

Zajęcia 1: Przestrzeń probabilistyczna

2024-10-04

Przykład. Weryfikacja równości dwóch wielomianów. $F(x) = \prod_{i=1}^d (x - a_i)$ i $G(x)$ w innej postaci. Wymnożenie i sprawdzenie współczynników działa w $O(d^2)$. Można lepiej: wybieramy losowo ze zbioru $\{1, \dots, 100 \cdot d\}$, liczymy wielomiany w $O(d)$ i porównujemy. Wielomian $F(x) - G(x)$ ma max d pierwiastków, czyli jak wielomiany są różne, to jest max d miejsc, gdzie algorytm się pomyli. Czyli mamy $\frac{1}{100}$ szansy na pomyłkę.

Definicja 1. (Ω, \mathcal{F}, P) jest przestrzenią probabilistyczną, jeśli:

- Ω – zbiór zdarzeń elementarnych
- $\mathcal{F} \subseteq 2^\Omega$ – zbiór zdarzeń mierzalnych. \mathcal{F} jest σ -algebrą nad Ω (w przypadku dyskretnym możemy zawsze wziąć po prostu 2^Ω)
- $P : \mathcal{F} \rightarrow \mathbb{R}$ jest funkcją prawdopodobieństwa. $\forall E \in \mathcal{F} 0 \leq P(E) \leq 1$, $P(\Omega) = 1$ oraz dla dowolnej przeliczalnej rodziny rozłącznych zbiorów $\{E_i\}_{i \in \mathbb{N}}$ zachodzi $P(\bigcup_{i \in \mathbb{N}} E_i) = \sum_{n=1}^{\infty} P(E_i)$

Uwaga. Dla skończonego Ω i $A \subseteq \Omega$ mamy $P(A) = \sum_{a \in A} P(\{a\})$.

Definicja 2. σ -algebra to zbiór \mathcal{F} taki, że

- $\emptyset \in \mathcal{F}$
- $X \in \mathcal{F} \implies \Omega \setminus X \in \mathcal{F}$
- $\forall \{X_i : X_i \in \mathcal{F}\}_{i \in \mathbb{N}} \bigcup_{i \in \mathbb{N}} X_i, \bigcap_{i \in \mathbb{N}} X_i \in \mathcal{F}$

Przykład. Przy sprawdzaniu równości wielomianów mamy $\Omega = \{1, \dots, 100d\}$, niech E – algorytm się pomylił, mamy $P(\{w\}) = \frac{1}{100d}$, więc $P(E) = \frac{\|E\|}{100d} \leq \frac{d}{100d} = \frac{1}{100}$. Wykonując k razy mamy prawdopodobieństwo $(\frac{1}{100})^k$.

Definicja 3. Zdarzenia E, F są niezależne, jeśli $P(E \cap F) = P(E) \cdot P(F)$. Ogólniej E_1, \dots, E_k są niezależne, jeśli $\forall I \subseteq [k] \neq \emptyset P(\bigcap_{i \in I} E_i) = \prod_{i \in I} P(E_i)$.

Definicja 4. Prawdopodobieństwo warunkowe zdarzenia E pod warunkiem $F : P(F) > 0$ to $P(E|F) = \frac{P(E \cap F)}{P(F)}$. Definiujemy tak, bo to oznacza ograniczenie się z całego Ω do samego F . Dla niezależnych E, F jest $P(E|F) = P(E)$.

Twierdzenie 1 (Prawdopodobieństwo całkowite). Jeśli E_1, \dots, E_n jest podziałem Ω oraz $\forall i \in [n] P(E_i) > 0$, to dla dowolnego $B \in \mathcal{F}$ jest

$$P(B) = \sum_{i=1}^n P(B \cap E_i) = \sum_{i=1}^n P(B|E_i)P(E_i).$$

Przykład. Weryfikacja mnożenia macierzy. Trzy macierze $A, B, C \in \{0, 1\}^{n \times n}$. Arytmetyka modulo 2. Sprawdzamy, czy $A \cdot B = C$. Wymnożenie daje $O(n^3)$. Bierzemy $R = (r_1, \dots, r_n) \in \{0, 1\}^n$ losowo. Obliczamy $A(BR)$ i CR , sprawdzamy czy równe. Działa w $O(n^2)$. Jeśli $AB \neq C$ i R jest losowe, to prawdopodobieństwo $P(ABR = CR) \leq \frac{1}{2}$.

