

Rachunek prawdopodobieństwa i statystyka

Komentarz do wykładu z 22. marca

Definicja 1. Dana jest 2-wymiarowa zmienna losowa (X, Y) o gęstości $f(x, y)$, zaś gęstości brzegowe to $f_1(x), f_2(y)$. Gęstościami warunkowymi nazywamy funkcje

$$f_{X|Y}(x) = \frac{f(x, y)}{f_2(y)} \quad \text{oraz} \quad f_{Y|X}(y) = \frac{f(x, y)}{f_1(x)}. \quad (1)$$

W wypadku dyskretnym mamy wzory

$$p_{i \leftarrow j} \equiv p_{x_i|y_j} = \frac{p_{ij}}{p_{\bullet j}} \quad \text{oraz} \quad p_{i \rightarrow j} \equiv p_{y_j|x_i} = \frac{p_{ij}}{p_{i \bullet}}. \quad (2)$$

Przykład:

Wracamy do (lekko zmienionego) przykładu z poprzedniej notatki.

(i) Gęstość dwuwymiarowa i gęstości brzegowe są następujące.

$(X, Y) =$	X/Y	2	3	5	$p_{i \bullet}$
	-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
	3	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	3
	$p_{i \bullet}$	0.22	0.16	0.23	0.39
$Y =$	y_j	2	3	5	
	$p_{\bullet j}$	0.54	0.15	0.31	

(ii) Gęstości warunkowe $p_{X|Y}$ są następujące. Dla każdej z kolumn powyższej tabeli obliczamy ilorazy $p_{x_i|y_j} = \frac{p_{ij}}{p_{\bullet j}}$ po to, aby wartości w kolumnach sumowały się do 1. Mówimy, że w kolumnie j -tej znajdują się ppb wartości x_i pod warunkiem iż $Y = y_j$.

$(X Y = y_j) =$	X/Y	2	3	5
	-2	$10/54$	$5/15$	$7/31$
	0	$5/54$	$3/15$	$8/31$
	1	$1/54$	$7/15$	$15/31$
	3	$38/54$	0	$1/31$
		1	1	1

Podobnie jest dla gęstości warunkowych $p_{Y|X}$. Dla każdego wiersza powyższej tabeli obliczamy ilorazy $p_{x_i|y_j} = \frac{p_{ij}}{p_{i \bullet}}$ po to, aby wartości w wierszach sumowały się do 1. Mówimy, że w wierszu

i -tym znajdują się ppb wartości y_j pod warunkiem iż $X = x_i$.

X/Y	2	3	5	
$(Y X = x_i) =$	$10/22$	$5/22$	$7/22$	1
0	$5/16$	$3/16$	$8/16$	1
1	$1/23$	$7/23$	$15/23$	1
3	$38/39$	0	$1/39$	1

Ponieważ kolumny (*respective* wiersze) powyższych tabeli opisują zmienne losowe ma sens zwrot “wartość oczekiwana warunkowa”, dla przykładu

$$E(X|Y = 2) = -2 \cdot 10/54 + 0 \cdot 5/54 + 1 \cdot 1/54 + 3 \cdot 38/54 = \frac{95}{54}.$$

(iii) Przejdźmy teraz do wyznaczenia rozkładu zmiennej losowej $Z = X + Y$. W lewym górnym rogu każdego elementu tabeli znajduje się wartość zmiennej Z , poniżej prawdopodobieństwo takiej wartości.

X/Y	2	3	5
$Z = X + Y =$	$0/0.10$	$1/0.05$	$3/0.07$
0	$2/0.05$	$3/0.03$	$5/0.08$
1	$3/0.01$	$4/0.07$	$6/0.15$
3	$5/0.38$	$6/0.00$	$8/0.01$

Po uporządkowaniu otrzymujemy następujący rozkład zmiennej Z :

z_i	0	1	2	3	4	5	6	8
p_i	0.10	0.05	0.05	0.11	0.07	0.46	0.15	0.01

Zauważmy, na zakończenie przykładu, że $4.04 = E(X + Y) = 0.96 + 3.08 = E(X) + E(Y)$.

