

ZMIENNA LOSOWA I JEJ PARAMETRY-powtórzenie

(Ω, S, P) – przestrzeń probabilistyczna

(matematyczny model zjawiska losowego),

Ω – zbiór wszystkich zdarzeń elementarnych,

S – zbiór zdarzeń, (podzbiory zbioru Ω , (dokładnie σ – ciało podzbiorów)),

P – prawdopodobieństwo (funkcja przyporządkowująca zdarzeniom szansę ich zajścia).

$$P : S \rightarrow R$$

Zmienną losową X nazywamy funkcję (borelowską czyli praktycznie każdą) przyporządkowującą zdarzeniom elementarnym liczby rzeczywiste.

$$X : \Omega \longrightarrow R$$

Dystrybuantą zmiennej losowej X nazywamy funkcję $F: R \longrightarrow R$ określoną wzorem:

$$F(x) = P(X < x) = P_x((-\infty, x))$$

Własności dystrybuanty:

a) F jest funkcją niemalejącą,

b) F jest funkcją lewostronnie ciągłą,

c) $F(-\infty) = 0$; $F(\infty) = 1$,

d) dystrybuenta zmiennej losowej wyznacza jednoznacznie jej rozkład,

e) $P(a \leq X < b) = F(b) - F(a)$; $a < b$

f) $P(X = a) = F(a^+) - F(a)$; gdzie $F(a^+)$ oznacza granicę prawostronną, (jeśli a jest punktem ciągłości dystrybuanty to $P(X = a) = 0$).

Zmienna losowa jest **skokowa (dyskretna)** jeśli zbiór wszystkich jej wartości jest skończony lub przeliczalny.

Rozkład zmiennej losowej skokowej często określamy za pomocą **funkcji prawdopodobieństwa**:

$$P(X = x_k) = p_k$$

$$(\text{własność: } \sum_k p_k = 1; p_k > 0)$$

Liczby p_k nazywamy **skokami**, a wartości x_k **punktami skokowymi**.

Zmienna losowa X o dystrybuancie F jest **ciągła** jeśli jej dystrybuenta da się przedstawić w postaci

$$F(x) = \int_{-\infty}^x f(t)dt \quad x \in R$$

gdzie f jest funkcją spełniającą warunki:

$$f(x) \geq 0; \quad x \in R; \quad \int_{-\infty}^{\infty} f(t)dt = 1$$

i nazywamy ją **gęstością prawdopodobieństwa** zmiennej losowej X .

Własności zmiennej losowej ciągłej:

$$a) P(X < a) = \int_{-\infty}^a f(x)dx = F(a),$$

$$b) P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = \\ = P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a)$$

$$c) P(X = a) = 0, \text{ dla dowolnego } a \in R; \text{ (brak punktów skokowych),}$$

d) F jest funkcją ciągłą i prawie wszędzie różniczkowalną $F'(x) = f(x)$ (równość zachodzi dla punktów ciągłości gęstości). Wyznaczając gęstość przez różniczkowanie dystrybuanty, w punktach w których F nie jest różniczkowalna można przyjąć, że gęstość jest równa zero.

Własności rozkładu zmiennej losowej często charakteryzujemy jej **parametrami**.

Jednym z podstawowych parametrów jest wartość oczekiwana.

Wartość oczekiwana. Oznaczenie EX lub m .

Dla zmiennej losowej skokowej

$$EX = \sum_i x_i p_i$$

(jeśli ewentualny szereg jest zbieżny bezwzględnie, takie szeregi są "odporne" np. na zmianę kolejności wyrazów).

Dla zmiennej losowej ciągłej

$$EX = \int_{-\infty}^{\infty} xf(x)dx$$

(jeśli ewentualna całka niewłaściwa jest zbieżna bezwzględnie).

Przykład

Dla zmiennej losowej o funkcji prawdopodobieństwa

x_k	-1	2	3
p_k	0,2	0,6	0,2

$$EX = -1 \cdot 0,2 + 2 \cdot 0,6 + 3 \cdot 0,2 = 1,6.$$

Przykład

Dla zmiennej losowej o gęstości

$$f(x) = \begin{cases} 2x & x \in <0,1> \\ 0 & x \notin <0,1> \end{cases}$$

$$EX = \int_0^1 x \cdot 2x dx = 2 \int_0^1 x^2 dx = 2 \frac{x^3}{3} \Big|_0^1 = \frac{2}{3}$$

Własności wartości oczekiwanej

$$a) Ec = c; \quad c - \text{stała},$$

$$b) E(aX) = aE(X),$$

$$c) E(X + Y) = EX + EY,$$

$$d) \text{ Jeśli } a \leq X \leq b, \text{ to } a \leq EX \leq b, \text{ jeśli } X \leq Y, \text{ to } EX \leq EY,$$

$$e) \quad EX \leq E|X|, \quad |EX| \leq E|X|$$

f) X, Y – niezależne, to $E(XY) = EX \cdot EY$.

