

Andrzej Sierociński

RACHUNEK PRAWDOPODOBIEŃSTWA

Zmienne losowe wielowymiarowe
Wykład 8

15 Zmienne losowe dwuwymiarowe

Często w doświadczeniu losowym mamy do czynienia z kilkoma wielkościami losowymi, które w jakiś sposób są ze sobą powiązane i do opisu eksperymentu niewystarczające jest podanie jedynie rozkładów pojedynczych zmiennych.

Najpierw rozważymy przypadek dwóch zmiennych, aby poznać różnice między przypadkiem jednowymiarowym a dwuwymiarowym. Następnie uzyskane wyniki rozszerzymy na przypadek większej liczby zmiennych.

15.1 Dystrybuanta dwuwymiarowa

Definicja.

Mówimy, że (X, Y) jest dwuwymiarową zmienną losową, jeżeli X i Y są dwiema zmiennymi losowymi zdefiniowanymi na tej samej przestrzeni zdarzeń elementarnych Ω

$$(X, Y) : \Omega \longrightarrow \mathbb{R}^2.$$

oraz

$$\{\omega \in \Omega : X(\omega) \leq x, Y(\omega) \leq y\} \in \mathcal{F} \quad \forall (x, y) \in \mathbb{R}^2$$

jest zdarzeniem losowym.

Analogicznie jak w przypadku jednowymiarowym, zdefiniujemy pojęcie dystrybuanty dwuwymiarowej zmiennych losowych X i Y .

Definicja.

Dystrybantą dwuwymiarową zmiennych losowych X i Y nazywamy taką funkcję dwóch zmiennych F , że dla dowolnego $(x, y) \in \mathbb{R}^2$

$$F(x, y) = P(X \leq x, Y \leq y) = P(\{\omega \in \mathcal{S} : X(\omega) \leq x, Y(\omega) \leq y\}).$$

Innymi słowy, $F(x, y)$ jest prawdopodobieństwem zdarzenia takim, że dwuwymiarowa zmienna losowa (X, Y) przyjmuje wartości ze zbioru $\{(s, t) \in \mathbb{R}^2 : s \leq x, t \leq y\}$.

Dystrybuanta dwuwymiarowa spełnia podobne warunki konieczne i dostateczne jak dystrybuanta jednowymiarowa.

Własności dystrybuanty dwuwymiarowej

1. $\lim_{x \rightarrow -\infty} F(x, y) = 0$, $\lim_{y \rightarrow -\infty} F(x, y) = 0$, oraz $\lim_{x, y \rightarrow \infty} F(x, y) = 1$.
2. Dystrybuanta dwuwymiarowa jest funkcją niemalejącą ze względu na każdą ze zmiennych, tzn. $\forall x \ F(x, y)$ oraz $\forall y \ F(x, y)$ są, odpowiednio, funkcjami niemalejącymi ze względu na y i x .
3. Dystrybuanta dwuwymiarowa jest funkcją prawostronnie ciągłą ze względu na każdą ze zmiennych.
4. Dla dowolnych $x_1 < x_2$ oraz $y_1 < y_2$

$$F(x_1, y_1) + F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) \geq 0.$$

Twierdzenie

Warunki 1-4 są warunkami koniecznymi i wystarczającymi na to, aby dwuwymiarowa funkcja F była dystrybuantą pewnej zmiennej losowej.

Przykład. Pokazać, że funkcja

$$F(x, y) = \begin{cases} 1 & \text{dla } x + y \geq 1, \ x \geq 0, \ y \geq 0 \\ 0 & \text{w przeciwnym wypadku} \end{cases}$$

spełnia warunki 1-3, ale nie spełnia warunku 4, zatem F nie jest dystrybuantą żadnej zmiennej losowej.

Rozwiązanie.

Istotnie, dla $x_1 = y_1 = 0$ oraz $x_2 = y_2 = 1$ mamy

$$F(0, 0) + F(1, 1) - F(1, 0) - F(0, 1) = -1.$$

15.2 Dwuwymiarowe zmienne losowe typu dyskretnego

W przypadku, gdy obie zmienne losowe X i Y są typu dyskretnego z nośnikami $\mathcal{X} = \{x_1, x_2, \dots\}$ oraz $\mathcal{Y} = \{y_1, y_2, \dots\}$, można zdefiniować funkcję prawdopodobieństwa rozkładu łącznego.

Definicja.

Funkcję $p(x, y)$ zdefiniowaną wzorem

$$p(x, y) = P(X = x, Y = y) \quad \forall (x, y) \in R^2.$$

nazywamy **funkcją prawdopodobieństwa rozkładu łącznego** dwuwymiarowej zmiennej losowej (X, Y) .

Własności funkcji prawdopodobieństwa $p(x, y)$

1. $p(x, y) \geq 0, \quad \forall (x, y) \in R^2.$
2. $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) = 1.$

Twierdzenie

Warunki 1-2 są warunkami koniecznymi i dostatecznymi na to, aby funkcja $p(x, y)$ była funkcją prawdopodobieństwa dwuwymiarowej zmiennej losowej typu dyskretnego.

Dla wszystkich $x \in \mathcal{X}$ mamy

$$\{X = x\} = \bigcup_{y \in \mathcal{Y}} \{X = x, Y = y\},$$

pozwala nam to zdefiniować funkcję prawdopodobieństwa zmiennej losowej X zwaną funkcją prawdopodobieństwa **rozkładu brzegowego**.

Definicja.

Funkcje prawdopodobieństwa **rozkładów brzegowych** zmiennych losowych X i Y , oznaczonych jako $p_X(x)$ and $p_Y(y)$, są równe odpowiednio

$$p_X(x) = P(X = x) = \sum_{y \in \mathcal{Y}} p(x, y) \quad p_Y(y) = P(Y = y) = \sum_{x \in \mathcal{X}} p(x, y).$$

Przykład. Towarzystwo ubezpieczeniowe świadczy ubezpieczenia komunikacyjne oraz ubezpieczenia mieszkań. Dla każdego typu ubezpieczenia są stosowane zniżki. Niech X będzie zniżką na polisę komunikacyjną (w zł) a Y zniżką na polisę mieszkaniową udzielaną klientowi. Łączny rozkład prawdopodobieństwa zmiennej losowej (X, Y) dany jest poprzez funkcję prawdopodobieństwa zdefiniowaną w tabeli poniżej:

	Y			
X	0	100	200	300
0	0.10	0.15	0.15	0.05
100	0.05	0.05	0.10	0.10
250	0.05	0.05	0.05	0.10

Znaleźć rozkłady brzegowe zmiennych losowych X i Y .

Obliczyć prawdopodobieństwo zdarzenia $\{Y \geq 200\}$.

Rozwiązanie.

Rozkłady brzegowe zmiennych losowych X i Y znajdujemy, sumując prawdopodobieństwa w wierszach i w kolumnach, odpowiednio.

	Y				
X	0	100	200	300	$p_X(x)$
0	0.10	0.15	0.15	0.05	0.45
100	0.05	0.05	0.10	0.10	0.30
250	0.05	0.05	0.05	0.10	0.25
$p_Y(y)$	0.20	0.25	0.30	0.25	

Z tabeli można odczytać, że

$$P(Y \geq 200) = p_Y(200) + p_Y(300) = 0.30 + 0.25 = 0.55.$$

15.3 Dwuwymiarowe zmienne losowe typu ciągłego

Niech (X, Y) będzie parą zmiennych losowych o rozkładach ciągłych przyjmującą wartości ze zbioru $\mathcal{X} \times \mathcal{Y}$, gdzie \mathcal{X} i \mathcal{Y} są, odpowiednio, nośnikami zmiennych losowych X i Y .

Podobnie jak w przypadku jednowymiarowym, możemy zdefiniować funkcję gęstości prawdopodobieństwa, tym razem rozkładu łącznego, która pozwoli nam wyznaczyć prawdopodobieństwo zdarzenia, że zmienna losowa (X, Y) przyjmie wartości z dwuwymiarowego zbioru A , za pomocą całki z funkcji gęstości po zbiorze A .

Definicja.

Niech X i Y będą dwiema zmiennymi losowymi typu ciągłego. Wówczas nieujemna funkcja $f(x, y)$ jest **gęstością prawdopodobieństwa rozkładu łącznego** X i Y , jeżeli dla dowolnego zdarzenia losowego $A \in \mathcal{R}^2$

$$P\{(X, Y) \in A\} = \iint_A f(x, y) dx dy.$$

W szczególności, dla dowolnych $(x, y) \in R^2$ otrzymujemy dystrybuantę łączną

$$F(x, y) = P\{(X, Y) \in (-\infty, x] \times (-\infty, y]\} = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv.$$

Różniczkując dystrybuantę w punktach ciągłości gęstości f mamy

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

Własności gęstości dwuwymiarowej

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

Twierdzenie 3.

