

Rachunek Prawdopodobieństwa

Anna Janicka

wykład X, 5.12.2018

**NIEZALEŻNOŚĆ ZMIENNYCH LOSOWYCH
REGRESJA LINIOWA**

Plan na dzisiaj

17. Niezależne zmienne losowe

18. Regresja liniowa



WARSAW UNIVERSITY
Faculty of Economic Sciences

17. Niezależne zmienne losowe

1. Definicja niezależności

Zmienne losowe $X_1, \dots, X_n: \Omega \rightarrow \mathbb{R}$ nazywamy **niezależnymi**, jeśli dla dowolnego ciągu $B_1, B_2, \dots, B_n$ zbiorów borelowskich zachodzi

$$\begin{aligned} \mathbb{P}(X_1 \in B_1, X_2 \in B_2, \dots, X_n \in B_n) \\ = \mathbb{P}(X_1 \in B_1) \mathbb{P}(X_2 \in B_2) \dots \mathbb{P}(X_n \in B_n). \end{aligned}$$

- 2. Niezależność zmiennych losowych a niezależność zdarzeń
 - 3. Niezależność parami a niezależność łączna
-



17. Niezależne zmienne losowe – cd.

4. Badanie niezależności dyskretnych zm.

Niech $X_1, X_2, \dots, X_n$ będą dyskretnymi zmiennymi losowymi, zaś S_{X_i} takimi zbiorami przeliczalnymi, że $\mathbb{P}(X_i \in S_{X_i}) = 1$. Wówczas zmienne $X_1, X_2, \dots, X_n$ są niezależne wtedy i tylko wtedy, gdy dla każdego ciągu $x_1, x_2, \dots, x_n$ takiego, że $x_i \in S_{X_i}$, $i = 1, 2, \dots, n$, zachodzi równość

$$\begin{aligned} \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \\ = \mathbb{P}(X_1 = x_1) \mathbb{P}(X_2 = x_2) \cdot \dots \cdot \mathbb{P}(X_n = x_n). \end{aligned}$$

5. Przykład



17. Niezależne zmienne losowe – cd(2)

6. Badanie niezależności ciągłych zm.

Niech $X_1, X_2, \dots, X_n: \Omega \rightarrow \mathbb{R}$ będą ciągłymi zmiennymi losowymi o gęstościach $g_1, g_2, \dots, g_n$, odpowiednio.

Wówczas zmienne $X_1, X_2, \dots, X_n$ są niezależne wtedy i tylko wtedy, gdy funkcja $g: \mathbb{R}^n \rightarrow [0, \infty)$ dana wzorem $g(x_1, x_2, \dots, x_n) = g_1(x_1)g_2(x_2) \cdot \dots \cdot g_n(x_n)$, jest gęstością rozkładu $\mu_{(X_1, X_2, \dots, X_n)}$.

7. Przykłady

- rozkład jednostajny na kwadracie $\{\pm 1, \pm 1\}$
 - rozkład jednostajny na kole
 - inny rozkład
-



17. Niezależne zmienne losowe – cd (3)

8. Niezależność funkcji zmiennych niezal.

*Rozważmy zmienne losowe $X_{1,1}, X_{1,2}, \dots, X_{1,k_1},$
 $X_{2,1}, X_{2,2}, \dots, X_{2,k_2}, \dots, X_{n,1}, X_{n,2}, \dots, X_{n,k_n}$
oraz funkcje borelowskie $\varphi_i: \mathbb{R}^{k_i} \rightarrow \mathbb{R}, i = 1, 2, \dots, n.$
Jeśli powyższe zmienne są niezależne, to również zmienne*

$$Y_1 = \varphi_1(X_{1,1}, X_{1,2}, \dots, X_{1,k_1}),$$

$$Y_2 = \varphi_2(X_{2,1}, X_{2,2}, \dots, X_{2,k_2}),$$

...

$$Y_n = \varphi_n(X_{n,1}, X_{n,2}, \dots, X_{n,k_n})$$

są niezależne.