Dowód. Mamy $D = AB - C \neq 0$, ale $ABR = CR \implies DR = 0 \implies \sum_{j=1}^n d_{1j}r_j = 0$. D ma niezerowy element, bso $d_{11} \neq 0$. Zatem $r_1 = \frac{-\sum_{j=2}^n d_{1j}r_j}{d_{11}}$ i jest wyznaczone przez pozostałe elementy wektora. Stosujemy *principle of deferred decision* – najpierw losujemy $\{r_2, \dots, r_n\}$, potem dopiero

r_1 . Teraz mamy:

$$\begin{aligned}
 P(ABR = CR) &= \sum_{(x_2, \dots, x_n) \in \{0,1\}^{n-1}} P((ABR = CR) \cap (r_2, \dots, r_n) = (x_2, \dots, x_n)) \\
 &\leq \sum_{(x_2, \dots, x_n) \in \{0,1\}^{n-1}} P\left(\left(r_1 = \frac{-\sum_{j=2}^n d_{1j} r_j}{d_{11}}\right) \cap (r_2, \dots, r_n) = (x_2, \dots, x_n)\right) \\
 &= \sum_{(x_2, \dots, x_n) \in \{0,1\}^{n-1}} P\left(r_1 = \frac{-\sum_{j=2}^n d_{1j} r_j}{d_{11}} \mid (r_2, \dots, r_n) = (x_2, \dots, x_n)\right) \cdot \\
 &\quad P((r_2, \dots, r_n) = (x_2, \dots, x_n)) = \sum_{(x_2, \dots, x_n) \in \{0,1\}^{n-1}} \frac{1}{2} \cdot \frac{1}{2^{n-1}} = \frac{1}{2}.
 \end{aligned}$$

Zajęcia 2: Zestaw 1

2024-10-08

Twierdzenie 2. Dla przestrzeni probabilistycznej (Ω, \mathcal{F}, P) zachodzi:

1. $P(\emptyset) = 0$
2. $P(\Omega \setminus A) = 1 - P(A)$
3. jeśli $A \subseteq B$, to $P(B \setminus A) = P(B) - P(A)$
4. jeśli $A \subseteq B$, to $P(A) \leq P(B)$
5. $P(A) \leq 1$
6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Dowód. 1. $P(\emptyset) = P(\bigcup_{n=1}^{\infty} \emptyset) = \sum_{n=1}^{\infty} P(\emptyset) \implies P(\emptyset) = 0$

2. $1 = P(\Omega) = P(\Omega \setminus A \cup A) = P(\Omega \setminus A) + P(A)$

3. jak wyżej

4. z powyższego i nieujemności $P(B \setminus A)$

5. z punktu drugiego

6. $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$.

□

Twierdzenie 3. Jeśli $\{A_n\}_{n \in \mathbb{N}}$ jest wstępującą rodziną zdarzeń, czyli $A_0 \subseteq A_1 \subseteq \dots$, to

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Natomiast dla zstępującej rodziny zdarzeń $\{A_n\}_{n \in \mathbb{N}}$, czyli $A_0 \supseteq A_1 \supseteq \dots$, zachodzi

$$P\left(\bigcap_{n=0}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n).$$

Dowód. Niech $A_{-1} = \emptyset$. Dostajemy

$$\begin{aligned}
 P\left(\bigcup_{i=0}^{\infty} A_i\right) &= P\left(\bigcup_{i=0}^{\infty} (A_i \setminus A_{i-1})\right) = \sum_{i=0}^{\infty} P(A_i \setminus A_{i-1}) = \lim_{n \rightarrow \infty} \left[\sum_{i=0}^n P(A_i \setminus A_{i-1}) \right] = \\
 &\quad \lim_{n \rightarrow \infty} P(A_n).
 \end{aligned}$$

