

Metody probabilistyczne i statystyka - wykład czwarty¹

dr Jarosław Kotowicz

Instytut Informatyki Uniwersytet w Białymstoku

7 marca 2020

¹©J.Kotowicz, 2020

Spis treści

- 1 **Zmienne losowe wielowymiarowe**
- 2 Dystrybuanta wielowymiarowej zmiennej losowej.
- 3 Rozkłady ciągłe i dyskretne
- 4 Rozkłady brzegowe
- 5 Niezależne zmienne losowe

Pojęcie wielowymiarowej zmiennej losowej. I

Niech (Ω, Σ, P) będzie ustaloną przestrzenią probabilistyczną i $d \in \mathbb{N}$.

Definicja 1

Zmienną losową d -wymiarową nazywamy każde odwzorowanie mierzalne X z (Ω, Σ) w $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ tzn.

$$\forall A \in \mathcal{B}(\mathbb{R}^d) X^{-1}(A) \in \Sigma. \quad (1)$$

Lemat 1

Jeżeli odwzorowanie $X : \Omega \rightarrow \mathbb{R}^d$ spełniające warunek

$$\forall t_1, \dots, t_d \in \mathbb{R} X^{-1}(] - \infty, t_1] \times \dots \times] - \infty, t_d]) \in \Sigma \quad (2)$$

jest d -wymiarową zmienną losową.

Pojęcie wielowymiarowej zmiennej losowej. II

Twierdzenie 1

Niech X będzie d -wymiarową zmienną losową, zaś $\varphi: \mathbb{R}^d \rightarrow \mathbb{R}^m$ będzie funkcją borelowską. Wtedy $\varphi(X)$ jest m -wymiarową zmienną losową.

Rozkład prawdopodobieństwa na \mathbb{R}^d . I

Definicja 2

Rozkładem prawdopodobieństwa na \mathbb{R}^d nazywamy każdą miarę probabilistyczną μ na $\mathcal{B}(\mathbb{R}^d)$.

Definicja 3

Rozkładem prawdopodobieństwa d -wymiarowej zmiennej losowej X nazywamy rozkład prawdopodobieństwa μ określony na $\mathcal{B}(\mathbb{R}^d)$ zależnością

$$\forall B \in \mathcal{B}(\mathbb{R}^d) \quad \mu(B) := P(X^{-1}(B)). \quad (3)$$

Uwaga 1

Oznaczamy

$$P(X^{-1}(B)) = P(\{\omega : X(\omega) \in B\}) = P(\{X \in B\}).$$

Spis treści

- 1 Zmienne losowe wielowymiarowe
- 2 **Dystrybuanta wielowymiarowej zmiennej losowej.**
- 3 Rozkłady ciągłe i dyskretne
- 4 Rozkłady brzegowe
- 5 Niezależne zmienne losowe

Dystrybuanta wielowymiarowej zmiennej losowej. I

Niech (Ω, Σ, P) będzie ustaloną przestrzenią probabilistyczną. Niech ponadto $d \in \mathbb{N}$.

Definicja 4

Dystrybuantą rozkładu prawdopodobieństwa μ na \mathbb{R}^d nazywamy funkcję $F: \mathbb{R}^d \rightarrow \mathbb{R}$, określoną wzorem

$$F(t_1, \dots, t_d) := \mu([-\infty, t_1] \times \dots \times [-\infty, t_d]). \quad (4)$$

Uwaga 2

Jeżeli μ jest rozkładem d -wymiarowej zmiennej losowej X , to dystrybuantę rozkładu μ będziemy nazywać dystrybuantą zmiennej losowej X i wtedy

$$F(t_1, \dots, t_d) = P(\{X_1 \leq t_1 \wedge \dots \wedge X_d \leq t_d\}) \quad (5)$$

Dystrybuanta wielowymiarowej zmiennej losowej. II

Uwaga 3

W literaturze spotyka się również definicję dystrybuanty rozkładu z przedziałami otwartymi.