Warunki 1-2 są warunkami koniecznymi i wystarczającymi na to, aby funkcja $f(x, y)$ była łączną gęstością prawdopodobieństwa pewnej zmiennej losowej typu ciągłego.

Jeżeli zmienne losowe X i Y mają łączny rozkład typu ciągłego, to rozkłady brzegowe są także typu ciągłego a ich gęstości można uzyskać w następujący sposób:

$$\begin{aligned} P\{X \in A\} &= P\{X \in A, Y \in (-\infty, \infty)\} = \int_A \int_{-\infty}^{\infty} f(x, y) dx dy = \\ &= \int_A f_X(x) dx, \end{aligned}$$

gdzie

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

jest gęstością prawdopodobieństwa zmiennej losowej X .

Definicja.

Gęstościami brzegowymi zmiennych losowych X i Y , oznaczonymi odpowiednio jako $f_X(x)$ oraz $f_Y(y)$, są

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad \text{dla} \quad -\infty < x < \infty,$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad \text{dla} \quad -\infty < y < \infty.$$

Przykład. Pewna sieć typu fast-food posiada okienka dla pieszych oraz okienka dla kierowców. W losowo wybranym dniu, niech X oznacza proporcję czasu, kiedy okienko dla zmotoryzowanych jest zajęte (co najmniej jeden klient jest obsługiwany lub czeka na obsługę) a Y oznacza proporcję czasu zajęcia okienka dla niezmotoryzowanych. Nośnikiem rozkładu łącznego (X, Y) jest kwadrat $D = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$.

Założmy, że gęstość dwuwymiarowa (X, Y) dana jest przez

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{jeżeli} \quad 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

- (a) Sprawdzić, czy $f(x, y)$ jest gęstością prawdopodobieństwa.
- (b) Obliczyć prawdopodobieństwo, że oba okienka nie są zajęte jednocześnie dłużej niż przez 25% czasu.
- (c) Wyznaczyć gęstości brzegowe oraz obliczyć $P(0, 25 \leq Y \leq 0, 75)$.

Rozwiązanie.

(a) Zauważmy, że $f(x, y) \geq 0$. Ponadto

$$\begin{aligned}\iint_{R^2} f(x, y) \, dx \, dy &= \int_0^1 \int_0^1 f(x, y) \, dx \, dy = \\&= \int_0^1 \int_0^1 \frac{6}{5}(x + y^2) \, dx \, dy = \\&= \int_0^1 \int_0^1 \frac{6}{5}x \, dx \, dy + \int_0^1 \int_0^1 \frac{6}{5}y^2 \, dx \, dy = \\&= \int_0^1 \frac{6}{5}x \, dx \, dy + \int_0^1 \frac{6}{5}y^2 \, dy = \\&= \frac{6}{10} + \frac{6}{15} = 1.\end{aligned}$$

(b) Prawdopodobieństwo tego, że oba okienka nie są zajęte jednocześnie dłużej niż przez 25% czasu wynosi

$$\begin{aligned}P(0 \leq X \leq 0.25, 0 \leq Y \leq 0.25) &= \\&= \int_0^{0.25} \int_0^{0.25} \frac{6}{5}(x + y^2) \, dx \, dy = \\&= \frac{6}{5} \int_0^{0.25} \int_0^{0.25} x \, dx \, dy + \frac{6}{5} \int_0^{0.25} \int_0^{0.25} y^2 \, dx \, dy = \\&= \frac{6}{20} \cdot \frac{x^2}{2} \Big|_0^{0.25} + \frac{6}{20} \cdot \frac{y^3}{3} \Big|_0^{0.25} = \frac{7}{640} = 0.0109.\end{aligned}$$

(c) Gęstość rozkładu brzegowego czasu zajętości okienka dla zmotoryzowanych X (bez odwoływania się do czasu zajętości okienka dla niezmotoryzowanych) dla $0 \leq x \leq 1$ jest równa

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) \, dy = \int_0^1 \frac{6}{5}(x + y^2) \, dy = \frac{6}{5}x + \frac{2}{5}.$$

Zatem

$$f_X(x) = \begin{cases} \frac{6}{5}x + \frac{2}{5} & \text{dla } 0 \leq x \leq 1 \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Podobnie otrzymujemy rozkład brzegowy zmiennej Y .

$$f_Y(y) = \begin{cases} \frac{6}{5}y^2 + \frac{3}{5} & \text{dla } 0 \leq y \leq 1 \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Zatem

$$P(0.25 \leq Y \leq 0.75) = \int_{0.25}^{0.75} f_Y(y) dy = \frac{37}{80} = 0.4625.$$

16 Niezależność zmiennych losowych

W wielu sytuacjach, informacja o wartości przyjmowanej przez jedną ze zmiennych daje informację o możliwych wartościach drugiej ze zmiennych. Mówimy wówczas, że zmienne są zależne.

Już wcześniej dowiedzieliśmy się, że niektóre zdarzenia losowe są wzajemnie niezależne. Używając podobnych argumentów możemy zdefiniować niezależne zmienne losowe.

Definicja.

Mówimy, że zmienne losowe X i Y są **niezależne**, jeżeli dla dowolnych zbiorów borelowskich A i B

$$P\{X \in A, Y \in B\} = P\{X \in A\} \cdot P\{Y \in B\}.$$

Innymi słowy, X i Y są niezależne, jeżeli dla dowolnych A i B , zdarzenia losowe $E_A = \{X \in A\}$ oraz $E_B = \{Y \in B\}$ są niezależne.

Zmienne losowe, które nie są niezależne nazywamy zmiennymi **zależnymi**.

Korzystając z aksjomatów prawdopodobieństwa można pokazać, że definicja ta jest równoważna warunkowi:

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\} \cdot P\{Y \leq y\}$$

dla dowolnych $x, y \in \mathcal{R}$.

Używając pojęcia łącznej dystrybuanty F zmiennych losowych X i Y , otrzymujemy następujące twierdzenie.

Twierdzenie

Zmienne losowe X i Y są niezależne wtedy i tylko wtedy, gdy

$$F(x, y) = F_X(x) \cdot F_Y(y) \quad \text{dla dowolnych } x, y \in \mathcal{R}.$$

W praktyce, warunek niezależności podany w ostatnim twierdzeniu, nie jest zbyt wygodny do sprawdzenia. Szczególnie w przypadku, gdy zmienne losowe X i Y mają rozkłady dyskretne.

Twierdzenie

Jeżeli zmienne losowe X i Y są zmiennymi losowymi typu dyskretnego, to X i Y są niezależne wtedy i tylko wtedy, gdy

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad \text{dla dowolnych } x \in \mathcal{X}, y \in \mathcal{Y},$$

gdzie p_X i p_Y są funkcjami rozkładów brzegowych X i Y .

Dowód.

(\Rightarrow) Jeżeli X i Y są niezależne, to kładąc $A = \{x\}$ oraz $B = \{y\}$, otrzymujemy warunek wymieniony w twierdzeniu.

(\Leftarrow) Jeżeli warunek wymieniony w twierdzeniu jest spełniony, to dla dowolnych zbiorów A i B

$$\begin{aligned} P\{X \in A, Y \in B\} &= \sum_{x \in A} \sum_{y \in B} p(x, y) = \sum_{x \in A} \sum_{y \in B} p_X(x) \cdot p_Y(y) = \\ &= \sum_{x \in A} p_X(x) \cdot \sum_{y \in B} p_Y(y) = \\ &= P\{X \in A\} \cdot P\{Y \in B\} \end{aligned}$$

a stąd wynika niezależność zmiennych X i Y . ■

Twierdzenie 6.

Jeżeli zmienne losowe X i Y są typu ciągłego, to niezależność X i Y jest równoważna warunkowi

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

dla dowolnych $x, y \in \mathcal{R}$, dla których gęstość f jest ciągła a f_X oraz f_Y są, odpowiednio, gęstościami brzegowymi zmiennych losowych X i Y .

Mówiąc nieprecyzyjnie, zmienne losowe X i Y są niezależne jeśli wiedza o wartości jednej zmiennej nie ma wpływu na rozkład prawdopodobieństwa drugiej.

Przykład. Pokazać, że zmienne losowe X i Y o rozkładzie łącznym:

	Y			
X	0	100	200	300
0	0.10	0.15	0.15	0.05
100	0.05	0.05	0.10	0.10
250	0.05	0.05	0.05	0.10

są zależne. Znaleźć rozkład niezależnych zmiennych losowych (X_1, Y_1) mających te same rozkłady brzegowe co zmienne losowe (X, Y) .

Rozwiązanie.