17. Niezależne zmienne losowe – cd. (2)

9. Charakterystyki niezależnych zm. los.

Niech $X_1, X_2, \dots, X_n$ będą niezależnymi zmiennymi losowymi, posiadającymi wartość oczekiwaną. Wówczas zmienna $X_1 \cdot X_2 \cdot \dots \cdot X_n$ również ma wartość oczekiwaną i zachodzi równość

$$\mathbb{E}(X_1 X_2 \cdot \dots \cdot X_n) = \mathbb{E}X_1 \cdot \mathbb{E}X_2 \cdot \dots \cdot \mathbb{E}X_n.$$

10. Przykład

11. Kowariancja niezależnych zm. los.

Jeśli X, Y są niezależnymi zmiennymi losowymi takimi, że zachodzi $\mathbb{E}|XY| < \infty$, to $\text{Cov}(X, Y) = 0$. Jeśli dodatkowo zmienne X, Y mają skończoną, niezerową wariancję, to zeruje się także ich współczynnik korelacji, $\rho(X, Y) = 0$.

17. Niezależne zmienne losowe – cd. (3)

12. Nieskorelowane zmienne losowe

Zmienne losowe X, Y nazwiemy nieskorelowanymi, jeżeli $\rho(X, Y) = 0$.

13. Niezależność $\Rightarrow$ nieskorelowanie, ale

$\Leftarrow$ **NIE ZACHODZI!**

14. Przykład – rozkład jednostajny na kole



17. Niezależne zmienne losowe – cd. (4)

15. Wariancja sumy nzał zm. los.

Niech $X_1, X_2, \dots, X_n$ będą ~~niezależnymi~~ zmiennymi losowymi o skończonej wariancji. Wówczas zmienna $X_1 + X_2 + \dots + X_n$ także ma skończoną wariancję oraz

$$\text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) + 2 \sum_{i < j} \text{Cov}(X_i, X_j).$$