Twierdzenie 2

Dystrybuanta F rozkładu prawdopodobieństwa μ na \mathbb{R}^d ma następujące własności:

- i. *F jest niemalejąca względem każdego argumentu,*
- ii.
$$\lim_{\inf_{1 \leq i \leq d} x_i \rightarrow -\infty} F(x_1, \dots, x_d) = 0,$$
- iii.
$$\lim_{\inf_{1 \leq i \leq d} x_i \rightarrow +\infty} F(x_1, \dots, x_d) = 1,$$
- iv. *F jest prawostronnie ciągła względem każdego argumentu.*

Dystrybuanta wielowymiarowej zmiennej losowej. III

Twierdzenie 3

Dla dystrybuanty F rozkładu prawdopodobieństwa μ na \mathbb{R}^d zachodzi również

$$\forall 1 \leq k \leq d, x_k \leq y_k \Rightarrow \sum_{\varepsilon \in \{0,1\}^{\overline{1,d}}} (-1)^{\sum_{k=1}^d \varepsilon_k} F(U(\varepsilon)) \geq 0, \quad (6)$$

gdzie $U(\varepsilon) = (\varepsilon_1 x_1 + (1 - \varepsilon_1) y_1, \dots, \varepsilon_d x_d + (1 - \varepsilon_d) y_d)$.

Twierdzenie 4

Jeżeli funkcja $F: \mathbb{R}^d \rightarrow \mathbb{R}$ spełnia warunki twierdzeń 2 i 3, to istnieje taki rozkład prawdopodobieństwa μ na \mathbb{R}^d , że F jest jego dystrybuantą.

Dystrybuanta wielowymiarowej zmiennej losowej. IV

Twierdzenie 5

Jeżeli μ i ν są rozkładami prawdopodobieństwa na \mathbb{R}^d o dystrybuantach F i G oraz $F = G$, to dla dowolnego zbioru $A \in \mathcal{B}(\mathbb{R}^d)$ zachodzi $\mu(A) = \nu(A)$.

Spis treści

- 1 Zmienne losowe wielowymiarowe
- 2 Dystrybuanta wielowymiarowej zmiennej losowej.
- 3 Rozkłady ciągłe i dyskretne**
- 4 Rozkłady brzegowe
- 5 Niezależne zmienne losowe

Rozkłady ciągłe. I

Niech (Ω, Σ, P) będzie ustaloną przestrzenią probabilistyczną. Niech ponadto $d \in \mathbb{N}$.

Definicja 5

Jeżeli μ jest rozkładem prawdopodobieństwa na \mathbb{R}^d i dla pewnej funkcji $f : \mathbb{R}^d \rightarrow \mathbb{R}$ całkownej w sensie Lebesgue'a mamy

$$\forall A \in \mathcal{B}(\mathbb{R}^d) \mu(A) = \int_A f(x) dm(x), \quad (7)$$

to f nazywamy gęstością rozkładu μ .

Definicja 6

Jeżeli dla rozkładu prawdopodobieństwa na \mathbb{R}^d μ istnieje gęstość, to taki rozkład nazywamy ciągłym.

Rozkłady ciągłe. II

Twierdzenie 6

Niech f będzie gęstością rozkładu prawdopodobieństwa μ na \mathbb{R}^d . Wtedy

- (i) $\int_{\mathbb{R}^d} f(x) dm(x) = \mu(\mathbb{R}^d) = 1$,
- (ii) $f \geq 0$ p.w.,
- (iii) Gęstość jest wyznaczona jednoznacznie z dokładnością do zbiorów miary Lebesgue'a zero.

Ponadto każda funkcja $f: \mathbb{R}^d \rightarrow \mathbb{R}$ spełniająca warunki (i) i (ii) jest gęstością.

Rozkłady dyskretne.