Ponieważ $p(0, 0) = 0.10 \neq 0.45 \cdot 0.20 = p_X(0) \cdot p_Y(0)$, to X i Y są zależne. Rozkładem łącznym zmiennych niezależnych X_1 i Y_1 jest tzw. rozkład produktowy

$$p(x, y) = p_X(x) \cdot p_Y(y) \quad \forall (x, y) \in \mathcal{R}^2.$$

Stąd

	Y ₁				
X ₁	0	100	200	300	p _{X₁} (x)
0	0.09	0.1125	0.135	0.1125	0.45
100	0.06	0.0750	0.090	0.0750	0.30
250	0.05	0.0625	0.075	0.0625	0.25
p _{Y₁} (y)	0.20	0.2500	0.300	0.2500	

Przykład. Pokazać, że zmienne losowe X i Y o gęstości łącznej

$$f(x, y) = \begin{cases} 24xy & \text{dla } 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

są zależne.

Rozwiązanie.

Jest oczywiste, że dla $(x, y) \in \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1, x + y > 1\}$ mamy

$$f(x, y) = 0 \neq 12x(1-x)^2 \cdot 12y(1-y)^2 = f_X(x) \cdot f_Y(y),$$

zatem X i Y są zależne.

Wniosek

Jeżeli (X, Y) jest dwuwymiarową zmienną losową typu ciągłego oraz X i Y są niezależne, to obszar w którym gęstość łączna (X, Y) jest niezerowa jest postaci $\mathcal{X} \times \mathcal{Y}$, gdzie \mathcal{X} jest nośnikiem X a \mathcal{Y} nośnikiem Y .

Przykład. Załóżmy, że czasy życia dwóch urządzeń są niezależne od siebie o rozkładach wykładniczych, $X \sim \text{Exp}(\lambda_1)$ $Y \sim \text{Exp}(\lambda_2)$. Wyznaczyć gęstość łączną zmiennej losowej (X, Y) oraz prawdopodobieństwo, że oba urządzenia nie zepsują się przed upływem 1500 godzin.

Rozwiązanie.

Z niezależności X i Y otrzymujemy gęstość łączną

$$f(x, y) = f_X(x) \cdot f_Y(y) = \begin{cases} \lambda_1 e^{-\lambda_1 x} \cdot \lambda_2 e^{-\lambda_2 y} & \text{dla } 0 \leq x, 0 \leq y \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Prawdopodobieństwo zdarzenia, że oba urządzenia nie ulegną uszkodzeniu przez co najmniej 1500 godzin jest równe

$$\begin{aligned} P(X \geq 1500, Y \geq 1500) &= P(X \geq 1500) \cdot P(Y \geq 1500) = \\ &= e^{-1500(\lambda_1 + \lambda_2)}. \end{aligned}$$

Na przykład, dla $\lambda_1 = 1/1000$ oraz $\lambda_2 = 1/1200$, tzn. oczekiwany czas bezawaryjnej pracy wynosi odpowiednio 1000 oraz 1200 godzin, to szukane prawdopodobieństwo wynosi

$$e^{-1500(\frac{1}{1000} + \frac{1}{1200})} = 0,2231 \cdot 0,2865 = 0,0639.$$

17 Zmienne losowe wielowymiarowe

Rozważymy przypadek $n > 2$.

Założmy, że mamy n zmiennych losowych X_1, X_2, \dots, X_n określonych na tej samej przestrzeni probabilistycznej (Ω, \mathcal{F}, P) .

Definicja.

Funkcję n -wymiarową $(X_1(\omega), X_2(\omega), \dots, X_n(\omega)) : \Omega \rightarrow \mathcal{R}^n$ taką, że dla dowolnego $(x_1, x_2, \dots, x_n) \in \mathcal{R}^n$

$$\{\omega \in \Omega : X_1(\omega) \leq x_1, X_2(\omega) \leq x_2, \dots, X_n(\omega) \leq x_n\} \in \mathcal{F}$$

jest zdarzeniem losowym, nazywamy n -wymiarową zmienną losową.

Podobnie definiujemy dystrybuantę n -wymiarową

Definicja.

Funkcję n -wymiarową $F : \mathcal{R}^n \rightarrow [0, 1]$ zdefiniowaną

$$F(x_1, x_2, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

dla dowolnych $(x_1, x_2, \dots, x_n) \in \mathcal{R}^n$ nazywamy n -wymiarową dystrybuantą łączną zmiennej losowej (X_1, X_2, \dots, X_n) .

Dystrybuanta n -wymiarowa, podobnie jak dystrybuanta dwuwymiarowa, spełnia 4 warunki konieczne i dostateczne, o których była mowa w twierdzeniu 1, odpowiednio zmodyfikowane dla $n > 2$.

17.1 Zmienne losowe typu dyskretnego

Definicja.

Jeżeli zmienna losowa (X_1, X_2, \dots, X_n) przyjmuje co najwyżej przeliczalną liczbę wartości, to zmienną nazywamy typu dyskretnego oraz funkcję

$$p(x_1, x_2, \dots, x_n) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n),$$

funkcją prawdopodobieństwa rozkładu łącznego.

Własności funkcji prawdopodobieństwa

1. $p(x_1, \dots, x_n) \geq 0, \forall (x_1, \dots, x_n) \in \mathcal{R}^n$.
2. $\sum_{x_1 \in \mathcal{X}_1} \dots \sum_{x_n \in \mathcal{X}_n} p(x_1, \dots, x_n) = 1$.

Twierdzenie

Warunki 1-2 są warunkami koniecznymi i wystarczającymi na to, aby funkcja $p(x_1, x_2, \dots, x_n)$ była funkcją prawdopodobieństwa rozkładu łącznego pewnej zmiennej losowej n -wymiarowej typu dyskretnego.

Dla każdego i , $1 \leq i \leq n$, można zdefiniować rozkład brzegowy zmiennej losowej X_i , który jest również typu dyskretnego o funkcji prawdopodobieństwa równej

$$\begin{aligned} p_{X_i}(x_i) &= P(X_i = x_i) = \\ &= \sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_{i-1} \in \mathcal{X}_{i-1}} \sum_{x_{i+1} \in \mathcal{X}_{i+1}} \cdots \sum_{x_n \in \mathcal{X}_n} p(x_1, x_2, \dots, x_n). \end{aligned}$$

17.2 Zmienne losowe typu ciągłego

Definicja.

Zmienną losową n -wymiarową (X_1, X_2, \dots, X_n) nazywamy zmienną losową typu ciągłego, jeżeli istnieje taka nieujemna funkcja $f(x_1, x_2, \dots, x_n) \geq 0$, że dla dowolnego zbioru borelowskiego $A \subset \mathcal{R}^n$

$$P\{(X_1, X_2, \dots, X_n) \in A\} = \int_A f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n.$$

W szczególności dla $(x_1, \dots, x_n) \in R^n$ otrzymujemy postać dystrybucyjną $F(x_1, \dots, x_n)$

$$\begin{aligned} F(x_1, \dots, x_n) &= P\{(X_1, \dots, X_n) \in (-\infty, x_1] \times \cdots \times (-\infty, x_n]\} = \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \dots, u_n) du_1 \cdots du_n. \end{aligned}$$

Podobnie jak w przypadku dwuwymiarowym otrzymujemy

$$f(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \dots, x_n),$$

dla wszystkich punktów, w których pochodna istnieje.

Własności gęstości wielowymiarowej

1. $f(x_1, \dots, x_n) \geq 0$, $\forall (x_1, \dots, x_n) \in R^n$.
2. $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$.

Twierdzenie

Warunki 1-2 są warunkami koniecznymi i wystarczającymi na to, aby funkcja $f(x_1, \dots, x_n)$ była gęstością łączną pewnej zmiennej losowej n -wymiarowej typu ciągłego.

18 Niezależność

Jeżeli zmienna losowa (X_1, X_2, \dots, X_n) jest typu ciągłego, to wszystkie rozkłady brzegowe są również typu ciągłego. W szczególności, jednowymiarowe rozkłady brzegowe zmiennych losowych X_i , $1 \leq i \leq n$ otrzymujemy w następujący sposób:

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty}}_{(n-1) \text{ razy}} f(x_1, \dots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Podobnie jak poprzednio definiujemy niezależność X_1, X_2, \dots, X_n .

Definicja.

Mówimy, że zmienne losowe X_1, X_2, \dots, X_n są niezależne, jeżeli dla dowolnych zbiorów borelowskich $A_1, A_2, \dots, A_n \in \mathcal{B}(R)$,

$$P\{X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n\} = \prod_{i=1}^n P\{X_i \in A_i\}.$$

Warunek niezależności dla dystrybuanty n -wymiarowej.

$$F(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i) \quad \text{dla dowolnego } (x_1, \dots, x_n) \in \mathcal{R}^n,$$

gdzie $F_{X_i}(x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$, $i = 1, \dots, n$ są dystrybuantami brzegowymi jednowymiarowymi.