Druga równość wynika z praw de Morgana. Mamy $\overline{A_0} \subseteq \overline{A_1} \subseteq \dots$, a więc

$$1 - \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} P(\overline{A_n}) = P\left(\bigcup_{n=0}^{\infty} \overline{A_n}\right) = 1 - P\left(\bigcap_{n=0}^{\infty} A_n\right).$$

□

Twierdzenie 4 (union bound). Dla dowolnego ciągu zdarzeń A_0, A_1, \dots , zachodzi

$$P\left(\bigcup_{n=0}^{\infty} A_n\right) \leq \sum_{n=0}^{\infty} P(A_n).$$

Dowód. Definiujemy rodzinę $\{B_n\}_{n \in \mathbb{N}}$ jako

$$B_i = \begin{cases} A_i & i = 1 \\ A_i \setminus (A_1 \cup \dots \cup A_{i-1}) & i \neq 1 \end{cases}$$

Zdarzenia z $\{B_n\}_{n \in \mathbb{N}}$ są rozłączne oraz $B_i \subseteq A_i$, więc jest

$$P\left(\bigcup_{i=0}^{\infty} A_i\right) = P\left(\bigcup_{i=0}^{\infty} B_i\right) = \sum_{i=0}^{\infty} P(B_i) \leq \sum_{i=0}^{\infty} P(A_i).$$

□

Twierdzenie 5. Niech A_0, A_1, \dots będzie nieskończonym ciągiem zdarzeń i niech $P(A_n) = 1$ dla każdego n . Zachodzi $P(\bigcap_{n=0}^{\infty} A_n) = 1$.

Dowód. Mamy

$$P\left(\bigcap_{n=0}^{\infty} A_n\right) = P\left(\overline{\bigcup_{n=0}^{\infty} \overline{A_n}}\right) = 1 - P\left(\bigcup_{n=0}^{\infty} \overline{A_n}\right) \geq 1 - \sum_{n=0}^{\infty} P(\overline{A_n}) = 1.$$

□

Twierdzenie 6. Mamy zadany ciąg $q \in \{0, 1\}^{19}$. Prawdopodobieństwo, że losując nieskończenie wiele zer i jedynek otrzymamy ciąg, którego spójnym podciągiem jest q , wynosi 1.

Dowód. Dzielimy ciąg na fragmenty po 19. X_i – zdarzenie, że q to i -ty taki fragment. Mamy $P(X_i) = \frac{1}{2^{19}}$. Zatem szansa, że q nie wystąpi na pierwszych k segmentach wynosi $P(\overline{X}) \leq P\left(\bigcap_{k=1}^{\infty} \bigcup_{i=1}^k \overline{X_i}\right) = \lim_{k \rightarrow \infty} P\left(\bigcup_{i=1}^k \overline{X_i}\right) = \lim_{k \rightarrow \infty} \left(1 - \frac{1}{2^{19}}\right)^k = 0$.

□

Twierdzenie 7. Istnieje przykład $n+1$ zdarzeń takich, że każde n z nich są niezależne, ale wszystkie $n+1$ nie są.

Dowód. Rzucamy monetą n razy. Dla $i \in [n]$ niech A_i – za i -tym razem wypadł orzeł. A_{n+1} – łączna liczba orłów jest parzysta. Prawdopodobieństwo każdego zdarzenia osobno to $\frac{1}{2}$, zdarzenie A_{n+1} jest niezależne od pozostałych, o ile są jakieś nieokreślone rzuty (połowa ciągów nieokreślonych rzutów będzie zawierać parzystą liczbę orłów), ale A_{n+1} jest zdeterminowane przez wszystkie poprzednie.

□

Twierdzenie 8. Oczekiwana liczba inwersji w losowej permutacji wynosi $\frac{n(n-1)}{4}$.

Dowód. Niech $X_{ij} = \begin{cases} 1, & \pi(i) > \pi(j) \\ 0, & \pi(i) < \pi(j) \end{cases}$, X – liczba inwersji. $E[X] = \sum_{\{i,j\} \in \binom{[n]}{2}} E[X_{ij}] = \binom{n}{2} \cdot \frac{1}{2}$.

□

Twierdzenie 9. Jeśli zdarzenia A_1, \dots, A_n są niezależne, to $\overline{A_1}, \dots, \overline{A_n}$ też są.

