Rachunek prawdopodobieństwa MAT1332

Wydział Matematyki, Matematyka Stosowana

Wykładowca: dr hab. Agnieszka Jurlewicz

Wykład 7: Rozkład łączny. Rozkłady brzegowe. Niezależność zmiennych losowych.

Wektor losowy, rozkład łączny, rozkłady brzegowe.

Definicja.

Wektor losowy to wektor, którego składowe są zmiennymi losowymi.
Np. (X, Y), gdzie X, Y to zmienne losowe.

Rozkład wektora losowego (X, Y) to funkcja $P((X, Y) \in C)$, gdzie C to borelowski podzbiór płaszczyzny $\mathbb{R}^2$. Nazywamy go **rozkładem łącznym** zmiennych losowych X, Y.

Rozkład zmiennej losowej X i rozkład zmiennej losowej Y nazywamy **rozkładami brzegowymi** wektora losowego (X, Y).

Pełna informacja o rozkładzie łącznym zmiennych losowych X, Y zawarta jest:

(a) w **dystrybuancie** tego rozkładu, czyli funkcji

$$F_{X,Y}(x, y) = P(X < x, Y < y)$$

(b) w przypadku **dyskretnego** wektora losowego (X, Y) zawarta jest także w **ciągu** trójek $\{(x_n, y_k, p_{nk}), n \in \mathbb{T}_1 \subset \mathbb{N}, k \in \mathbb{T}_2 \subset \mathbb{N}\}$, gdzie $\{x_n, n \in \mathbb{T}_1\}$ oraz $\{y_k, k \in \mathbb{T}_2\}$ to ciągi wszystkich wartości przyjmowanych odpowiednio przez X i Y z dodatnimi prawdopodobieństwami, natomiast $p_{nk} = P(X = x_n, Y = y_k)$, $\ n \in \mathbb{T}_1, k \in \mathbb{T}_2$.

(c) w przypadku **ciągłego** rozkładu wektora losowego (X, Y) zawarta jest także w **gęstości łącznej** $f(x, y)$, czyli takiej funkcji $f(x, y) \geqslant 0$ dla każdego (x, y), że

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} ds \int_{-\infty}^{y} f(s, t)dt$$

(d) w **funkcji charakterystycznej** wektora losowego (X, Y), czyli zespolonej funkcji

$$\varphi_{X,Y}(s, t) = \mathrm{E}e^{i(sX+tY)} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(sx+ty)} dF_{X,Y}(x, y)$$

Opracowanie: dr hab. Agnieszka Jurlewicz

FAKT: Jeśli znamy rozkład łączny, to znamy też rozkłady brzegowe, gdyż:

$$F_X(x) = P(X < x) = P(X < x, Y < \infty) = \lim_{y\to\infty} F_{X,Y}(x,y),$$
$$F_Y(y) = P(Y < y) = P(X < \infty, Y < y) = \lim_{x\to\infty} F_{X,Y}(x,y)$$

W przypadku **dyskretnego** wektora losowego (X,Y) zadanego ciagiem $\{(x_n, y_k, p_{nk}), n \in \mathbb{T}_1, k \in \mathbb{T}_2\}$:

rozkład zmiennej losowej X zadany jest ciągiem $\{(x_n, p_{n\cdot}), n \in \mathbb{T}_1\}$, gdzie $p_{n\cdot} = P(X = x_n) = \sum\limits_{k\in\mathbb{T}_2} P(X = x_n, Y = y_k) = \sum\limits_{k\in\mathbb{T}_2} p_{nk}$

Podobnie, rozkład zmiennej losowej Y zadany jest ciągiem $\{(y_k, p_{\cdot k}), k \in \mathbb{T}_2\}$, gdzie $p_{\cdot k} = P(Y = y_k) = \sum\limits_{n\in\mathbb{T}_1} P(X = x_n, Y = y_k) = \sum\limits_{n\in\mathbb{T}_1} p_{nk}$

W przypadku wektora o rozkładzie **ciągłym** o gęstości łącznej $f(x,y)$ można pokazać, że:

rozkład zmiennej losowej X jest ciągły o gęstości $f_X(x) = \int\limits_{-\infty}^{\infty} f(x,y)dy$,

rozkład zmiennej losowej Y jest ciągły o gęstości $f_Y(y) = \int\limits_{-\infty}^{\infty} f(x,y)dx$.