Innymi słowy, zmienne losowe X_1, X_2, \dots, X_n są niezależne jeśli dla dowolnego podzbioru $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ zmiennych losowych (każda para, trójka itd.), ma rozkład łączny, który jest produktem rozkładów brzegowych.

Rozważmy ciąg n niezależnych, identycznych doświadczeń, kończących się jednym z r możliwych wyników $\{a_1, a_2, \dots, a_r\}$. Niech

$$p_i = P(\text{zaobserwowano } a_i \text{ w pojedynczym doświadczeniu}), \quad i = 1, 2, \dots, r.$$

Zdefiniujmy zmienną losową

$$X_i = \text{liczba doświadczeń kończących się wynikiem } a_i, \quad i = 1, 2, \dots, r.$$

Takie doświadczenie nazywamy **doświadczeniem wielomianowym**, a funkcję prawdopodobieństwa zmiennej losowej X_1, X_2, \dots, X_r nazywamy **rozkładem wielomianowym**.

Zatem funkcja prawdopodobieństwa zmiennej losowej X_1, X_2, \dots, X_r dana jest wzorem

$$p(x_1, \dots, x_r) = \frac{n!}{(x_1!) \cdot \dots \cdot (x_r!)} p_1^{x_1} \cdot \dots \cdot p_r^{x_r}, \quad \forall i \ 0 \leq x_i \leq r,$$

gdzie $\sum_{i=1}^r p_i = 1$ oraz $\sum_{i=1}^r x_i = n$.

Przykład. Załóżmy, że X_1, X_2, \dots, X_n oznaczają czasy życia n niezależnie pracujących elementów pewnego systemu. Każda ze zmiennych losowych X_i ma rozkład wykładniczy z parametrem λ . Załóżmy, że system ulega awarii w chwili, gdy jeden z jego elementów ulega uszkodzeniu. Wyznaczyć rozkład prawdopodobieństwa zmiennej losowej T czasu życia (bezawaryjnej pracy) systemu.

Rozwiązanie.

Zmienne losowe X_1, X_2, \dots, X_n są niezależne, zatem dla dowolnych $x_1 \geq 0, \dots, x_n \geq 0$ łączna gęstość jest iloczynem gęstości brzegowych

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}.$$

Tym samym prawdopodobieństwo, że system nie ulegnie awarii przed chwilą $t \geq 0$, jest równe

$$\begin{aligned} P(X_1 > t, \dots, X_n > t) &= \int_t^\infty \dots \int_t^\infty f(x_1, \dots, x_n) dx_1 \dots dx_n = \\ &= \left(\int_t^\infty \lambda e^{-\lambda x_1} dx_1 \right) \cdot \dots \cdot \left(\int_t^\infty \lambda e^{-\lambda x_n} dx_n \right) = \\ &= (e^{-\lambda t})^n = e^{-n\lambda t}. \end{aligned}$$

Dystrybuanta zmiennej losowej T jest równa

$$F_T(t) = P(T \leq t) = \begin{cases} 0 & \text{dla } t < 0 \\ 1 - e^{-n\lambda t} & \text{dla } t \geq 0, \end{cases}$$

czyli T ma rozkład wykładniczy z parametrem $n\lambda$, ($T \sim \text{Exp}(n\lambda)$).

Taki system nazywamy systemem szeregowym, $T = \min\{X_1, X_2, \dots, X_n\}$. Za-uważmy ponadto, że $E[T] = \frac{1}{n\lambda}$.

19 Funkcje zmiennych losowych

Założmy, że X i Y są dwiema zmiennymi losowymi a funkcja $g(x, y)$ funkcją rzeczywistą dwóch zmiennych. Zdefiniujmy nową zmienną losową

$$Z = g(X, Y).$$

Znając rozkłady zmiennych losowych X i Y chcemy wyznaczyć rozkład prawdopodobieństwa zmiennej losowej Z .

19.1 Funkcje zmiennych losowych typu dyskretnego

Jeżeli X i Y są dwiema zmiennymi losowymi typu dyskretnego o rozkładzie łącznym $p(x, y)$, $(x, y) \in \mathcal{X} \times \mathcal{Y}$, to $Z = g(X, Y)$ jest również typu dyskretnego i dla dowolnego $z \in \mathcal{Z} = g(\mathcal{X} \times \mathcal{Y})$ mamy

Rozkład funkcji dwóch zmiennych dyskretnych

$$\forall z \in \mathcal{Z} \quad P(Z = z) = \sum_{\{(x,y) \in \mathcal{X} \times \mathcal{Y} : g(x,y)=z\}} p(x, y).$$

Przykład. Wyznaczyć rozkład zmiennych losowych $T = X + Y$ oraz $U = \max(X, Y)$, gdzie rozkład (X, Y) podany jest w tabeli poniżej.

	Y			
X	0	1	2	3
0	0.1	0.1	0.2	0.1
1	0.1	0.1	0.1	0.0
2	0.1	0.0	0.0	0.1

Rozwiązanie.

Nośnikiem zmiennej losowej T jest $\mathcal{T} = \{0, 1, 2, 3, 4, 5\}$.

$$\{(x, y) : x + y = 0\} = \{(0, 0)\}, \quad \text{zatem} \quad P(T = 0) = p(0, 0) = 0.1.$$

Podobnie otrzymujemy, że

$$\begin{aligned}
 \{(x, y) : x + y = 1\} &= \{(0, 1), (1, 0)\} \\
 P(T = 1) &= p(0, 1) + p(1, 0) = 0.2, \\
 \{(x, y) : x + y = 2\} &= \{(0, 2), (2, 0), (1, 1)\} \\
 P(T = 2) &= p(0, 2) + p(2, 0) + p(1, 1) = 0.4, \\
 \{(x, y) : x + y = 3\} &= \{(1, 2), (2, 1), (0, 3)\} \\
 P(T = 2) &= p(1, 2) + p(2, 1) + p(0, 3) = 0.2, \\
 \{(x, y) : x + y = 4\} &= \{(2, 2), (1, 3)\} \\
 P(T = 2) &= p(2, 2) + p(1, 3) = 0.0, \\
 \{(x, y) : x + y = 5\} &= \{(2, 3)\} \\
 P(T = 5) &= p(2, 3) = 0.1,
 \end{aligned}$$

Ostatecznie

t	0	1	2	3	5
$p_T(t)$	0.1	0.2	0.4	0.2	0.1

Dla zmiennej losowej U nośnikiem jest $\mathcal{U} = \{0, 1, 2, 3\}$.

$$\begin{aligned}
 \{(x, y) : \max(x, y) = 0\} &= \{(0, 0)\} \\
 \{(x, y) : \max(x, y) = 1\} &= \{(0, 1), (1, 0), (1, 1)\} \\
 \{(x, y) : \max(x, y) = 2\} &= \{(0, 2), (2, 0), (1, 2), (2, 1), (2, 2)\} \\
 \{(x, y) : \max(x, y) = 3\} &= \{(0, 3), (1, 3), (2, 3)\}
 \end{aligned}$$

Funkcja prawdopodobieństwa zmiennej U

u	0	1	2	3
$p_U(u)$	0.1	0.3	0.4	0.2

Twierdzenie

Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o jednakowych rozkładach Bernoulliego $b(1, p)$, gdzie $p \in (0, 1)$, to zmienna losowa

$$S_n = \sum_{i=1}^n X_i \sim b(n, p)$$

ma rozkład dwumianowy z parametrami n i p .

Dowód (indukcyjny).

n = 2

Istotnie dla $n = 2$ nośnikiem zmiennej losowej S_2 jest $\{0, 1, 2\}$ oraz

$$P(S_2 = 0) = P(X_1 + X_2 = 0) = P(X_1 = 0) \cdot P(X_2 = 0) = (1 - p)^2.$$

Podobnie

$$\begin{aligned} P(S_2 = 1) &= P(X_1 + X_2 = 1) = P(X_1 = 0) \cdot P(X_2 = 1) + \\ &\quad + P(X_1 = 1) \cdot P(X_2 = 0) = \\ &= (1 - p)p + p(1 - p) = 2p(1 - p), \\ P(S_2 = 2) &= P(X_1 + X_2 = 2) = P(X_1 = 1) \cdot P(X_2 = 1) = p^2 \end{aligned}$$

Krok indukcyjny

Założmy, że $S_n \sim b(n, p)$ oraz $X_{n+1} \sim b(1, p)$ są niezależne. Pokażemy, że $S_{n+1} = S_n + X_{n+1} \sim b(n + 1, p)$.