Dowód. Dla dowolnego zbioru $\emptyset \neq N \in [n]$ mamy

$$1 - P\left(\bigcap_{i \in N} \overline{A_i}\right) = P\left(\bigcup_{i \in N} A_i\right) = \sum_{\emptyset \neq I \subseteq N} (-1)^{|I|+1} \prod_{i \in I} P(A_i)$$

oraz

$$1 - \prod_{i \in N} (1 - P(A_i)) = \sum_{\emptyset \neq I \subseteq N} (-1)^{|I|+1} \prod_{i \in I} P(A_i),$$

co daje

$$P\left(\bigcap_{i \in N} \overline{A_i}\right) = \prod_{i \in N} (1 - P(A_i)) = \prod_{i \in N} P(\overline{A_i}).$$

□

Zajęcia 3: Zmienna losowa

2024-10-04

Definicja 5. X jest zmienną losową na przestrzeni (Ω, \mathcal{F}, P) , jeśli $X : \Omega \rightarrow \mathbb{R}$ i $\forall F \in \mathcal{F} \ X(F)$ jest zbiorem borelowskim nad \mathbb{R} (w przypadku dyskretnym ten warunek nie jest konieczny). Dyskretna zmienna losowa przyjmuje co najwyżej przeliczalnie wiele wartości.

Uwaga. Często będziemy chcieli wyznaczyć prawdopodobieństwo $P(X = a)$ osiągnięcia jakiejś wartości przez zmienną losową. W przypadku dyskretnym wynosi ona

$$P(X = a) = \sum_{s \in \Omega: X(s)=a} P(\{s\}).$$

Definicja 6. Zmienne losowe X, Y są niezależne, jeśli $\forall x, y \in \mathbb{R} \ P((X = x) \cap (Y = y)) = P(X = x) \cdot P(Y = y)$. Ogólniej, zmienne losowe X_1, \dots, X_n są niezależne, jeśli

$$\forall I \subseteq [n] \ \forall \{x_i\}_{i \in [n]} \ P\left(\bigcap_{i \in I} (X_i = x_i)\right) = \prod_{i \in I} P(X_i = x_i).$$

Definicja 7. Wartość oczekiwana zmiennej losowej X , oznaczana $E[X]$, to

$$E(X) = \sum_i i \cdot P(X = i),$$

gdzie suma przebiega po obrazie X . Wartość oczekiwana może być nieokreślona, jeśli jest nieskończenie wiele wartości w obrazie i szereg określający wartość oczekiwaną nie jest zbieżny.

Przykład. Jeśli zmienna losowa X przyjmuje wartość 2^i z prawdopodobieństwem $\frac{1}{2^i}$ dla $i \in \mathbb{N}_1$, to $E[X] = \sum_{i=1}^{\infty} 2^i \cdot \frac{1}{2^i} = \infty$. Zatem wartość oczekiwana nie jest określona.

Twierdzenie 10 (Liniowość wartości oczekiwanej). Dla dowolnych zmiennych losowych X_1, \dots, X_n z ograniczonymi wartościami oczekiwanymi zachodzi

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i].$$

Dowód. Dowodzimy dla $n = 2$, reszta z indukcji. Mamy

$$\begin{aligned} E[X + Y] &= \sum_k k \cdot P(X + Y = k) = \sum_{i,j} (i + j) P((X = i) \cap (Y = j)) \\ &= \sum_i i \sum_j P((X = i) \cap (Y = j)) + \sum_j j \sum_i P((Y = j) \cap (X = i)) \\ &= \sum_i i P(X = i) + \sum_j j P(Y = j) = E[X] + E[Y]. \end{aligned}$$

□

Twierdzenie 11 (Mnożenie wartości oczekiwanej przez stałą). Dla dowolnej liczby rzeczywistej c i zmiennej losowej X zachodzi $E[cX] = cE[X]$.

Dowód. Dla $c = 0$ oczywiste. Dla $c \neq 0$ przekształcamy:

$$E[cX] = \sum_j j P(cX = j) = c \sum_j \frac{j}{c} P\left(X = \frac{j}{c}\right) = c \sum_k k P(X = k) = cE[X].$$

□

Przykład. Urna z 4 kulami różnych kolorów. 4 razy wyciągamy kulę i zwracamy. Jaka jest oczekiwana liczba zaobserwowanych kolorów?

