

Rachunek prawdopodobieństwa i statystyka

Komentarz do wykładu z 6. marca

Definicja 1. Funkcję $f(x, y)$ nazywamy funkcją gęstości (gęstością) 2-wymiarowej zmiennej losowej (X, Y) iff

1. $f(x, y) \geq 0$, dla $(x, y) \in \mathbb{R}^2$,
2. $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) dy dx = 1$.

W wypadku dyskretnym funkcję $p(i, j)$ nazywamy gęstością (prawdopodobieństwem) iff

1. $P(X = x_i, Y = y_j) = p_{ij} \geq 0$, dla $(i, j) \in \mathbb{N}^2$,
2. $\sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} p_{ij} = 1$.

Definicja 2. Gęstościami brzegowymi zmiennej losowej (X, Y) o gęstości $f(x, y)$ nazywamy funkcje $f_1(x) = \int_{\mathbb{R}} f(x, y) dy$ oraz $f_2(y) = \int_{\mathbb{R}} f(x, y) dx$.

W wypadku dyskretnym używamy oznaczeń $p_{i\bullet} = \sum_{j \in \mathbb{N}} p_{ij}$ oraz $p_{\bullet j} = \sum_{i \in \mathbb{N}} p_{ij}$.

Definicja 3.

- Momentem zwykłym rzędu k (k -tym momentem zwykłym) zmiennej losowej X nazywamy wartość $m_k = E(X^k)$.
- Momentem centralnym rzędu k (k -tym momentem centralnym) nazywamy wartość $\mu_k = E[(X - EX)^k]$.
- Dla 2-wymiarowej zmiennej losowej (X, Y) momentem mieszanym rzędu (k, l) nazywamy wartości $m_{kl} = E(X^k Y^l)$ oraz $\mu_{kl} = E[(X - EX)^k \cdot (Y - EY)^l]$.

Uwagi.

Rozkład ciągły	Rozkład dyskretny
$m_k = \int_{\mathbb{R}} x^k f(x) dx$	$m_k = \sum_{i \in \mathbb{N}} x_i^k p_i$
$\mu_k = \int_{\mathbb{R}} (x - EX)^k f(x) dx$	$\mu_k = \sum_{i \in \mathbb{N}} (x_i - EX)^k p_i$
$m_{kl} = \int_{\mathbb{R}} \int_{\mathbb{R}} x^k y^l f(x, y) dy dx$	$m_{kl} = \sum_{i, j \in \mathbb{N}} x_i^k y_j^l p_{ij}$

Wartość oczekiwana EX to m_1 , wariancja VX to μ_2 , moment mieszany m_{11} to kowariancja zmiennych X, Y , oznaczenie $\text{Cov}(X, Y)$. Symbole EX , $E(X)$ oznaczają to samo (wartość oczekiwaną). Podobnie symbole VX , $V(X)$ oznaczają wariancję. Dla odróżnienia: $E(X^2)$ oznacza wartość oczekiwaną zmiennej X^2 , natomiast $E^2(X)$ lub $[EX]^2$ – kwadrat wartości oczekiwanej zmiennej X .

Definicja 4. Dana jest 2-wymiarowa zmienna losowa (X, Y) . Zmienne (1-wymiarowe) X, Y nazywamy niezależnymi iff

- $\forall x, y \in \mathbb{R} \quad f(x, y) = f_1(x) \cdot f_2(y)$ (vide def. 2), lub
- $\forall i, j \in \mathbb{N} \quad p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$.

Przykład:

(i) Rozważamy funkcję $f(x, y) = \frac{3xy}{16}$ określoną na obszarze ograniczonym prostymi $y = 0$, $x = 2$ oraz krzywą $y = x^2$. Bezpośrednim rachunkiem sprawdzamy, że

$$\int_0^2 \int_0^{x^2} x \cdot y \, dy \, dx = \frac{16}{3}, \quad (1)$$

czyli nieujemna na wybranym obszarze funkcja $f(x, y)$ może być uważana za gęstość. Równanie (1) zapiszmy jako

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) \, dy \, dx = 1,$$

gdzie domyślnie przyjmujemy, że $f(x, y) = 0$ wszędzie poza zdefiniowanym obszarem.

(ii) 2-wymiarową dystrybuantą jest $F(s, t) \equiv F_{XY}(x, y) = \int_{-\infty}^s \int_{-\infty}^t f(x, y) \, dy \, dx$.