Nośnikiem S_{n+1} jest $\{0, 1, 2, \dots, n, n + 1\}$.

$$\begin{aligned} P(S_{n+1} = 0) &= P(S_n = 0, X_{n+1} = 0) = P(S_n = 0) \cdot P(X_{n+1} = 0) = \\ &= (1 - p)^n \cdot (1 - p) = (1 - p)^{n+1}. \end{aligned}$$

Podobnie otrzymujemy, że

$$P(S_{n+1} = n + 1) = P(S_n = n, X_{n+1} = 1) = p^n \cdot p = p^{n+1}.$$

Niech k będzie dowolną liczbą naturalną taką, że $1 \leq k \leq n$.

Zmienna losowa X_{n+1} może przyjąć jedną z dwóch wartości (0 lub 1). Zatem

$$\begin{aligned} P(S_{n+1} = k) &= P(S_n = k, X_{n+1} = 0) + P(S_n = k - 1, X_{n+1} = 1) = \\ &= \binom{n}{k} p^k (1 - p)^{n-k} \cdot (1 - p) + \binom{n}{k-1} p^{k-1} (1 - p)^{n-k+1} \cdot p = \\ &= \left[\binom{n}{k} + \binom{n}{k-1} \right] p^k (1 - p)^{n-k+1} = \\ &= \binom{n+1}{k} p^k (1 - p)^{n+1-k}. \end{aligned}$$

■

Z tego twierdzenia wynika, że dowolna zmienna losowa o rozkładzie dwumianowym $b(n, p)$ może być przedstawiona jako suma n niezależnych zmiennych losowych o jednakowych rozkładach Bernoulliego $b(1, p)$.

Prawdziwe jest zatem następujące twierdzenie

Twierdzenie o dodawaniu rozkładów dwumianowych.

Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładach dwumianowych $X_i \sim B(n_i, p)$ z tym samym prawdopodobieństwem sukcesu $p \in (0, 1)$, to

$$S_n = \sum_{i=1}^n X_i \sim B\left(\sum_{i=1}^n n_i, p\right)$$

ma rozkład dwumianowy z parametrami $\sum_{i=1}^n n_i$ i p .

Twierdzenie o dodawaniu rozkładów Poissona. Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładach Poissona $P(\lambda_i)$, gdzie $\lambda_i > 0$, dla $i = 1, 2, \dots, n$, to

$$S_n = \sum_{i=1}^n X_i \sim P\left(\sum_{i=1}^n \lambda_i\right)$$

ma rozkład Poissona z parametrem $\lambda = \sum_{i=1}^n \lambda_i$.

Dowód (indukcyjny).

n=2

Dla $n = 2$ nośnikiem S_2 jest $\{0, 1, 2, \dots\}$.

$$\begin{aligned} P(S_2 = 0) &= P(X_1 = 0, X_2 = 0) = P(X_1 = 0) \cdot P(X_2 = 0) = \\ &= e^{-\lambda_1} \cdot e^{-\lambda_2} = e^{-(\lambda_1 + \lambda_2)}. \end{aligned}$$

Niech $k \geq 1$ będzie dowolną liczbą naturalną. Zmienne losowe X_1 i X_2 są niezależne o wartościach nieujemnych. Zatem

$$\begin{aligned} P(S_2 = k) &= P(X_1 + X_2 = k) = \sum_{j=0}^k P(X_1 = j, X_2 = k - j) = \\ &= \sum_{j=0}^k e^{-\lambda_1} \frac{\lambda_1^j}{j!} \cdot e^{-\lambda_2} \frac{\lambda_2^{k-j}}{(k-j)!} = \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{1}{k!} \sum_{j=0}^k \frac{k!}{j!(k-j)!} \lambda_1^j \lambda_2^{k-j} = \\ &= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}. \end{aligned}$$

Krok indukcyjny.

Jeżeli założymy, że $S_n \sim P(\lambda)$ oraz $X_{n+1} \sim P(\lambda_n)$, gdzie $\lambda = \sum_{i=1}^n \lambda_i$ są niezależne, to dowód jest identyczny jak dla $n = 2$. ■

Przykład. W wyniku przeprowadzonych badań, ustalono, że w pewnej fabryce wytwarzającej sprzęt RTV, liczba wyprodukowanych w ciągu dnia niesprawnych odbiorników telewizyjnych ma rozkład Poissona o średniej 2. Podobnie, liczba niesprawnych zestawów stereo ma również rozkład Poissona o wartości oczekiwanej 3.

Wyznaczyć prawdopodobieństwo zdarzenia, że w ciągu 2 dni, liczba niesprawnych produktów nie przekroczy 5.

Rozwiązanie.

Niech X_1 i X_2 oznaczają liczbę niesprawnych zestawów stereofonicznych wyprodukowanych odpowiednio pierwszego i drugiego dnia.

Podobnie, niech X_3 i X_4 będą liczbami niesprawnych odbiorników TV, wyprodukowanych pierwszego i drugiego dnia, odpowiednio. Załóżmy, że wszystkie zmienne są niezależne. Wówczas

$$S = X_1 + X_2 + X_3 + X_4$$

ma rozkład Poissona z parametrem $\lambda = 2 + 2 + 3 + 3 = 10$.

Otrzymujemy zatem

$$P(S \leq 5) = F(5; 10) = 0,067086.$$

19.2 Funkcje zmiennych losowych typu ciągłego

Jeżeli X i Y są dwiema zmiennymi losowymi typu ciągłego o łącznej gęstości $f(x, y)$, $(x, y) \in \mathcal{R}^2$ oraz $Z = g(X, Y)$ jest jednowymiarową zmienną losową to dla dowolnego $z \in \mathcal{R}$ dystrybuenta zmiennej losowej Z dana jest wzorem

Dystrybuenta funkcji zmiennej losowej typu ciągłego

$$F_Z(z) = P(Z \leq z) = \iint_{\{(x,y): g(x,y) \leq z\}} f(x, y) dx dy.$$

Na przykład, jeżeli $Z = X + Y$ jest suma zmiennych X i Y , to podstawiając

$y = u - x$, otrzymujemy

$$\begin{aligned} F_Z(z) &= \iint_{\{(x,y): x+y \leq z\}} f(x,y) dx dy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} f(x,y) dy = \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^z f(x, u-x) du = \int_{-\infty}^z du \int_{-\infty}^{\infty} f(x, u-x) dx. \end{aligned}$$

Zatem suma zmiennych losowych typu ciągłego Z jest typu ciągłego o gęstości

Gęstość sumy $Z = X + Y$

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x) dx.$$

W szczególności,

Gęstość sumy $X + Y$, X i Y **niezależne** Jeżeli zmienne losowe X i Y są niezależne, to

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x) dx = (f_X * f_Y)(z),$$

gdzie $f_X * f_Y$ oznacza splot gęstości brzegowych zmiennych X i Y .

Przykład. Niech X i Y będą niezależnymi zmiennymi losowymi o jednakowym rozkładzie $N(0, 1)$. Znaleźć rozkład sumy $T = X + Y$.

Rozwiązanie.

Stosując wzór na gęstość sumy dwóch niezależnych zmiennych losowych, otrzymujemy

$$\begin{aligned} f_{X+Y}(z) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}} dx = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2 + (z-x)^2}{2}} dx = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[(\sqrt{2}x - \frac{z}{\sqrt{2}})^2 + \frac{z^2}{2}]} dx = \\ &= \frac{1}{2\pi} e^{-\frac{z^2}{2 \cdot 2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(\sqrt{2}x - \frac{z}{\sqrt{2}})^2} dx. \end{aligned}$$

Podstawiając $\sqrt{2}x - \frac{z}{\sqrt{2}} = t$ otrzymujemy

$$f_{X+Y}(z) = \frac{1}{2\pi} e^{-\frac{z^2}{2 \cdot 2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} \frac{dt}{\sqrt{2}} = \frac{1}{\sqrt{2\pi} \cdot \sqrt{2}} e^{-\frac{z^2}{2 \cdot 2}}.$$

Zatem

$$X + Y \sim N(0, (\sqrt{2})^2).$$

Wynik ten można bez trudu uogólnić na sumę dowolnej liczby zmiennych normalnych o dowolnych parametrach.

Twierdzenie o dodawaniu dla zmiennych normalnych.

Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi losowymi o rozkładach normalnych $N(\mu_i, \sigma_i^2)$, odpowiednio, gdzie $\mu_i \in \mathcal{R}$, $\sigma_i > 0$, $i = 1, 2, \dots, n$, to zmienna losowa

$$S_n = \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

ma rozkład normalny z parametrami

$$\mu = \sum_{i=1}^n \mu_i \quad \text{oraz} \quad \sigma^2 = \sum_{i=1}^n \sigma_i^2.$$

Podobne twierdzenie o dodawaniu można udowodnić dla rozkładów gamma.