Miarą rozrzutu wartości zmiennej losowej jest wariancja.

Wariancja. Oznaczenie D^2X lub σ^2 .

$$D^2X = E(X - EX)^2$$

$$\text{Dla zmiennej losowej skokowej} \quad D^2X = \sum (x_i - EX)^2 p_i$$

$$\text{Dla zmiennej losowej ciągłej} \quad D^2X = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx$$

Własności wariancji

$$a) \quad D^2c = 0; \quad c - \text{stała},$$

$$b) \quad D^2(aX) = a^2 D^2(X),$$

$$c) \quad D^2(X + b) = D^2X, \quad b - \text{stała},$$

$$d) \quad X, Y - \text{niezależne, to } D^2(X \pm Y) = D^2X + D^2Y$$

$$e) \quad D^2X = E(X^2) - (EX)^2.$$

Uzasadnienie e)

$$\begin{aligned} D^2X &= E(X - EX)^2 = E(X^2 - 2XEX + (EX)^2) = EX^2 - 2EXEX + (EX)^2 = \\ &= E(X^2) - (EX)^2. \end{aligned}$$

Jeśli rozrzut wartości zmiennej losowej chcemy (np. z powodu interpretacji w zastosowaniach) mierzyć w tych samych jednostkach co X to stosujemy odchylenie standardowe.

Odchylenie standardowe. Oznaczenie DX lub σ .

$$DX = \sqrt{D^2X}$$

Podstawowe rozkłady.

Rozkład dwupunktowy (zerojedynkowy)

Niech $p \in (0, 1)$ będzie ustaloną liczbą. Określamy:

$$P(X = 0) = q, \quad P(X = 1) = p; \quad \text{gdzie } q = 1 - p.$$

Rozkład ten jest wykorzystywany w statystycznej kontroli jakości. Można np. przyjąć, że $X = 0$ gdy wyrób dobry, $X = 1$ gdy wyrób jest wadliwy, wtedy $p = P(X = 1)$ traktujemy jako wadliwość wyrobu.

Rozkład dwumianowy

Dla danych $p \in (0, 1)$, $n \in N$ określamy funkcję prawdopodobieństwa

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \quad \text{gdzie } q = 1 - p$$

$$k = 0, 1, 2, \dots, n.$$

Zauważmy, że gdy $n = 1$ to rozkład dwumianowy jest rozkładem zerojedynkowym.

Jeśli przyjmiemy, że n oznacza liczbę niezależnych doświadczeń z których każde kończy się jednym z dwóch wyników: „sukcesem” (z prawdopodobieństwem p w każdym doświadczeniu) lub „porażką” i zmienna losowa X oznacza liczbę „sukcesów” to powyższy wzór wyznacza **prawdopodobieństwo uzyskania dokładnie k sukcesów w n doświadczeniach** (próbach).

Rozkład Poissona

Dla $\lambda > 0$ określamy funkcję prawdopodobieństwa

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad k = 0, 1, 2, \dots$$

(wartości tych prawdopodobieństw zawiera tablica rozkładu Poissona)

Rozkład Poissona (możliwość odczytu w tablicy) może dla dużych n (praktycznie $n \geq 30$) i małych p (praktycznie $p \leq 0,2$) przybliżać rozkład dwumianowy (przybliżenie Poissona)

$$\binom{n}{k} p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{gdzie } \lambda = n \cdot p$$

Rozkłady ciągłe

Rozkład jednostajny

Rozkład którego gęstość jest stała w pewnym przedziale nazywamy **jednostajnym**.

$$\text{Gęstość rozkładu jednostajnego w } (a, b) \quad f(x) = \begin{cases} \frac{1}{b-a} & x \in (a; b) \\ 0 & x \notin (a; b) \end{cases}$$

Ponieważ gęstość ta ma oś symetrii w punkcie $x = (a+b)/2$ to

$$\begin{aligned} EX &= (a+b)/2 \\ D^2X &= (b-a)^2/12 \end{aligned}$$

Rozkład wykładniczy

Rozkład ten występuje często w zagadnieniach rozkładu czasu między zgłoszeniami (awariami) lub czasu oczekiwania na obsługę w systemach kolejkowych.