Dowód. X – liczba zaobserwowanych kolorów. X_i – zaobserwowanie koloru i . Mamy $P(X_i = 1) = 1 - \left(\frac{3}{4}\right)^4$. Zatem $E[X] = E[X_1 + X_2 + X_3 + X_4] = 4 \left(1 - \left(\frac{3}{4}\right)^4\right)$.

Przykład. n mrówek stoi na kiju długości $1m$. Są zwrócone w różne strony. Zaczynają iść przed siebie, w momencie zderzenia odwracają się i idą dalej. Po zetknięciu się z końcem kija spadają z niego. Jaka jest oczekiwana liczba zderzeń?

Dowód. Zauważmy, że można rozważać sytuację, w której mrówki się nie odbijają, lecz przenikają – rezultat przeniknięcia i zderzenia jest taki sam. W takiej sytuacji dwie mrówki zderzą się tylko, gdy lewa jest zwrócona w prawo, a prawa w lewo. Zatem wartość oczekiwana to $\binom{n}{2} \cdot \frac{1}{4}$.

Zajęcia 4: Rozkład zmiennej losowej

2024-10-11

Definicja 8. Rozkład dyskretnej zmiennej losowej to zbiór (przeliczalny) $\{x \in \mathbb{R} : P(X = x) > 0\}$. Po nim są sumy w definicji wartości oczekiwanej. $\sum_x P(X = x) = 1$.

Propozycja 1. Zachodzi $E[f(X)] = \sum_y y \cdot P(f(X) = y) = \sum_x f(x) \cdot P(X = x)$, podobnie $E[g(X, Y)] = \sum_{x,y} g(x, y) \cdot P(X = x, Y = y)$.

Przykład. Niech $P(X = n) = \frac{1}{99}$ dla każdego $n \in [99]$. $E[X]^2 = \left(\sum_n \frac{n}{99}\right)^2 = 50^2 = 2500$ oraz $E[X^2] = \sum_n n^2 \cdot \frac{1}{99} = \frac{1}{99} \cdot \frac{99 \cdot 100 \cdot 199}{6} \approx 3300$. Mamy przykład $E[X^2] > E[X]^2$.

Propozycja 2. Zachodzi $E[X^2] \geq E[X]^2$.

Dowód. Niech $Y = (X - E[X])^2$. Mamy

$$\begin{aligned} 0 \leq E[Y] &= E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2. \end{aligned}$$

□

Definicja 9. Funkcja $f : \mathbb{R} \rightarrow \mathbb{R}$ jest wypukła, jeśli

$$\forall x_1, x_2, 0 \leq \lambda \leq 1 \quad f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

Dla funkcji różniczkowalnej jest to równoważne dodatniości drugiej pochodnej.

Twierdzenie 12 (Nierówność Jensena). Jeśli f jest wypukła, to

$$E[f(X)] \geq f(E[X]).$$

Dowód. Dowodzimy zakładając, że f ma rozkład Taylora. Niech $\mu = E[X]$. Z rozwinięcia Taylora istnieje takie c , że

$$f(x) = f(\mu) + (x - \mu)f'(\mu) + \frac{(x - \mu)^2}{2}f''(c).$$

Z nieujemności drugiej pochodnej jest $f(x) \geq f(\mu) + (x - \mu)f'(\mu)$, a więc

$$\begin{aligned} E[f(X)] &\geq E[f(\mu) + (X - \mu)f'(\mu)] = E[f(\mu)] + f'(\mu)E[X - \mu] \\ &= E[f(E[X])] + f'(\mu)(E[X] - E[X]) = f(E[X]), \end{aligned}$$

bo $E[X - E[X]] = 0$. □

Definicja 10. Indykator zdarzenia A to zmienna losowa $Y = \begin{cases} 1 & A \text{ zaszło} \\ 0 & \text{wpp} \end{cases}$. Mamy

$$E[Y] = 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) = P(Y = 1) = P(A).$$

Definicja 11. Zmienna X ma rozkład dwumianowy z parametrami (n, p) dla $n \in \mathbb{N}_1$ i $p \in (0, 1)$, jeśli dla $i \in \{0, \dots, n\}$ jest $P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}$. Mamy

$$\sum_{i \in \{0, \dots, n\}} P(X = i) = \sum_{i \in \{0, \dots, n\}} \binom{n}{i} p^i (1 - p)^{n-i} = (p + (1 - p))^n = 1,$$

więc jest to rozkład.