Twierdzenie o dodawaniu dla zmiennych o rozkładach gamma z jednakowym parametrem skali.

Jeżeli X_1, X_2, \dots, X_n są niezależnymi zmiennymi o rozkładach gamma $\gamma(\alpha_i, \beta)$, odpowiednio, gdzie $\alpha_i > 0$, $\beta > 0$, $i = 1, 2, \dots, n$, to zmienna losowa

$$S_n = \sum_{i=1}^n X_i \sim \gamma\left(\sum_{i=1}^n \alpha_i, \beta\right)$$

ma rozkład gamma z parametrem kształtu $\alpha = \sum_{i=1}^n \alpha_i$ oraz parametrem skali β .

20 Wartość oczekiwana, kowariancja, korelacja

Dla zmiennej losowej jednowymiarowej X zdefiniowaliśmy wartość oczekiwaną $E[h(X)]$ dowolnej funkcji $h(X)$ od tej zmiennej losowej jako średnią ważoną wartości tej funkcji, z wagami będącymi funkcją prawdopodobieństwa $p(x)$ lub gęstością prawdopodobieństwa $f(x)$ zmiennej X .

Podobną definicję podamy dla funkcji dwuwymiarowej $h : R^2 \rightarrow R$ od zmiennej losowej dwuwymiarowej (X, Y) .

Niech X i Y będą zmiennymi losowymi o rozkładzie łącznym zdefiniowanym przez funkcję prawdopodobieństwa $p(x, y)$ lub funkcję gęstości $f(x, y)$ w zależności czy X i Y są typu dyskretnego czy typu ciągłego.

Definicja.

Wartością oczekiwaną funkcji $h(X, Y)$, oznaczoną przez $E[h(X, Y)]$ nazywamy

$$E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) \cdot p(x, y) & X, Y \text{ typu dyskretnego} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y) dx dy & X, Y \text{ typu ciągłego} \end{cases}$$

pod warunkiem, że szereg (całka) po prawej stronie jest zbieżny (zbieżna) absolutnie.

Rozważmy zmienną losową

$$h(X, Y) = X,$$

w przypadku, gdy (X, Y) jest typu ciągłego o gęstości łącznej $f(x, y)$, wówczas

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) dx dy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx = \\ &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx, \end{aligned}$$

gdzie $f_X(x)$ jest gęstością brzegową zmiennej X .

Wynika stąd, że wartość oczekiwana rozkładu brzegowego X jest taka sama jak w przypadku jednowymiarowym.

Wykorzystując definicję wartości oczekiwanej funkcji zmiennych losowych dwuwymiarowych wyznaczymy wartość oczekiwaną oraz wariancję kombinacji liniowej dwóch zmiennych losowych.

Twierdzenie.

Dla dowolnych zmiennych losowych (X, Y) oraz dowolnych stałych $a, b \in R$

$$E(aX + bY) = aE[X] + bE[Y].$$

Dowód.

Założmy, że (X, Y) jest typu ciągłego o gęstości $f(x, y)$ (w przypadku dyskretnym dowód jest identyczny, jedynie całkę należy zastąpić sumą), wówczas

$$\begin{aligned} E[aX + bY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (ax + by) \cdot f(x, y) dx dy = \\ &= a \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) dx dy + b \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \cdot f(x, y) dx dy \\ &= aE[X] + bE[Y]. \end{aligned}$$

■

Twierdzenie.

Dla dowolnych zmiennych losowych (X, Y) oraz dowolnych stałych $a, b \in R$

$$V(aX + bY) = a^2V[X] + b^2V[Y] + 2ab \cdot E\{(X - E[X])(Y - E[Y])\}.$$

Dowód.

Z definicji wariancji otrzymujemy

$$\begin{aligned} V(aX + bY) &= E\{[(aX + bY) - E(aX + bY)]^2\} = \\ &= E\{[a(X - E[X]) + b(Y - E[Y])]^2\} = \\ &= E\{a^2(X - E[X])^2 + b^2(Y - E[Y])^2 + \\ &\quad + 2ab(X - E[X])(Y - E[Y])\} = \\ &= a^2V[X] + b^2V[Y] + \\ &\quad + 2ab \cdot E\{(X - E[X])(Y - E[Y])\}. \end{aligned}$$

■

Wariancja kombinacji liniowej dwóch zmiennych losowych zależy od obu wariancji oraz dodatkowo od wielkości

$$E\{(X - E[X])(Y - E[Y])\},$$

która mówi o wzajemnej zależności pomiędzy zmiennymi X i Y .

Definicja.

Kowariancję zmiennych losowych X i Y nazywamy wyrażenie

$$Cov(X, Y) = E\{(X - E[X])(Y - E[Y])\}.$$

W ogólności, dodatnia wartość $Cov(X, Y)$ oznacza, że zmienna losowa Y raczej rośnie wraz ze wzrostem wartości X , natomiast w przypadku, gdy kowariancja jest ujemna, raczej maleje. Można powiedzieć, że kowariancja jest miarą zależności pomiędzy X i Y . Jednakże jej wartość mocno zależy od bezwzględnej wielkości X i Y .

Aby usunąć tę wadę, zamiast kowariancji zmiennych X i Y rozważa się kowariancję zmiennych standaryzowanych X^* i Y^* , która mierzy siłę wzajemnej korelacji pomiędzy X i Y .

Definicja.

Współczynnikiem korelacji zmiennych losowych X i Y , oznaczonym jako $Corr(X, Y)$, $\rho_{X,Y}$ lub ρ , nazywamy

$$\rho_{X,Y} = Cov\left(\frac{X - E[X]}{\sigma_X}, \frac{Y - E[Y]}{\sigma_Y}\right) = \frac{Cov(X, Y)}{\sigma_X \cdot \sigma_Y}.$$

Twierdzenie.

Jeżeli (X, Y) są zmiennymi losowymi niezależnymi, dla których istnieje

$$E(X \cdot Y),$$

to

$$E(X \cdot Y) = E[X] \cdot E[Y].$$

Dowód.

Dowód przeprowadzimy w przypadku, gdy (X, Y) jest typu ciągłego o gęstości $f(x, y) = f_X(x)f_Y(y)$ (w przypadku dyskretnym jest podobny)

$$\begin{aligned} E[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x, y) dx dy = \\ &= \left(\int_{-\infty}^{\infty} x \cdot f_X(x) dx \right) \cdot \left(\int_{-\infty}^{\infty} y \cdot f_Y(y) dy \right) \\ &= E[X] \cdot E[Y]. \end{aligned}$$

■

Definicja.

Mówimy, że zmienne losowe X i Y są **nieskorelowane** jeżeli

$$\text{Cov}(X, Y) = 0.$$

Wniosek

Jeżeli zmienne losowe X i Y są niezależne oraz istnieje ich kowariancja, to są one nieskorelowane.

Twierdzenie odwrotne nie jest prawdziwe.

Przykład

Niech $X \sim N(0, 1)$ oraz $Y = X^2$. Nietrudno sprawdzić, że $\text{Cov}(X, Y) = E(X^3) - E(X) \cdot E(X^2) = 0$. Jednocześnie jest oczywiste, że zmienne losowe X i X^2 są zależne.

Własności współczynnika korelacji

Dla dowolnych zmiennych losowych X i Y , dla których istnieje współczynnik korelacji ρ , mamy

1. $|\rho_{XY}| \leq 1$.
2. Jeżeli X i Y są niezależne, to $\rho_{XY} = 0$.
3. $|\rho_{XY}| = 1$ wtedy i tylko wtedy, gdy istnieją takie $a \neq 0$ oraz b , że $P(Y = aX + b) = 1$.
4. $\forall a \neq 0, c \neq 0$ i $\forall b, d$ mamy

$$|\text{Corr}(X, Y)| = |\text{Corr}(aX + b, cY + d)|.$$

Dowód.

1. Niech X^* i Y^* oznaczają odpowiednio standaryzowane zmienne X i Y . Ponieważ wariancja dowolnej zmiennej losowej jest nieujemna, to

$$0 \leq V(X^* + Y^*) = V(X^*) + V(Y^*) + 2 \cdot \text{Cov}(X^*, Y^*) = 2 + 2\rho_{XY}.$$

Zatem $\rho_{XY} \geq -1$. Podobnie z warunku $0 \leq V(X^* - Y^*)$ otrzymujemy, że $\rho_{XY} \leq 1$.

