\, dx\, dy$$
- Własności:

  $$\forall x, y \quad f_{X,Y}(x, y) \ge 0$$

  $$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x, y)\, dx\, dy = 1$$

  $$P(x < X \le x + dx, y < Y \le y + dy) = f_{X,Y}(x, y)\, dx\, dy$$

11

Łączna dystrybuanta

- **Łączna dystrybuanta**

  $$F_{X,Y}(x, y) = P[X \le x, Y \le y] = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, du\, dv$$
- Własności

1. $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$
2. $F_{X,Y}(+\infty, +\infty) = 1$
3. $dla\ x_1 \le x_2\ i\ y_1 \le y_2 \quad F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$
4. $\lim_{\delta \to 0^+} F_{X,Y}(x + \delta, y) = \lim_{\delta \to 0^+} F_{X,Y}(x, y + \delta) = F_{X,Y}(x, y)$
5. $f_{X,Y}(x, y) = \dfrac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$

12

Łączna dystrybuanta

- Własności

6. $P[a < X \le b, c < Y \le d] = F_{X,Y}(b,d) - F_{X,Y}(b,c) - F_{X,Y}(a,d) + F_{X,Y}(a,c)$

Dowód:

$$P[a < X \le b, c < Y \le d] = \int_a^b \int_c^d f_{X,Y}(x,y)\,dy\,dx =$$

$$= \int_{-\infty}^{b} \int_c^d f_{X,Y}(x,y)\,dy\,dx - \int_{-\infty}^{a} \int_c^d f_{X,Y}(x,y)\,dy\,dx =$$

$$= \int_{-\infty}^{b} \int_{-\infty}^{d} f_{X,Y}(x,y)\,dy\,dx - \int_{-\infty}^{b} \int_{-\infty}^{c} f_{X,Y}(x,y)\,dy\,dx +$$

$$- \int_{-\infty}^{a} \int_{-\infty}^{d} f_{X,Y}(x,y)\,dy\,dx + \int_{-\infty}^{a} \int_{-\infty}^{c} f_{X,Y}(x,y)\,dy\,dx =$$

$$= F_{X,Y}(b,d) - F_{X,Y}(b,c) - F_{X,Y}(a,d) + F_{X,Y}(a,c)$$