Zmienna o rozkładzie dwumianowym modeluje sytuację, w której wykonujemy ciąg n niezależnych prób, z których każda ma prawdopodobieństwo sukcesu p i zliczamy liczbę sukcesów.

Propozycja 3. Zmienna losowa X o rozkładzie dwumianowym z parametrami (n, p) ma wartość oczekiwaną $E[X] = np$.

Dowód. Niech X_i – indyktor sukcesu w i -tym kroku. Mamy $E[X] = \sum_i E[X_i] = np$.

Można też wprost:

$$\begin{aligned} E[X] &= \sum_{i=0}^n i \binom{n}{i} p^i (1 - p)^{n-i} = np \sum_{i=1}^n \frac{(n-1)!}{(i-1)! \cdot ((n-1) - (i-1))!} p^{i-1} (1 - p)^{n-1-(i-1)} \\ &= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k! (n-1-k)!} p^k (1 - p)^{n-1-k} = np (p + (1 - p))^{n-1} = np. \end{aligned}$$

□

Definicja 12. Warunkowa wartość oczekiwana (dla $P(Z = z) > 0$) to

$$E[Y | Z = z] = \sum_y y P(Y = y | Z = z).$$

Przykład. Rzucamy dwiema kostkami. Mamy X_1 – wynik na pierwszej kostce, tak samo X_2 . $X = X_1 + X_2$. Mamy $E[X | X_1 = 2] = 5.5$, bo wynik na jednej jest znany a na drugiej ma wartość oczekiwaną 3.5.

$$E[X | X_1 = 2] = \sum_x x \cdot P(X = x | X_1 = 2) = \frac{1}{6} \sum_{x=3}^8 x = 5.5$$

Podobnie $E[X_1 | X = 5] = 2.5$, bo na każdej tyle samo a razem 5.

$$E[X_1 | X = 5] = \sum_{x=1}^4 x P(X_1 = x | X = 5) = \sum x \frac{P(X_1=x \cap X=5)}{P(X=5)} = \sum_1^4 x \frac{\frac{1}{36}}{\frac{4}{36}} = 2.5.$$

Propozycja 4. Zmienne losowe X, Y . Mamy $E[X] = \sum_y E[X | Y = y] \cdot P(Y = y)$.

Dowód.

$$\begin{aligned} \sum_y E[X | Y = y] P(Y = y) &= \sum_y \sum_x x P(X = x | Y = y) P(Y = y) \\ &= \sum_x x \sum_y P(X = x | Y = y) P(Y = y) = \sum_x x \sum_y P(X = x \cap Y = y) \\ &= \sum_x x P(X = x) = E[X]. \end{aligned}$$

□

Propozycja 5. X_1, X_2 zmienne losowe o skończonej wartości oczekiwanej. Y zmienna losowa.

$$E[X_1 + X_2 | Y = y] = E[X_1 | Y = y] + E[X_2 | Y = y].$$

Dowód. Dowód jak liniowość.

□

Definicja 13. $E[Y | Z]$ to funkcja zmiennej losowej Z , a zatem zmienna losowa. Definiujemy dla $\omega \in \Omega$:

$$E[Y | Z](\omega) = E[Y | Z = Z(\omega)].$$

Przykład. Dla kostek $E[X | X_1] = X_1 + 3.5$, bo druga kostka ma zwykłą wartość, a pierwsza zależy od pierwszej (xd).

$$E[X | X_1](\omega) = E[X | X_1 = z] = \sum_x x P(X = x | X_1 = z) = \sum_{x=z+1}^{z+6} x \cdot \frac{1}{6} = z + 3.5$$

Propozycja 6. $E[Y] = E[E[Y | Z]]$.