2. Wynika z zerowania się kowariancji.
3. Załóżmy, że $\rho = 1$. Wówczas $V(X^* - Y^*) = 0$ tym samym rozkład $X^* - Y^*$ jest zdegenerowany.
Podobnie dla $\rho = -1$ rozkład $X^* + Y^*$ jest zdegenerowany.
Założmy, że dla pewnych $a \neq 0$ oraz b

$$P(Y = aX + b) = 1.$$

Stąd $E(Y) = E(aX + b) = aE(X) + b$, $V(Y) = a^2V(X)$ oraz

$$Cov(X, aX + b) = aCov(X, X) = aV(X).$$

Tym samym, otrzymujemy, że

$$\rho_{XY} = \frac{aV(X)}{|a|V(X)} = \text{sign}(a),$$

oraz $|\rho| = 1$.

4. Niezmienniczość współczynnika korelacji przy transformacjach liniowych zmiennych X i Y jest prostą konsekwencją własności wartości oczekiwanej oraz wariancji kombinacji liniowej zmiennych losowych.

■

W szczególności trzecia własność mówi, że współczynnik korelacji jest miarą liniowej zależności pomiędzy X i Y .

Wprowadzone pojęcia kowariancji i korelacji oraz ich własności można uogólnić na przypadek n -wymiarowy.

Twierdzenie.

Jeżeli X_1, X_2, \dots, X_n są zmiennymi losowymi, dla których istnieją kowariancje $Cov(X_i, X_j)$ dla dowolnych i i j , oraz a_1, a_2, \dots, a_n są dowolnymi rzeczywistymi stałymi, to

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$$

oraz

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i) + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} Cov(X_i, X_j).$$

Jeżeli dodatkowo założymy, że X_1, X_2, \dots, X_n są parami niezależne, to

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i).$$

Przykład

Niech X_1, X_2, \dots, X_n będą niezależnymi zmiennymi losowymi o jednakowych rozkładach z wartością oczekiwaną $E(X_i) = \mu$ oraz wariancją $V(X_i) = \sigma^2$, $i = 1, 2, \dots, n$.

Zdefiniujmy średnią arytmetyczną zmiennych losowych X_1, X_2, \dots, X_n

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Wówczas

$$E(\bar{X}) = \mu \quad \text{oraz} \quad V(\bar{X}) = \frac{\sigma^2}{n}.$$

Jest to prosta konsekwencja poprzedniego twierdzenia.

21 Macierz kowariancji

Oznaczmy przez \underline{X} wektor kolumnowy złożony ze zmiennych losowych X_1, X_2, \dots, X_n .

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Zdefiniujmy wartość oczekiwaną wektora losowego \underline{X} jako wektor wartości oczekiwanych

$$E[\underline{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}.$$

Używając notacji macierzowej otrzymujemy, że

$$E[\underline{a}^T \cdot \underline{X}] = \underline{a}^T \cdot E[\underline{X}],$$

gdzie $\underline{a}^T = [a_1, a_2, \dots, a_n]$ jest n -wym. wektorem rzeczywistym.

Definicja.

Macierz kowariancji wektora losowego \underline{X} nazywamy macierz C utworzoną z kowariancji zmiennych losowych X_i i X_j , tzn.

$$C = \begin{bmatrix} V(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & V(X_2) & \cdots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \cdots & V(X_n) \end{bmatrix}$$

Macierz kowariancji wektora los. \underline{X} oznaczamy zwykle jako $\mathcal{D}(\underline{X})$ lub Σ .

Własności macierzy kowariancyjnej

- (1) **Macierz kowariancyjna jest macierzą symetryczną, tzn.**

$$C^T = C,$$

ponieważ dla dowolnych i i j , $Cov(X_i, X_j) = Cov(X_j, X_i)$.

- (2) **Macierz kowariancyjna jest nieujemnie określona.**

Istotnie, dla dowolnego wektora stałych $\underline{a}^T = [a_1, a_2, \dots, a_n]$ mamy

$$\begin{aligned} 0 \leq V(\underline{a}^T \underline{X}) &= V\left(\sum_{i=1}^n a_i X_i\right) = \\ &= \sum_{i=1}^n a_i^2 V(X_i) + 2 \sum_{i=2}^n \sum_{j=1}^{i-1} Cov(X_i, X_j) = \\ &= \underline{a}^T C \underline{a}. \end{aligned}$$

Zauważmy, że $\underline{a}^T C \underline{a}$ jest formą kwadratową.

- (3) **Warunek $\det(C) = 0$ jest warunkiem koniecznym i dostatecznym na to, aby istniały takie stałe a_1, a_2, \dots, a_n, b , $a_1^2 + a_2^2 + \dots + a_n^2 > 0$, że**

$$P(\underline{a}^T \underline{X} = b) = 1,$$

tzn. zm. los. X_1, X_2, \dots, X_n są liniowo zależne.

$\det(C) = 0$ oznacza, że macierz C jest osobliwa, a forma kwadratowa $\underline{a}^T C \underline{a}$ nie jest dodatnio określona. Wynika stąd, że dla pewnego niezerowego wektora stałych \underline{a} wariancja $V(\underline{a}^T \underline{X}) = 0$, co jest możliwe tylko wówczas, gdy rozkład zmiennej losowej $\underline{a}^T \underline{X}$ jest zdegenerowany.

- (4) **Macierz kowariancyjna jest macierzą diagonalną tylko dla zmiennych nieskorelowanych.**
- (5) **W notacji macierzowej macierz kowariancji jest równa**

$$C = E \left[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^T \right].$$

- (6) **Macierz kowariancyjna C_Y zmiennej losowej $\underline{Y} = A\underline{X}$, gdzie A jest $m \times n$ macierzą ($m \leq n$), jest**

$$C_Y = A C A^T.$$

Istotnie, jest to prosta konsekwencja własności poprzedniej.

$$\begin{aligned} C_Y &= E[(A\underline{X} - AE[\underline{X}])(A\underline{X} - AE[\underline{X}])^T] = \\ &= E[A(\underline{X} - E[\underline{X}])(A(\underline{X} - E[\underline{X}]))^T] = \\ &= AE[(\underline{X} - E[\underline{X}])(\underline{X} - E[\underline{X}])^T] A^T = ACA^T. \end{aligned}$$

Przykład

Niech (X, Y) będzie dwuwymiarową zmienną losową, wówczas wyznacznik macierzy kowariancyjnej C jest równy

$$\begin{aligned} \det(C) &= \begin{vmatrix} V(X) & Cov(X, Y) \\ Cov(Y, X) & V(Y) \end{vmatrix} = \\ &= V(X)V(Y) - Cov^2(X, Y) = \\ &= V(X)V(Y)(1 - \rho^2). \end{aligned}$$

Zatem $\det(C) = 0$ wtedy i tylko wtedy, gdy $\rho^2 = 1$. Oznacza to, że istnieje zależność liniowa pomiędzy zmiennymi, tzn. istnieją takie $a \neq 0$ oraz b , że $P(Y = aX + b) = 1$.

22 Wielowymiarowy rozkład normalny

Niech $\underline{X} = (X_1, X_2, \dots, X_n)^T$ będzie wektorem losowym o wartości oczekiwanej $E[\underline{X}] = \underline{\mu}$ i macierzy kowariancyjnej $\mathcal{D}(\underline{X}) = C$.

Założmy, że $\det(C) > 0$, tzn. macierz kowariancji jest nieosobliwa. Oznaczmy przez $L = C^{-1}$ macierz odwrotną do macierzy kowariancji C .

Definicja.

Mówimy, że wektor losowy \underline{X} ma rozkład normalny o wektorze wartości oczekiwanych $\underline{\mu}$ i macierzy kowariancji C , $C > 0$ (oznaczenie $\underline{X} \sim N_n(\underline{\mu}, C)$) jeżeli jego gęstość ma postać

$$f(x_1, \dots, x_n) = \frac{\sqrt{\det(L)}}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n l_{ij}(x_i - \mu_i)(x_j - \mu_j) \right\},$$

gdzie $\mu_i = E(X_i)$ oraz $L = (l_{ij})$.

Twierdzenie.

Jeżeli wektor losowy \underline{X} ma rozkład normalny, dla którego zmienne losowe X_i , $i = 1, 2, \dots, n$ są nieskorelowane, to X_i są niezależne.

Dowód.

Ponieważ X_i są nieskorelowane, to macierz kowariancji jest diagonalna oraz

$$C = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2),$$

gdzie $\sigma_i^2 = V(X_i)$, $i = 1, 2, \dots, n$. Zatem

$$L = \text{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \dots, \frac{1}{\sigma_n^2}\right),$$

oraz $l_{ij} = 0$ for all $i \neq j$. Ponieważ

$$\sqrt{\det(L)} = \frac{1}{\prod_{i=1}^n \sigma_i}$$

mamy

$$\begin{aligned} f(x_1, \dots, x_n) &= \frac{1}{(2\pi)^{n/2} \prod_{i=1}^n \sigma_i} \exp\left\{-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu_i)^2}{\sigma_i^2}\right\} = \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left\{-\frac{1}{2} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2\right\} = \prod_{i=1}^n f_{X_i}(x_i), \end{aligned}$$

co oznacza, że rozkłady brzegowe X_i są rozkładami normalnymi $N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, n$. ■

Wynika stąd, że dla zmiennych normalnych X_1, X_2, \dots, X_n brak korelacji jest równoważny niezależności.