DZIAŁANIA NA ZMIENNYCH LOSOWYCH

(X,Y) to wektor losowy. Definiujemy zmienną losową $Z = g(X,Y)$, gdzie g jest odpowiednią funkcją. Aby określić rozkład Z, potrzebna jest znajomość rozkładu łącznego zmiennych losowych X, Y.

NAJWAŻNIEJSZE PRZYKŁADY:

(a) **suma** $Z = X + Y$

Gdy wektor losowy (X,Y) ma rozkład **dyskretny** zadany ciągiem $\{(x_n, y_k, p_{nk}), n \in \mathbb{T}_1, k \in \mathbb{T}_2\}$, zmienna losowa Z ma także rozkład dyskretny zadany ciągiem $\{(z_j, p_j^{(Z)}), j \in \mathbb{T} \subset \mathbb{N}\}$, gdzie różnowartościowy ciąg $\{z_j\}$ utworzony jest z wszystkich liczb postaci $x_n + y_k$ oraz $p_j^{(Z)}$ to suma wyrazów ciągu $\{p_{nk}\}$ o takich numerach nk, dla których $z_j = x_n + y_k$.
Można zapisać $p_j^{(Z)}$ jako:

$$p_j^{(Z)} = \sum_{n\in\mathbb{T}_1} P(X = x_n, Y = z_j - x_n) = \sum_{k\in\mathbb{T}_2} P(X = z_j - y_k, Y = y_k)$$

Gdy wektor losowy (X,Y) ma rozkład **ciągły** o gęstości $f(x,y)$, zmienna losowa Z też ma rozkład ciągły o gęstości

$$f_Z(z) = \int\limits_{-\infty}^{\infty} f(x, z-x)dx = \int\limits_{-\infty}^{\infty} f(z-y, y)dy$$

Opracowanie: dr hab. Agnieszka Jurlewicz

(b) **iloczyn** $Z = XY$

Gdy wektor losowy (X, Y) ma rozkład **dyskretny** zadany ciągiem $\{(x_n, y_k, p_{nk}), n \in \mathbb{T}_1, k \in \mathbb{T}_2\}$, zmienna losowa Z ma także rozkład dyskretny zadany ciągiem $\{(z_j, p_j^{(Z)}), j \in \mathbb{T} \subset \mathbb{N}\}$, gdzie różnowartościowy ciąg $\{z_j\}$ utworzony jest z wszystkich liczb postaci $x_n y_k$ oraz $p_j^{(Z)}$ to suma wyrazów ciągu $\{p_{nk}\}$ o takich numerach nk, dla których $z_j = x_n y_k$.
Przy założeniu, że $P(X = 0) = 0$, można zapisać $p_j^{(Z)}$ jako:

$$p_j^{(Z)} = \sum_{n \in \mathbb{T}_1} P\left(X = x_n, Y = \frac{z_j}{x_n}\right)$$

albo - gdy $P(Y = 0) = 0$:

$$p_j^{(Z)} = \sum_{k \in \mathbb{T}_2} P\left(X = \frac{z_j}{y_k}, Y = y_k\right)$$

Gdy wektor losowy (X, Y) ma rozkład **ciągły** o gęstości $f(x, y)$, zmienna losowa Z też ma rozkład ciągły o gęstości

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f\left(x, \frac{z}{x}\right) dx = \int_{-\infty}^{\infty} \frac{1}{|y|} f\left(\frac{z}{y}, y\right) dy$$

(c) **iloraz** $Z = \dfrac{X}{Y}$, przy założeniu, że $P(Y = 0) = 0$.

Gdy wektor losowy (X, Y) ma rozkład **dyskretny** zadany ciągiem $\{(x_n, y_k, p_{nk}), n \in \mathbb{T}_1, k \in \mathbb{T}_2\}$, zmienna losowa Z ma także rozkład dyskretny zadany ciągiem $\{(z_j, p_j^{(Z)}), j \in \mathbb{T} \subset \mathbb{N}\}$, gdzie różnowartościowy ciąg $\{z_j\}$ utworzony jest z wszystkich liczb postaci $\dfrac{x_n}{y_k}$ oraz $p_j^{(Z)}$ to suma wyrazów ciągu $\{p_{nk}\}$ o takich numerach nk, dla których $z_j = \dfrac{x_n}{y_k}$
Można zapisać $p_j^{(Z)}$ jako:

$$p_j^{(Z)} = \sum_{k \in \mathbb{T}_2} P(X = z_j y_k, Y = y_k)$$

Gdy wektor losowy (X, Y) ma rozkład **ciągły** o gęstości $f(x, y)$, zmienna losowa Z też ma rozkład ciągły o gęstości

$$f_Z(z) = \int_{-\infty}^{\infty} |y| f(zy, y) dy$$

Opracowanie: dr hab. Agnieszka Jurlewicz

FAKT: Ogólnie,

$$\mathrm{E}Z = \mathrm{E}g(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)dF_{X,Y}(x,y) =$$

$$= \begin{cases} \sum\limits_{n\in\mathbb{T}_1}\sum\limits_{k\in\mathbb{T}_2} g(x_n,y_k)p_{nk}, & \text{gdy } X \text{ ma rozkład dyskretny} \\ & \text{zadany ciągiem } \{(x_n,y_k,p_{nk}), n\in\mathbb{T}_1, k\in\mathbb{T}_2\}; \\ \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x,y)f(x,y)dxdy, & \text{gdy } X \text{ ma rozkład ciągły o gęstości } f(x,y). \end{cases}$$

o ile całka (szereg) zbieżne.

Stąd jeśli istnieją $\mathrm{E}X$ i $\mathrm{E}Y$, to

$$\mathrm{E}(X+Y) = \mathrm{E}X + \mathrm{E}Y$$

oraz jeśli istnieją D^2X i D^2Y, to

$$\mathrm{D}^2(X+Y) = \mathrm{D}^2X + \mathrm{D}^2Y + 2(\mathrm{E}(XY) - \mathrm{E}X\mathrm{E}Y).$$

WSPÓŁCZYNNIK KORELACJI

DEFINICJA: Przy założeniu, że istnieją $\mathrm{D}^2X > 0$ i $\mathrm{D}^2Y > 0$, określamy **współczynnik korelacji** zmiennych losowych X i Y jako:

$$\rho_{XY} = \frac{\mathrm{E}(XY) - \mathrm{E}X\mathrm{E}Y}{\sqrt{\mathrm{D}^2X\cdot\mathrm{D}^2Y}}.$$

WŁASNOŚCI WSPÓŁCZYNNIKA KORELACJI:

- $|\rho_{XY}| \leqslant 1$.
- $|\rho_{XY}| = 1$ wtedy i tylko wtedy, gdy $Y = aX + b$ dla pewnych stałych $a \neq 0, b$, przy czym $\rho_{XY} = 1$ odpowiada $a > 0$, a $\rho_{XY} = -1$ odpowiada $a < 0$ (pełna liniowa zależność Y od X).
- Gdy $\rho_{XY} = 0$, mówimy, że X i Y są **nieskorelowane**.

Opracowanie: dr hab. Agnieszka Jurlewicz

Niezależność zmiennych losowych

Definicja: Zmienne losowe X i Y są **niezależne**, gdy dla dowolnych borelowskich zbiorów B_1 i B_2 zdarzenia $\{X \in B_1\}$ i $\{Y \in B_2\}$ są niezależne,
tzn. $P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2)$.

Zmienne losowe $X_1, X_2, \ldots, X_n$ są **niezależne**, gdy dla dowolnych borelowskich zbiorów B_1, $B_2, \ldots, B_n$ rodzina $\{\{X_i \in B_i\}, i = 1, 2, \ldots, n\}$ jest rodziną zdarzeń niezależnych.

Fakt: Zmienne losowe X i Y są **niezależne** wtedy i tylko wtedy, gdy

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \forall\, x, y$$

Wówczas

$$\mathrm{E}XY = \mathrm{E}X\mathrm{E}Y$$

a stąd

$$\mathrm{D}^2(X + Y) = \mathrm{D}^2X + \mathrm{D}^2Y$$

oraz $\rho_{XY} = 0$,
o ile wartości oczekiwane i wariancje istnieją, wariancje są niezerowe.

Zatem jeśli zmienne losowe o skończonych i niezerowych wariancjach są niezależne, to są też nieskorelowane. Implikacja odwrotna nie jest prawdziwa.

Opracowanie: dr hab. Agnieszka Jurlewicz