(iii) Wyznamy gęstości brzegowe:

$$\begin{aligned} f_1(x) \equiv f_X(x) &= \int_{\mathbb{R}} \frac{3xy}{16} \, dy = \int_0^{x^2} \frac{3xy}{16} \, dy = \frac{3x^5}{32}, \\ f_2(y) \equiv f_Y(y) &= \int_{\mathbb{R}} \frac{3xy}{16} \, dx = \int_{\sqrt{y}}^2 \frac{3xy}{16} \, dx = \frac{3(4-y)y}{32}. \end{aligned}$$

Zauważmy (w pierwszym równaniu), że dla ustalonego x jest $y \in [0, x^2]$. Podobnie, dla ustalonego y jest $x \in [\sqrt{y}, 2]$. Oprócz tego $f_1(x), f_2(x)$ mogą być uważane za gęstości, bo są nieujemne (w zdefiniowanym obszarze) oraz

$$\int_0^2 f_1(x) \, dx = \int_0^2 \frac{3x^5}{32} \, dx = 1, \quad \int_0^4 f_2(y) \, dy = \int_0^4 \frac{3(4-y)y}{32} \, dy = 1.$$

Kolejne całkowania dają wyniki: $EX = \int_0^2 x \cdot f_1(x) \, dx = \frac{12}{7}$ oraz $EY = \int_0^4 y \cdot f_2(y) \, dy = 2$ (ale to taki dodatek).

(iv) Sprawdźmy, czy zmienne X, Y są niezależne. Weźmy pod rozwagę punkt $(x, y) = (1, 1.5)$. Jest teraz

$$0 = f(1, 1.5) \neq f_1(1) \cdot f_2(1.5) = \frac{3}{32} \cdot \frac{45}{128}.$$

To oznacza, że zmienne X, Y nie są niezależne (krócej: są zależne).

Przykład:

Przykład kolejny pokazuje, że dla zmiennych dyskretnych jest o wiele prościej. Również **z tego powodu** (są też inne powody) będziemy tym zmiennym poświęcać mniej uwagi.

(i) Definiujemy zmienną losową (X, Y) następująco:

	y_1	y_2	y_3	
$(X, Y) =$	x_1	p_{11}	p_{12}	p_{13}
	x_2	p_{21}	p_{12}	p_{23}
	x_3	p_{31}	p_{32}	p_{33}
	x_4	p_{41}	p_{42}	p_{43}

W powyższym wzorze $p_{ij} \geq 0$ oraz $\sum_{i,j} p_{ij} = 1$, czyli na przykład:

X/Y	2	3	5
$(X, Y) = -2$	0.10	0.05	0.07
0	0.05	0.03	0.08
1	0.01	0.07	0.15
$\sqrt{2}$	0.38	0.00	0.01

(ii) Dystrybuanta $F(s, t) = \sum_{x_i \leq s, y_j \leq t} p_{ij}$.

(iii) Gęstości brzegowe:

X/Y	2	3	5	$p_{i\bullet}$
$(X, Y) = -2$	0.10	0.05	0.07	0.22
0	0.05	0.03	0.08	0.16
1	0.01	0.07	0.15	0.23
$\sqrt{2}$	0.38	0.00	0.01	0.39
$p_{\bullet j}$	0.54	0.15	0.31	1.00

Zmienne X, Y mają zatem rozkłady brzegowe:

$X =$	x_i	-2	0	1	$\sqrt{2}$
	$p_{i\bullet}$	0.22	0.16	0.23	0.39
$Y =$	y_j	2	3	5	
	$p_{\bullet j}$	0.54	0.15	0.31	

(iv) Niezależność zmiennych X, Y . Zmienne są zależne, ponieważ

$$0.01 = p_{31} = P(X = 1, Y = 2) \neq P(X = 1) \cdot P(Y = 2) = p_{3\bullet} \cdot p_{\bullet 1} = 0.23 \cdot 0.54.$$

Polski	Angielski
dystybuanta	cdf (cumulative distribution function)
gęstość	density (pdf)
gęstość brzegowa	marginal density
gęstość warunkowa	conditional density
i.i.d. ¹	i.i.d. ²
kowariancja	covariance
rozkład	distribution
rozkład ciągły	continuous distribution
rozkład dyskretny	discrete distribution
wariancja	variance
wartość oczekiwana	expected value
zmienna losowa	random variable
zmienne niezależne	independent variables

Z poważaniem,
Witold Karczewski

¹(zmienne) niezależne o tym samym rozkładzie

²independent and of identical distribution (variables)