Gęstość rozkładu wykładniczego o parametrze $a > 0$ ma postać

$$f(x) = \begin{cases} ae^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

dystrybuantą tego rozkładu jest funkcja

$$F(x) = \begin{cases} 1 - e^{-ax} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

(uzasadnienie: $F'(x) = f(x)$)

Własność.

- 1) Jeśli liczba zgłoszeń w systemie kolejkowym w przedziale czasu $(t, t + T)$ ma rozkład Poissona o parametrze λT , oraz liczby zgłoszeń przychodzące w rozłącznych przedziałach czasu są niezależne to czas X między kolejnymi zgłoszeniami ma rozkład wykładniczy o parametrze $a = 1/\lambda$.
- 2) Dla dowolnych $t, T > 0$ mamy
 $P(X \geq t + T \mid X \geq t) = P(X \geq T)$ (własność braku pamięci)

Uzasadnienie.

$$P(X \geq t+T | X \geq t) = \frac{P(X \geq t+T \wedge X \geq t)}{P(X \geq t)} = \frac{P(X \geq t+T)}{P(X \geq t)} =$$

$$= \frac{e^{-(t+T)a}}{e^{-ta}} = e^{-Ta} = P(X \geq T)$$

Jest to jedyny rozkład ciągły o tej własności.

Dyskretnym odpowiednikiem rozkładu wykładniczego jest rozkład geometryczny.

Rozkład normalny

Dla $m \in R, \sigma \in (0, +\infty)$

Określamy gęstość rozkładu

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in R$$

Wartości dystrybuanty dla argumentów ujemnych wyznaczamy na podstawie zależności

$$\Phi(-x) = 1 - \Phi(x)$$

Uwaga

Jeśli X ma rozkład $N(m, \sigma)$ to zmienna losowa $Y = (X - m)/\sigma$ ma rozkład $N(0, 1)$ (takie przekształcenie nazywamy **standaryzacją**).

ROZKŁAD ZMIENNEJ LOSOWEJ N WYMIAROWEJ. CIĄGI LOSOWE

(Ω, S, P) - ustalona przestrzeń probabilistyczna.

$X = (X_1, X_2, \dots, X_n)$ - zmienna losowa n - wymiarowa (wektor losowy, ciąg losowy).

$X : \Omega \rightarrow R^n$ (funkcja borelowska)

$P_X : B(R^n) \rightarrow [0, 1]$ - rozkład zmiennej losowej X .

Dystrybuanta

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n)$$

X nazywamy **zmienną losową skokową** jeśli jej zbiór wartości jest skończony lub przeliczalny.

X nazywamy **zmienną losową ciągłą** jeśli jej dystrybuanta da się przedstawić w postaci

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(u_1, \dots, u_n) du_1 \dots du_n$$

dla pewnej nieujemnej funkcji f zwanej gęstością.

Uwaga.

1. W punktach ciągłości funkcji f zachodzi:

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

2. Dla $A \in \mathcal{B}(R^n)$ mamy $P_X(A) = \int \int \dots \int_A f(x_1, \dots, x_n) dx_1 \dots dx_n$.

Funkcja charakterystyczna zmiennej losowej n - wymiarowej.

$$\varphi(t) = \varphi(t_1, \dots, t_n) = E(e^{itX}) = E(\exp(i(t_1 X_1 + \dots + t_n X_n))).$$

Rozkłady warunkowe.

Jeśli $P_{1, \dots, k}(X_1 = x_{1j}, \dots, X_k = x_{kj}) > 0$ to rozkład zmiennej losowej skokowej $(n - k)$ wymiarowej określonej wzorem:

$$P(X_{k+1} = x_{k+1,j}, \dots, X_n = x_{nj} \mid X_1 = x_{1j}, \dots, X_k = x_{kj}) = \frac{P(X_1 = x_{1j}, \dots, X_n = x_{nj})}{P_{1, \dots, k}(X_1 = x_{1j}, \dots, X_k = x_{kj})}$$

nazywamy rozkładem warunkowym zmiennej losowej $(X_{k+1}, \dots, X_n)$ pod warunkiem, że $(X_1 = x_{1j}, \dots, X_k = x_{kj})$.