Przykład

Niech (X, Y) będzie dwuwymiarową zmienną losową o rozkładzie normalnym. Niech $E[X] = \mu_X$, $E[Y] = \mu_Y$, $V(X) = \sigma_X^2$, $V(Y) = \sigma_Y^2$ a współczynnik korelacji ρ . Wówczas macierz kowariancji ma postać

$$C = \begin{bmatrix} V(X) & \text{Cov}(X, Y) \\ \text{Cov}(Y, X) & V(Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

Wówczas

$$L = \frac{1}{1 - \rho^2} \begin{bmatrix} \frac{1}{\sigma_X^2} & -\frac{\rho}{\sigma_X\sigma_Y} \\ -\frac{\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{bmatrix}$$

i dwuwymiarowa gęstość jest równa

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right\}.$$

Całkując gęstość łączną względem zmiennej y otrzymujemy gęstość brzegową X . Wynika stąd, że $X \sim N(\mu_X, \sigma_X^2)$

Podobnie $Y \sim N(\mu_Y, \sigma_Y^2)$.

Powyższy rezultat można uogólnić na przypadek n -wymiarowy.

Twierdzenie.

Jeżeli wektor losowy $\underline{X} \sim N_n(\underline{\mu}, C)$, gdzie $C > 0$ a $\underline{\mu} \in R^n$ to dla dowolnej macierzy $A_{m \times n}$ zmienna losowa $\underline{Y} = A\underline{X}$ ma rozkład normalny

$$\underline{Y} \sim N_m(A\underline{\mu}, ACA^T).$$

W szczególności, jeżeli A jest rzutem, to otrzymujemy, że wszystkie rozkłady brzegowe zmiennej losowej o rozkładzie normalnym mają rozkłady normalne.

Przykład

Jeżeli X_1 i X_2 są niezależnymi zmiennymi losowymi o rozkładzie $N(0, 1)$, to zmienne losowe $Y_1 = X_1 + X_2$ oraz $Y_2 = X_1 - X_2$ są niezależne o rozkładach normalnych.

Rozwiązanie.

$\underline{X} \sim N_2(\underline{0}, I_2)$, gdzie I_2 jest macierzą jednostkową a $\underline{0} = [0, 0]^T$.

$$\underline{Y} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

Zatem macierz kowariancji wektora \underline{Y} jest również diagonalna

$$C_Y = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

co oznacza niezależność Y_1 i Y_2 . Dokładniej, mamy

$$\underline{Y} \sim N_2(\underline{0}, C_Y).$$

Twierdzenie.

Jeżeli zmienne losowe X_1, X_2, \dots, X_n są niezależne o jednakowym rozkładzie $N(0, 1)$ a macierz A jest macierzą ortonormalną ($AA^T = I_n$), to zmienne losowe Y_1, Y_2, \dots, Y_n , gdzie $\underline{Y} = A\underline{X}$, są również niezależne o jednakowym rozkładzie $N(0, 1)$.

Dowód.

Ponieważ $\underline{X} \sim N(\underline{0}, I_n)$, to

$$\underline{Y} = A\underline{X} \sim N(A\underline{0}, AI_nA^T) \equiv N(\underline{0}, I_n).$$

■

23 Zagadnienie regresji liniowej

Założmy, że (X, Y) jest zmienną losową dwuwymiarową o znanej łącznej funkcji prawdopodobieństwa lub gęstości. Chcemy w najlepszy (w pewnym sensie) sposób wyznaczyć zależność liniową pomiędzy Y i X , tzn. znając wartości zmiennej losowej X chcemy móc przewidywać wartości zmiennej Y .

Jako kryterium wyboru, przyjmiemy minimum **błędu średniokwadratowego**, tzn. chcemy znaleźć taką funkcję liniową $h(x) = a \cdot x + b$, że wyrażenie

$$E \left\{ [Y - (aX + b)]^2 \right\}.$$

osiąga minimum po wszystkich możliwych wyborach a i b . Funkcję h nazywamy **funkcją regresji liniowej** zmiennej Y przy $X = x$.

Twierdzenie.

Jeżeli (X, Y) jest dwuwymiarową zmienną losową o znanej łącznej funkcji prawdopodobieństwa lub gęstości, dla której istnieją wszystkie momenty rzędu drugiego (σ_X^2 , σ_Y^2 oraz $Cov(X, Y)$) to wyrażenie

$$E \left\{ [Y - (aX + b)]^2 \right\}.$$

osiąga minimum wtedy i tylko wtedy, gdy

$$a = \rho_{XY} \frac{\sigma_Y}{\sigma_X} \quad \text{oraz} \quad b = m_Y - a \cdot m_X,$$

tzn. funkcja regresji liniowej zmiennej Y przy $X = x$ jest równa

$$\frac{y - m_Y}{\sigma_Y} = \rho_{XY} \frac{x - m_X}{\sigma_X}.$$

Dowód.

Dla dowolnych wartości rzeczywistych a i b , zdefiniujmy funkcję

$$S(a, b) = E \left\{ [Y - (aX + b)]^2 \right\}.$$

Wówczas

$$\begin{aligned} S(a, b) &= E \left\{ [(Y - m_Y) - a\{X - m_X\} + (m_Y - am_X - b)]^2 \right\} = \\ &= E \left\{ (Y - m_Y)^2 + a^2(X - m_X)^2 + (m_Y - am_X - b)^2 + \right. \\ &\quad \left. - 2a(X - m_X)(Y - m_Y) - 2a(X - m_X)(m_Y - am_X - b) + \right. \\ &\quad \left. + 2(Y - m_Y)(m_Y - am_X - b) \right\} = \\ &= \sigma_Y^2 + a^2\sigma_X^2 + (m_Y - am_X - b)^2 - 2aCov(X, Y). \end{aligned}$$

Ponieważ $S(a, b)$ jest funkcją kwadratową od a i b , to do wyznaczenia minimum niezbędne jest rozwiązanie następującego układu równań

$$\begin{cases} \frac{\partial S(a,b)}{\partial a} = 2a \cdot \sigma_X^2 - 2m_X(m_Y - am_X - b) - 2\rho_{XY}\sigma_X\sigma_Y = 0 \\ \frac{\partial S(a,b)}{\partial b} = -2(m_Y - am_X - b) = 0 \end{cases}$$

Z drugiego równania mamy $b = m_Y - a \cdot m_X$. Podstawiając b w pierwszym równaniu mamy $a = \rho_{XY} \frac{\sigma_Y}{\sigma_X}$, co kończy dowód. ■

Podobnie uzyskujemy linię regresji X przy $Y = y$,

$$\frac{x - m_X}{\sigma_X} = \rho_{XY} \frac{y - m_Y}{\sigma_Y}.$$

Dla $|\rho_{XY}| = 1$ obie linie się pokrywają.

Przykład

Założmy, że X, Y ma dyskretny rozkład jednostajny na zbiorze $\mathcal{X} \times \mathcal{Y} = \{(x_i, y_i), i = 1, 2, \dots, n\}$, tzn.

$$P(X = x_i, Y = y_i) = \frac{1}{n}.$$

Wyznaczyć linię regresji zmiennej Y dla $X = x$.

Rozwiązanie.

$$m_X = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}, \quad m_Y = \frac{1}{n} \sum_{i=1}^n y_i = \bar{Y},$$

$$\sigma_X^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{X})^2, \quad \sigma_Y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 - (\bar{Y})^2,$$

oraz

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - \bar{X} \cdot \bar{Y}.$$

Stąd linia regresji Y dla $X = x$ ma postać

$$y = \bar{Y} + \frac{\sum_{i=1}^n x_i \cdot y_i - n\bar{X} \cdot \bar{Y}}{\sum_{i=1}^n x_i^2 - n(\bar{X})^2} (x - \bar{X})$$

lub

$$y = \hat{\alpha} + \hat{\beta} \cdot x,$$

gdzie

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i \cdot y_i - n\bar{X} \cdot \bar{Y}}{\sum_{i=1}^n x_i^2 - n(\bar{X})^2} \quad \text{and} \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \cdot \bar{X}.$$

W statystyce jest to znany problem **prostej regresji liniowej** wyznaczenia zależności liniowej pomiędzy dwiema zmiennymi na podstawie n par obserwacji.

Przykład dopasowania linii regresji do n punktów.