Definicja 7

Rozkład μ na \mathbb{R}^d nazywamy dyskretnym wtedy i tylko wtedy, gdy istnieje zbiór co najwyżej przeliczalny $S \subset \mathbb{R}^d$ dla którego $\mu(S) = 1$.

Spis treści

- 1 Zmienne losowe wielowymiarowe
- 2 Dystrybuanta wielowymiarowej zmiennej losowej.
- 3 Rozkłady ciągłe i dyskretne
- 4 Rozkłady brzegowe**
- 5 Niezależne zmienne losowe

Jednowymiarowy rozkład brzegowy. I

Lemat 2

Niech μ będzie rozkładem prawdopodobieństwa na \mathbb{R}^d . Określmy dla dowolnego $B \in \mathcal{B}(\mathbb{R})$ i $1 \leq j \leq n$ funkcję

$$\mu_j(B) := \mu(\mathbb{R} \times \dots \times \mathbb{R} \times \underbrace{B}_j \times \mathbb{R} \times \dots \times \mathbb{R}). \quad (8)$$

Wtedy μ_j jest rozkładem prawdopodobieństwa na \mathbb{R} .

Definicja 8

Funkcję μ_j występującą w lemacie 2 nazywamy jednowymiarowym rozkładem brzegowym.

Jednowymiarowy rozkład brzegowy. II

Twierdzenie 7

Rozkład prawdopodobieństwa μ na \mathbb{R}^d jest dyskretny wtedy i tylko wtedy, gdy dla każdego $1 \leq j \leq n$ rozkład μ_j jest dyskretny.

Twierdzenie 8

Jeżeli rozkład prawdopodobieństwa μ na \mathbb{R}^d jest ciągły, to wszystkie jednowymiarowe rozkłady brzegowe są ciągłe.

Przykład 1

Niech $A = \{(x, y) : x = y\} \cap [0, 1]^2$. Niech μ będzie rozkładem prawdopodobieństwa na \mathbb{R}^2 określonym następująco $\mu(A) = 1$. Wówczas rozkłady brzegowe są ciągłe, lecz μ nie jest rozkładem ciągłym.

Wielowymiarowy rozkład brzegowy.

Lemat 3

Niech μ będzie rozkładem prawdopodobieństwa na \mathbb{R}^d . Niech $1 \leq k \leq n$ i niech $1 \leq i_1 < \dots < i_k \leq n$ będzie k -wyrazowym ciągiem indeksów. Określmy dla dowolnych $B_i \in \mathcal{B}(\mathbb{R})$ dla $i = 1, \dots, k$

$$\mu_{i_1, \dots, i_k}(B_1 \times \dots \times B_k) := \mu(\mathbb{R} \times \dots \times \mathbb{R} \times \underbrace{B}_{i_1} \times \mathbb{R} \times \dots \times \mathbb{R} \times \underbrace{B}_{i_k} \times \mathbb{R} \times \dots \times \mathbb{R}). \quad (9)$$

Wtedy μ_{i_1, \dots, i_k} jest rozkładem prawdopodobieństwa na \mathbb{R}^k .

Definicja 9

Funkcję μ_{i_1, \dots, i_k} występującą w lemacie 3 nazywamy k -wymiarowym rozkładem brzegowym.

Spis treści

- 1 Zmienne losowe wielowymiarowe
- 2 Dystrybuanta wielowymiarowej zmiennej losowej.
- 3 Rozkłady ciągłe i dyskretne
- 4 Rozkłady brzegowe
- 5 **Niezależne zmienne losowe**

Pojęcie niezależności. Podstawowe własności. I

Dana jest przestrzeń probabilistyczna (Ω, Σ, P) oraz jednowymiarowe zmienne losowe X_1, \dots, X_n określone na niej.