Jeśli gęstość $f_{1, \dots, k} > 0$ to rozkład zmiennej losowej ciągłej $(n - k)$ wymiarowej określonej wzorem:

$$f(x_{k+1}, \dots, x_n \mid x_1, \dots, x_k) = \frac{f(x_1, \dots, x_n)}{f(x_1, \dots, x_k)}$$

nazywamy rozkładem warunkowym zmiennej losowej $(X_{k+1}, \dots, X_n)$ pod warunkiem, że $(X_1 = x_1, \dots, X_k = x_k)$.

Niezależność zmiennych losowych.

Zmienne losowe $X_1, X_2, \dots, X_n$ są niezależne jeśli

$$F(x_1, \dots, x_n) = F_1(x_1) \cdot F_2(x_2) \dots F_n(x_n) \text{ dla dowolnych } x_1, x_2, \dots, x_n \in R^n.$$

gdzie F_i - dystrybuanty rozkładów brzegowych jednowymiarowych.

Dla zmiennych losowych skokowych odpowiedni warunek ma postać:

$$P(X_1 = x_{1j}, \dots, X_n = x_{nj}) = P_1(X_1 = x_{1j}) \cdot \dots \cdot P_n(X_n = x_{nj})$$

dla dowolnych $x_{1j}, \dots, x_{nj} \in R^n$

Dla zmiennych losowych ciągłych odpowiedni warunek ma postać:

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \dots f_n(x_n)$$

dla dowolnych $x_1, x_2, \dots, x_n \in R^n$.

Parametry (mogą nie istnieć)

Wartość oczekiwana $E(X) = [EX_1, EX_2, \dots, EX_n]$.

Wariancja $D^2(X) = [D^2 X_1, D^2 X_2, \dots, D^2 X_n]$.

Moment (zwyczajny) rzędu $l_1 + l_2 + \dots + l_n$

$$m_{l_1 l_2 \dots l_n} = E(X_1^{l_1} X_2^{l_2} \dots X_n^{l_n}),$$

Moment centralny rzędu $l_1 + l_2 + \dots + l_n$

$$\mu_{l_1 l_2 \dots l_n} = E((X_1 - EX_1)^{l_1} \dots (X_n - EX_n)^{l_n}),$$

Macierz kowariancji $K = [k_{ij}]$, gdzie
 $k_{ij} = \text{cov}(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)] = E(X_i X_j) - E(X_i)E(X_j)$

Uwaga $k_{ii} = D^2 X_i$, jest wariancją i -tej składowej.

Macierz K jest kwadratowa, symetryczna i słabo dodatnio określona (w szczególności ma wyznacznik nieujemny).

Macierz korelacji $R = [\rho_{ij}]$, gdzie $\rho_{ij} = \frac{\text{cov}(X_i, X_j)}{DX_i \cdot DX_j} =$

Uwaga $\rho_{ii} = 1$.

Wielowymiarowy rozkład Bernoulliego.

Dla danych $k \in \mathbb{N}$, $p = [p_1, p_2, \dots, p_n]^T$ takiego, że $0 \leq \sum_{i=1}^n p_i < 1$ oraz

$i = [i_1, i_2, \dots, i_n]^T$ gdzie $i_j \in \{0, 1, \dots, n\}$ $\sum_{j=1}^n i_j \leq k$ określamy

$$P(X = i) = \frac{k!}{i_0! i_1! i_2! \dots i_n!} p_0^{i_0} \cdot p_1^{i_1} \cdot \dots \cdot p_n^{i_n}$$

gdzie $p_0 = 1 - \sum_{i=1}^n p_i$; $i_0 = k - \sum_{j=1}^n i_j$.

Przykład.

Badanie systemu telekomunikacyjnego polega na wielokrotnych próbach uzyskania połączenia. Rozpatrujemy trzy możliwe wyniki każdego połączenia:

- A_0 - połączenie bez zakłóceń,
- A_1 - połączenie z zakłóceniami,
- A_2 - brak połączenia.

Wiadomo, że $P(A_0) = 0,7$; $P(A_1) = 0,2$; $P(A_2) = 0,1$.

Wykonano 50 prób łączności, obliczyć prawdopodobieństwo tego, że w tych próbach co najwyżej raz nie uzyskamy połączenia i co najwyżej raz uzyskamy połączenie z zakłóceniami.