Twierdzenie 1. Niech (X, Y) będzie 2-wymiarową zmienną losową. Wówczas wartość oczekiwana sumy zmiennych losowych X, Y jest równa sumie wartości oczekiwanych tych zmiennych: $E(X + Y) = E(X) + E(Y)$.

Dowód. Niech (X, Y) będzie zmienną losową typu dyskretnego. Jest wówczas:

$$\begin{aligned} E(X + Y) &= \sum_i \sum_j (x_i + y_j) \cdot p_{ij} = \sum_i \sum_j x_i p_{ij} + \sum_j \sum_i y_j p_{ij} = \\ &= \sum_i \left(x_i \sum_j p_{ij} \right) + \sum_j \left(y_j \sum_i p_{ij} \right) = \sum_i x_i \cdot p_{i\bullet} + \sum_j y_j \cdot p_{\bullet j} = E(X) + E(Y) \end{aligned} \quad (3)$$

W wypadku ciągłym jest:

$$\begin{aligned} E(X + Y) &= \int_{\mathbb{R}} \int_{\mathbb{R}} (x + y) f(x, y) dy dx = \int_{\mathbb{R}} \int_{\mathbb{R}} x f(x, y) dy dx + \int_{\mathbb{R}} \int_{\mathbb{R}} y f(x, y) dx dy = \\ &= \int_{\mathbb{R}} \left(x \int_{\mathbb{R}} f(x, y) dy \right) dx + \int_{\mathbb{R}} \left(y \int_{\mathbb{R}} f(x, y) dx \right) dy = \int_{\mathbb{R}} x f_1(x) dx + \int_{\mathbb{R}} y f_2(y) dy = E(X) + E(Y). \end{aligned} \quad (4)$$

□

Przypomnijmy, że kowariancją zmiennych X, Y nazywamy wartość wyrażenia $\mu_{11} \equiv \text{Cov}(X, Y) = E[(X - EX) \cdot (Y - EY)]$. Dla zmiennych dyskretnych $\text{Cov}(X, Y) = \sum_i \sum_j (x_i - EX) \cdot (y_j - EY) p_{ij}$,

dla zmiennych typu ciągłego $\int_{\mathbb{R}} \int_{\mathbb{R}} (x - EX)(y - EY) f(x, y) dy dx$. Poniższe twierdzenie podaje związek między niezależnością zmiennych i ich kowariancją.

Twierdzenie 2. Dana jest zmienna (X, Y) której zmienne brzegowe X, Y są niezależne. Wówczas $\text{Cov}(X, Y) = 0$.

Dowód. (Dla zmiennych typu dyskretnego).

$$\begin{aligned}\text{Cov}(X, Y) &= \sum_i \sum_j (x_i - EX) \cdot (y_j - EY) p_{ij} = \\ &= \sum_i \sum_j x_i y_j p_{ij} - EY \sum_i \sum_j x_i p_{ij} - EX \sum_i \sum_j y_j p_{ij} + EX \cdot EY \sum_i \sum_j p_{ij} = \\ &= \sum_i x_i \left(\sum_j y_j p_{i\bullet} \right) - EY \sum_i \left(x_i \sum_j p_{ij} \right) - EX \left(\sum_j y_j \sum_i p_{ij} \right) + EX \cdot EY = \\ &= \left(\sum_i x_i p_{i\bullet} \right) \left(\sum_j y_j p_{\bullet j} \right) - EY \sum_i x_i p_{i\bullet} - EX \sum_j y_j p_{\bullet j} + EX \cdot EY = 0.\end{aligned}$$

□

UWAGI:

1. Jeżeli zmienne są niezależne, to $E(X \cdot Y) = EX \cdot EY$ (por. zadanie 1.7b).
2. Odwroćenie twierdzenia (2) nie jest prawdziwe.
3. Zadanie 1.7a to szczególny przypadek twierdzenia $V(X) = E(X^2) - (EX)^2$

Przykład:

(Kontynuacja przykładu z notatki 3. ze strony 2.)

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

2-wymiarową dystrybuantą jest $F(s, t) \equiv F_{XY}(x, y) = \int_{-\infty}^s \int_{-\infty}^t f(x, y) dy dx$. Niestety, przy obliczaniu dystrybuanty powinniśmy określić precyzyjnie przedziały całkowania.