Definicja 10

Mówimy, że zmienne losowe X_1, \dots, X_n są niezależne wtedy i tylko wtedy, gdy dowolnych zbiorów $\{A_1, \dots, A_n\} \subset \mathcal{B}(\mathbb{R})$ zachodzi równość

$$P(\{\omega : X_i(\omega) \in A_i \wedge i \in \overline{1, n}\}) = \prod_{i=1}^n P(\{\omega : X_i(\omega) \in A_i\}). \quad (10)$$

Pojęcie niezależności. Podstawowe własności. II

Twierdzenie 9

Dyskretne zmienne losowe X_1, \dots, X_n są niezależne wtedy i tylko wtedy, gdy dla dowolnego ciągu $(a_1, \dots, a_n) \in (W_{X_1} \times \dots \times W_{X_n})^{\overline{1,n}}$ zachodzi równość

$$P(\{\omega : X_i(\omega) = a_i \wedge i \in \overline{1,n}\}) = \prod_{i=1}^n P(\{\omega : X_i(\omega) = a_i\}). \quad (11)$$

Niezależność i wielowymiarowe zmienne losowe. I

Dana jest przestrzeń probabilistyczna (Ω, Σ, P) oraz jednowymiarowe zmienne losowe X_1, \dots, X_n określone na niej.

Twierdzenie 10

Dla zmiennych losowych X_1, \dots, X_n następujące warunki są równoważne

- i) zmienne są niezależne,*
- ii) $\mu_{(X_1, \dots, X_n)} = \mu_{X_1} \otimes \dots \otimes \mu_{X_n}$,*
- iii) dla dowolnego $t_1, \dots, t_n \in \mathbb{R}$ zachodzi równość*

$$F_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{i=1}^n F_{X_i}(t_i).$$

Niezależność i wielowymiarowe zmienne losowe. II

Twierdzenie 11

Niech dane będą zmienne losowe X_1, \dots, X_n o rozkładach ciągłych z gęstościami równymi odpowiednio g_1, \dots, g_n . Zmienne losowe są niezależne wtedy i tylko wtedy, gdy $\mu_{(X_1, \dots, X_n)}$ jest rozkładem ciągłym o gęstości

$$g(x_1, \dots, x_n) = \prod_{i=1}^n g_i(x_i). \quad (12)$$

Twierdzenie 12

Niech dane będą niezależne zmienne losowe X_1, \dots, X_n , dla których istnieją wartości oczekiwane. Wówczas istnieje wartość oczekiwana ich iloczynu oraz

$$\mathbb{E}(X_1 \cdot \dots \cdot X_n) = \prod_{i=1}^n \mathbb{E}(X_i). \quad (13)$$

Uogólnienie niezależności. I

Definicja 11

Zmienne losowe $(X_\alpha)_{\alpha \in \Lambda}$ określone na tej samej przestrzeni probabilistycznej nazywamy niezależnymi wtedy i tylko wtedy, gdy dla każdego skończonego ciągu indeksów i_1, \dots, i_n , gdzie $i \in \mathbb{N}$, zmienne losowe X_{i_1}, \dots, X_{i_n} są niezależne.

Twierdzenie 13

Niech $\{\mathcal{H}_t : t \in T\}$, $\mathcal{H}_t \subset \mathcal{H}$ będzie rodziną niezależnych σ -ciał i niech zbiór T będzie sumą parami rozłącznych zbiorów T_i , gdzie $i \in I$. Wtedy niezależna jest rodzina σ -ciał

$$\mathfrak{G}_i := \sigma \left(\bigcup_{t \in T_i} \mathcal{H}_t \right) \quad i \in I.$$

Uogólnienie niezależności. II

Twierdzenie 14

Założmy, że zmienne losowe $X_{1,1}, \dots, X_{1,k_1}, \dots, X_{n,1}, \dots, X_{n,k_n}$ są niezależne. Wówczas zmienne losowe $Y_j = \varphi_j(X_{j,1}, \dots, X_{j,k_j})$, gdzie $j = 1, \dots, n$ i φ_j – funkcje borelowskie takie, że Y_j są dobrze zdefiniowane, są niezależne.