X - liczba prób z brakiem łączności,

Y - liczba prób z połączeniami z zakłóceniami.

$$P(X = i, Y = j) = \frac{50!}{i! j! (50 - i - j)!} 0,7^{50-i-j} 0,2^i 0,1^j \quad \text{gdzie } i, j = 0, 1, \dots, 50; i + j \leq 50.$$

Zatem

$$P(X \leq 1, Y \leq 1) = P(X = 0, Y = 0) + P(X = 0, Y = 1) + P(X = 1, Y = 0) + P(X = 1, Y = 1) = 0,7^{50} + 50 \cdot 0,1 \cdot 0,7^{49} + 50 \cdot 0,2 \cdot 0,7^{49} + 50 \cdot 49 \cdot 0,1 \cdot 0,2 \cdot 0,7^{48} \approx 0,0000022.$$

Wielowymiarowy rozkład wielomianowy.

Jeśli w definicji rozkładu Bernoulliego mamy $p_0 = 0$, $\sum_{j=1}^n i_j = k$ to otrzymany rozkład nazywamy rozkładem wielomianowym.

Wielowymiarowy rozkład Poissona.

Dla danego $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$ oraz

$i = [i_1, i_2, \dots, i_n]^T$ gdzie $i_j \in \{0, 1, \dots, n\}$ określamy

$$P(X = i) = \frac{\lambda_1^{i_1}}{i_1!} \cdot \dots \cdot \frac{\lambda_n^{i_n}}{i_n!} e^{-\lambda_0} \quad \text{gdzie } \lambda_0 = \sum_{i=1}^n \lambda_i.$$

Rozkład normalny n - wymiarowy.

K - macierz kowariancyjna, niech $\det K \neq 0$.

Zmienna losowa n - wymiarowa ma rozkład normalny n - wymiarowy gdy gęstość tej zmiennej losowej wyraża się wzorem:

$$\begin{aligned} f(x) = f(x_1, x_2, \dots, x_n) &= \frac{\sqrt{|L|}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{j,k=1}^n l_{jk} (x_j - m_j)(x_k - m_k)\right) = \\ &= \frac{\sqrt{|L|}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} (x - m)^T L (x - m)\right) \end{aligned}$$

gdzie

$$m_i = E(X_i) \quad \text{dla } i = 1, 2, \dots, n$$

$L = [l_{jk}]$ j, k = 1, 2, ..., n jest macierzą odwrotną do K.

Dla $n = 2$ warunek $|K| \neq 0$ jest równoważny warunkowi $\rho^2 \neq 1$.

Ponieważ macierz K ma wtedy postać $K = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$ to

$$L = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

Zatem gęstość rozkładu normalnego 2-wymiarowego $N(m_1, m_2, \sigma_1, \sigma_2, \rho)$ można zapisać następująco:

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho \frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2} \right] \right\}$$

Powyższa funkcja gęstości ma stałą wartość $f(x, y) = h$ na elipsie:

$$\frac{(x-m_1)^2}{\sigma_1^2} - 2\rho \frac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \frac{(y-m_2)^2}{\sigma_2^2} = \text{const} = \lambda^2$$

o środku w punkcie (m_1, m_2) .

gdzie $\lambda^2 = -2(1-\rho^2) \ln(h2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})$.

Dla $\rho \neq 0$ osie główne mają równania:

$$y - m_2 = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2 \pm \sqrt{(\sigma_1^2 - \sigma_2^2)^2 + 4\rho^2\sigma_1^2\sigma_2^2}} (x - m_1)$$

Dla $\rho = 0$ osie rozpatrywanej elipsy są równoległe do osi układu współrzędnych.

Zauważmy, że gdy $\rho^2 \rightarrow 1$ to jedna oś się wydłuża, a druga skraca, zależność między zmiennymi staje się ściśle liniowa.

Osie powyższej elipsy tworzą z osią OX kąty α i $\alpha + \pi/2$ gdzie

$$\text{tg } 2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}$$

Funkcja charakterystyczna:

$$\varphi(t) = \exp\left(im^T t - \frac{1}{2} t^T K t\right)$$

gdy $n = 2$ to

$$\varphi(t_1, t_2) = \exp\left(i(t_1 m_1 + t_2 m_2) - \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)\right)$$

Twierdzenie.