(i) Niech A oznacza (w terminologii geometrycznej \equiv licealnej) II, II i IV “ćwiartkę” płaszczyzny. Jeżeli $(s, t) \in A$ to oczywiście $F(s, t) \equiv F_{XY}(x, y) = \int_{-\infty}^s \int_{-\infty}^t f(x, y) dy dx = 0$.

(ii) Niech B oznacza obszar ograniczony prostymi $y = 0$, $x = 2$ oraz krzywą $y = x^2$. Jest teraz:

$$\begin{aligned}F(s, t) &= \int_{-\infty}^s \int_{-\infty}^t f(x, y) dy dx = \\ &= \int_0^{\sqrt{t}} \int_0^{x^2} f(x, y) dy dx + \int_{\sqrt{t}}^s \int_0^t f(x, y) dy dx.\end{aligned}$$

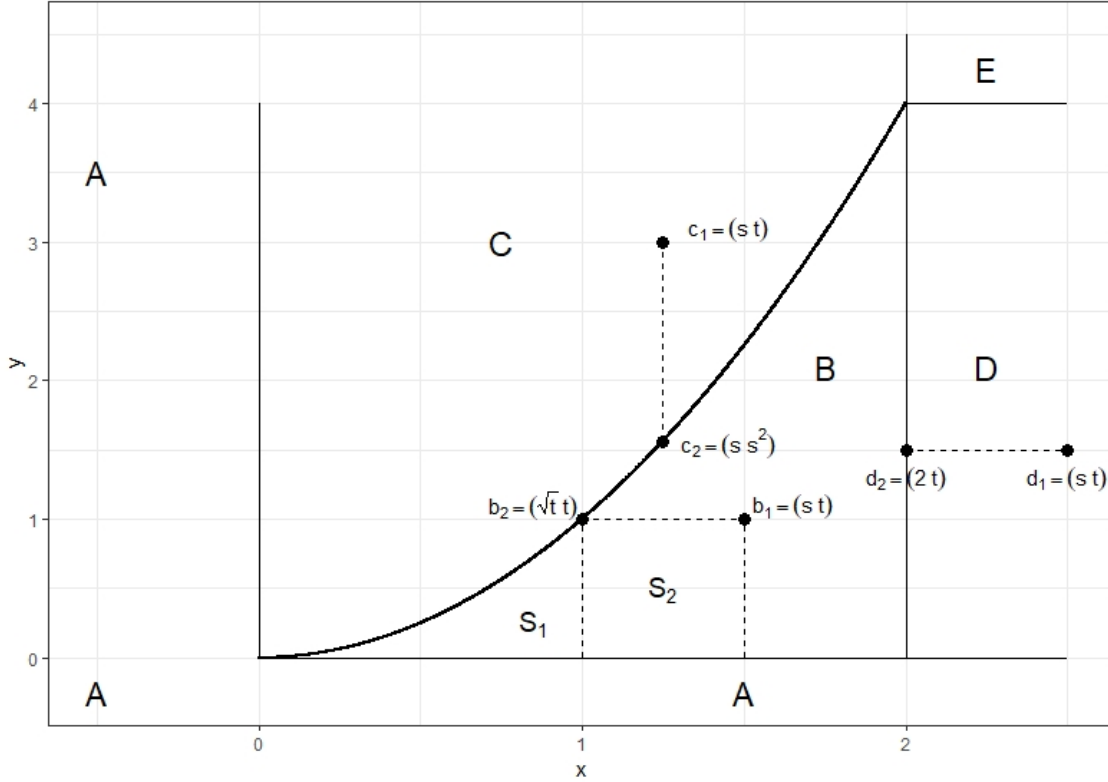
Intuicja: najpierw liczymy pole (całkę, ppb) pod krzywą $y = x^2$, dla $x \in (0, \sqrt{t})$ (obszar S_1) i dodajemy pole (całkę, ppb) pod prostą $y = t$ dla $x \in (\sqrt{t}, s)$ (obszar S_2). W pierwszej całce y zmienia się (dla ustalonego x) od 0 do x^2 , w drugiej – (również dla ustalonego x) od 0 do t . Obszar B będzie też użyteczny w punkcie (iii) i kolejnych.