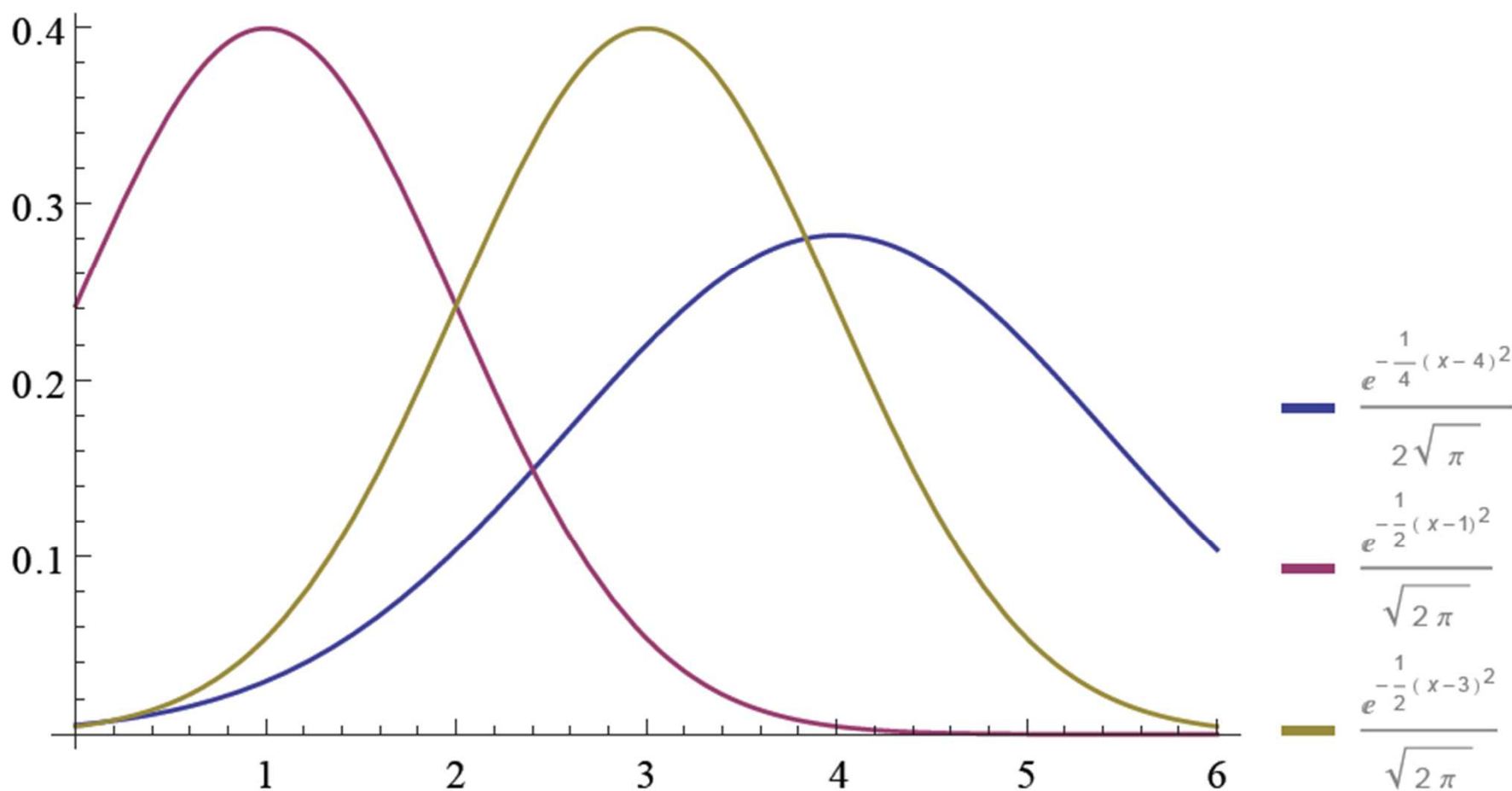
16. Przykład – suma oczek w n rzutach

17. Gęstość sumy – splot funkcji gęstości

Założmy, że X, Y są niezależnymi zmiennymi losowymi o rozkładach ciągłych z gęstościami g_X, g_Y . Wówczas zmienna $X + Y$ ma rozkład z gęstością będącą splotem funkcji g_X i g_Y , tzn.

$$g_{X+Y}(x) = g_X * g_Y(x) = \int_{\mathbb{R}} g_X(y) g_Y(x - y) dy = \int_{\mathbb{R}} g_X(x - y) g_Y(y) dy.$$


Suma niezależnych zmiennych losowych



Wielowymiarowy rozkład normalny

1. Standardowy rozkład dwuwymiarowy

$$g(x, y) = \frac{1}{2\pi} \exp \left[-\frac{1}{2}(x^2 + y^2) \right].$$

2. Ogólny o średniej $m = (m_1, m_2)$ i macierzy kowariancji $Q = A^{-1}$, gdzie

$$g(x, y) = \frac{\sqrt{\det A}}{2\pi} \cdot \exp \left(-\frac{1}{2} (a_{11}(x - m_1)^2 + 2a_{12}(x - m_1)(y - m_2) + a_{22}(y - m_2)^2) \right)$$
$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$



$$\langle A(x - m_1, y - m_2), (x - m_1, y - m_2) \rangle$$

Wielowymiarowy rozkład normalny – cd.

3. Definicja ogólna

$$g(x) = \frac{\sqrt{\det A}}{(2\pi)^{n/2}} \exp \left(-\frac{\langle A(x - m), (x - m) \rangle}{2} \right), \quad x \in \mathbb{R}^n,$$

4. Własności

- jednoznacznie wyznaczony przez średnią i macierz kowariancji
- każdy rozkład to afiniczne przekształcenie standardowej zmiennej; każde afiniczne przekształcenie to taki rozkład



Wielowymiarowy rozkład normalny – cd(2).

5. Nieskorelowane zmienne normalne

Założmy, że $X = (X_1, X_2, \dots, X_n)$ ma rozkład normalny i zmienne $X_1, X_2, \dots, X_n$ są nieskorelowane.

Wówczas są one niezależne.



18. Regresja liniowa

Szukamy liniowego przybliżenia zmiennej Y zmienną X , postaci $aX + b$. Przybliżenie ma być najlepsze możliwe w sensie błędu średniokwadratowego, tj. minimalizujące

$$f(a, b) = \mathbb{E}(Y - aX - b)^2$$

Wówczas, o ile odp. wyrażenia istnieją i mają sens:

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}X} \quad b = \mathbb{E}Y - \mathbb{E}X \cdot \frac{\text{Cov}(X, Y)}{\text{Var}X}$$



$$a = \rho_{X,Y} \cdot \frac{\sigma_Y}{\sigma_X}$$



WARSAW UNIVERSITY
Faculty of Economic Sciences