Dowolny rozkład brzegowy normalnego rozkładu n -wymiarowego jest rozkładem normalnym.

Twierdzenie.

Jeśli składowe normalnego rozkładu n -wymiarowego są parami nieskorelowane to są niezależne.

Zbieżność ciągów losowych

Zbieżność ciągu zmiennych losowych z prawdopodobieństwem 1 (prawie na pewno)

Ciąg zmiennych losowych (X_n) jest zbieżny do zmiennej losowej X z prawdopodobieństwem 1 jeśli

$$P\left(\left\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

Średniokwadratowa zbieżność ciągu zmiennych losowych

Ciąg zmiennych losowych (X_n) jest średniokwadratowo zbieżny do zmiennej losowej X jeśli

$$\lim_{n \rightarrow \infty} E\left(|X_n - X|^2\right) = 0$$

Rozpatrując ten rodzaj zbieżności zakładamy, że dla występujących tu zmiennych losowych (X_n) , X istnieje skończony moment rzędu 2.

Niekiedy stosuje się zapis $\text{l.i.m. } X_n = X$ (skrót od „limit in mean”).

Stochastyczna zbieżność ciągu zmiennych losowych

Ciąg zmiennych losowych (X_n) jest stochastycznie (wg prawdopodobieństwa) zbieżny do zmiennej losowej X jeśli

$$\bigwedge_{\varepsilon > 0} \lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$$

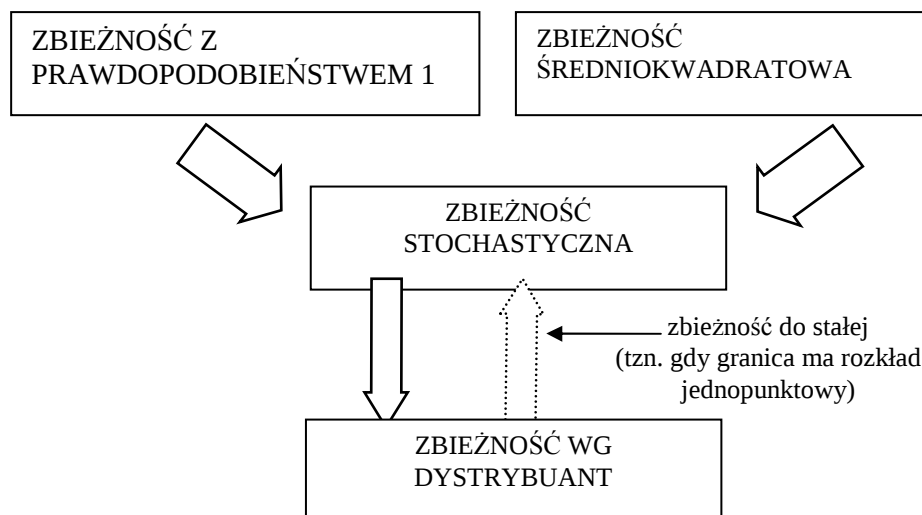
lub równoważnie

$$\bigwedge_{\varepsilon > 0} \lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$$

Zbieżność ciągu zmiennych losowych wg dystrybuant (wg rozkładu)

Ciąg zmiennych losowych (X_n) jest zbieżny do zmiennej losowej X wg dystrybuant jeśli ciąg ich dystrybuant F_n jest zbieżny do dystrybuanty F w każdym punkcie jej ciągłości (F jest dystrybuantą zmiennej losowej X).

Zależności między zbieżnościami.



Przykład.

Rozpatrzmy ciąg zmiennych losowych skokowych określonych na przedziale $[0, 1)$ w następujący sposób

$$X_{kn}(\omega) = \begin{cases} 1 & \text{gdy } \omega \in \left[\frac{k}{n}; \frac{k+1}{n}\right) \\ 0 & \text{gdy } \omega \in [0, 1) - \left[\frac{k}{n}; \frac{k+1}{n}\right) \end{cases}$$

$$P(X_{kn} = 1) = \frac{1}{n}; \quad P(X_{kn} = 0) = 1 - \frac{1}{n}$$

Ciąg $X_{01}, X_{02}, X_{12}, X_{03}, X_{13}, X_{23}, \dots$ jest zbieżny stochastycznie do zera bo

$$\bigwedge_{0 < \varepsilon < 1} \lim_{n \rightarrow \infty} P(|X_n| \geq \varepsilon) = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

Natomiast ciąg ten nie jest zbieżny w żadnym punkcie przedziału $[0, 1)$ bowiem dla każdego ustalonego punktu otrzymujemy rozbieżny ciąg zer i jedynek (zera i jedynki występują na dowolnie dalekich miejscach).