(iii) Obszar $C = [0, 2] \times [x^2, \infty)$. Jest teraz $F(s, t) \equiv F_{XY}(x, y) = \int_{-\infty}^s \int_{-\infty}^t f(x, y) dy dx = F(s, s^2)$.

Intuicyjnie: obszar na lewo i poniżej punktu $c_1 = (s, t)$ z uwzględnieniem nośnika funkcji $f(x, y)$ jest taki sam jak obszar wyznaczony przez punkt $c_2 = (s, s^2)$. Można zatem odwołać się do wzoru z punktu (ii).

(iv) Obszar $D = [2, \infty) \times [0, 4]$. Tutaj $F(s, t) = F(2, t)$. Proszę porównać przecięcie zbioru $(-\infty, s] \times (-\infty, t]$ z obszarem na którym $f(x, y)$ jest różne od 0. Graficznie: zamiast punktu

$d_1 = (s, t)$ trzeba wziąć do obliczeń punkt $d_2 = (2, t)$. Tutaj również zastosowanie ma wynik z punktu (ii).



(v) Obszar $E = [2, \infty) \times [4, \infty)$. To, na równi z obszarem A, najprostszy wypadek. Jest tutaj $F(s, t) = 1$, ponieważ całkujemy gęstość po całym obszarze “niezerowości”.

Ostatecznie, wzór na dystrybuantę to:

$$F(s, t) = \begin{cases} 0, & \text{dla } (s, t) \in A, \\ \frac{3s^2t^2}{64} - \frac{t^3}{96}, & \text{dla } (s, t) \in B, \\ \frac{s^6}{64}, & \text{dla } (s, t) \in C, \\ \frac{3t^2}{16} - \frac{t^3}{96}, & \text{dla } (s, t) \in D, \\ 1, & \text{dla } (s, t) \in E. \end{cases}$$

Kolejny przykład ilustrujący wyznaczanie sumy zmiennych losowych.

Przykład:

Zmienna losowa (X, Y) ma rozkład o gęstości $f(x, y) = 3x\sqrt{y}$ na obszarze $[0, 1] \times [0, 1]$. Wyznaczyć rozkład zmiennej $Z = X + Y$.

Rozpoczynamy od przejścia $(X, Y) \mapsto (Z, T)$. Niech $Z = X + Y$, $T = Y$ (wzór na T może być inny). Najpierw odwracamy przekształcenie i otrzymujemy $X = Z - T$, $Y = T$. Wyznaczamy Jacobian odwrócenia:

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial t} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1.$$

Kolejny krok to wstawienie odwrotu do gęstości $f(x, y)$ i pomnożenie przez moduł Jacobianu: $g(z, t) = f(x(z, t), y(z, t)) \cdot |J| = 3(z - t) \sqrt{t}$.

Kluczowe zdanie: rozkład zmiennej Z to jeden z rozkładów brzegowych 2-wymiarowej zmiennej (Z, T) . Należy zatem scałkować funkcję $g(z, t)$ po zmiennej t . Dla ustalonego z ($z \in [0, 2]$) należy wyznaczyć obszar zmienności zmiennej t .

$$\begin{cases} 0 < x < 1 \\ 0 < y < 1 \end{cases} \quad \begin{cases} 0 < z - t < 1 \\ 0 < t < 1 \end{cases} \quad \begin{cases} z - 1 < t < z \\ 0 < t < 1 \end{cases}.$$

Przedział całkowania dla zmiennej t to $[\max\{0, z - 1\}, \min\{1, z\}]$. Dla $z \in [1, 2]$ mamy $t \in [0, z]$; dla $z \in [0, 1]$ mamy $t \in [z - 1, 1]$.

Obliczamy całkę nieoznaczoną

$$\int g(z, t) dt = \int 3(z\sqrt{t} - t\sqrt{t}) dt = t\sqrt{t} (2z - \frac{6}{5}t + C).$$

Ostatecznie

$$g_1(z) = \begin{cases} t\sqrt{t} (2z - \frac{6}{5}t) \Big|_{t=0}^z, & z \in [0, 1], \\ t\sqrt{t} (2z - \frac{6}{5}t) \Big|_{t=z-1}^1, & z \in [1, 2]. \end{cases}$$

←

Z poważaniem,
Witold Karczewski