Przykład.

Ciąg zmiennych losowych X_n ciągłych o rozkładach jednostajnych na przedziałach $(0, 1/n)$ jest zbieżny do rozkładu jednopunktowego X ($P(X = 0) = 1$) wg dystrybuant.

Uwaga.

Punktowa granica ciągu dystrybuant nie musi być dystrybuantą.

Jeśli ciąg funkcji charakterystycznych odpowiadających rozpatrywanemu ciągowi dystrybuant jest punktowo zbieżny do funkcji ciągłej to granica tych dystrybuant jest dystrybuantą.

ZADANIA

Zadanie 1.

Dystrybuanta zmiennej losowej (X, Y) wyraża się wzorem:

$$F(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{dla } x > 0, y > 0 \\ 0 & \text{dla innych } x, y \end{cases}$$

Wyznacz gęstość tej zmiennej losowej.

Zadanie 2.

Wyznaczyć wartość parametru c aby funkcja

$$f(x, y) = \begin{cases} c & \text{dla } |x| \leq 2, |y| \leq 2 \\ 0 & \text{dla innych } x, y \end{cases}$$

była gęstością prawdopodobieństwa pewnej zmiennej losowej dwuwymiarowej.

Oblicz a) $P(X < -1, Y > 1)$, b) $P(X > 0, Y > 0)$, c) $P(X < Y)$.

Wyznacz dystrybuantę tej zmiennej losowej.

Wyznacz wektor wartości oczekiwanych tej zmiennej losowej.

Wyznacz macierz kowariancji i macierz korelacji tej zmiennej losowej.

Czy X, Y są niezależne?

Zadanie 3.

Wyznaczyć wartość parametru c aby funkcja

$$f(x, y, z) = \begin{cases} c & \text{dla } |x| \leq 1, |y| \leq 2, |z| \leq 3 \\ 0 & \text{dla innych } x, y, z \end{cases}$$

była gęstością prawdopodobieństwa pewnej zmiennej losowej trójwymiarowej.

Oblicz a) $P(Y > 1)$, b) $P(X > 0, Y > 0, Z > 0)$, c) $P(X < 2Y)$.

Wyznacz wektor wartości oczekiwanych tej zmiennej losowej.

Wyznacz macierz kowariancji i macierz korelacji tej zmiennej losowej.

Zadanie 4.

Wykazać, że macierz $\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ jest macierzą kowariancji.

Zapisz odpowiadającą jej macierz korelacji R .

Zadanie 5.

Sprawdź, że macierz $\begin{bmatrix} 2 & 1 & 2 \\ 1 & 3 & 2 \\ 2 & 2 & 1 \end{bmatrix}$ nie może być macierzą kowariancji.

Zadanie 6

Wyznaczyć $E(X_n)$, $\text{cov}(n,m)$, $D^2(X_n)$, dla ciągu losowego $X_n = An+B$, gdzie A, B to zmienne losowe o parametrach: $EA = 1$; $EB = -1$, $D^2A = 2$, $D^2B = 3$, $\rho = -0,5$.

Zadanie 7

Wyznaczyć $E(X_n)$, $\text{cov}(n,m)$, $D^2(X_n)$, dla ciągu losowego $X_n = A\cos(n+B)$, gdzie A, B to niezależne zmienne losowe o jednostajnym rozkładzie w przedziale $[-\pi, \pi]$.

Zadanie 8

Wykazać, że ciąg zmiennych losowych

$$Y_n = \frac{\sum_{i=1}^n (X_i - m)}{n}$$

jest zbieżny stochastycznie do zera.

Zakładamy, że zmienne losowe są niezależne o takim samym rozkładzie i skończonych momentach rzędu 2.

(Wsk. Wykazać zbieżność średniokwadratową)

Zadanie 9

Sprawdź, że punktowa granica ciągu dystrybuant

$$F_n(x) = \begin{cases} 0 & \text{gdy } x \leq -n \\ \frac{x+n}{2n} & \text{gdy } -n < x \leq n \\ 1 & \text{gdy } x > n \end{cases}$$

jest funkcją która nie jest dystrybuantą.

L.Kowalski 9.10.2009