Dowód. Korzystamy z $E[f(Z)] = \sum_z f(z) P(Z = z)$. Mamy

$$E[E[Y | Z]] = \sum_z E[Y | Z = z] P(Z = z) = E[Y].$$

□

Przykład (Prosty proces gałązkowy). Mamy proces S . On forkuje niezależnie kilka procesów S . Procesy są niezależne, liczba dzieci każdego to zmienna oczekiwana o rozkładzie dwumianowym z parametrami (n, p) . Jaka jest oczekiwana liczba wszystkich odpalanych procesów?

Dowód. Niech Y_i – liczba procesów w i -tym pokoleniu. Mamy $Y_0 = 1$. $E[Y_1] = np$. Zakładamy, że $Y_{i-1} = y_{i-1}$. Niech $Z_1, \dots, Z_{y_{i-1}}$ to liczby dzieci procesów w $(i-1)$ -szym pokoleniu.

$$\begin{aligned}
E[Y_i | Y_{i-1} = y_{i-1}] &= E\left[\sum_{k=1}^{y_{i-1}} Z_k | Y_{i-1} = y_{i-1}\right] = \sum_{j \geq 0} j \cdot P\left(\sum Z_k = j | Y_{i-1} = y_{i-1}\right) \\
&= \sum_{j \geq 0} j \cdot P\left(\sum Z_k = j\right) = E\left[\sum Z_k\right] = \sum E[Z_k] = y_{i-1}np.
\end{aligned}$$

Trzecie przejście działa, bo każda zmienna Z_i jest niezależna od przeszłości, więc ich liczba jest zależna od Y_{i-1} , ale sama suma jest niezależna. Zatem $E[Y_i | Y_{i-1}] = Y_{i-1}np$.

Teraz $E[Y_i] = E[E[Y_i | Y_{i-1}]] = E[Y_{i-1}np] = npE[Y_{i-1}]$. Indukcja daje $E[Y_i] = (np)^i$.

Ostatecznie $E[\sum Y_i] = \sum E[Y_i] = \sum (np)^i = \frac{1}{1-np}$. Problem jest z pierwszym przejściem, bo ta suma jest nieskończona. O ile $np < 1$, to jest dobrze, bo przeliczalna liniowość działa, o ile wychodzący szereg jest zbieżny. Inaczej wartość oczekiwana jest nieograniczona.

Definicja 14. Zmienna X ma rozkład geometryczny z parametrem p , jeśli $P(X = n) = (1-p)^{n-1}p$ dla $n \in \mathbb{N}_1$. Mamy $\sum_{n \in \mathbb{N}_1} P(X = n) = 1$ (szereg geometryczny), więc jest to rozkład.

Zmienna o rozkładzie geometrycznym modeluje sytuację, w której mamy prawdopodobieństwo sukcesu p i sprawdzamy, czy pierwszy sukces wystąpi po n -tej próbie.

Propozycja 7. Rozkład geometryczny jest bez pamięci (*memoryless*), czyli fakt wystąpienia nieudanych prób nie zmienia rozkładu.

Dowód. $\forall n \in \mathbb{N}_1, k \in \mathbb{N} \quad P(X = n+k | X > k) = P(X = n)$ Jest tak bo

$$P(X = n+k | X > k) = \frac{P(X = n+k \cap X > k)}{P(X > k)} = \frac{(1-p)^{n+k-1}p}{\sum_{i=k+1}^{\infty} (1-p)^{i-1}p} = (1-p)^{n-1}p.$$

□

Propozycja 8. Jeśli zmienna losowa X przyjmuje wartości nieujemne całkowite, to

$$E[X] = \sum_{i=1}^{\infty} P(X \geq i).$$

Dowód.

$$\sum_{i=1}^{\infty} P(X \geq i) = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} P(X = j) = \sum_{j=1}^{\infty} \sum_{i=1}^j P(X = j) = \sum_{j=1}^{\infty} jP(X = j) = E[X].$$

□

Zajęcia 5: Zestaw 2

2024-10-15

Twierdzenie 13. Każdy graf G zawiera podgraf dwudzielny zawierający co najmniej połowę wierzchołków G .

Dowód. Losujemy podzbiór wierzchołków, które leżą w jednym zbiorze niezależnym. Pozostałe idą do drugiego. Zachowujemy tylko te krawędzie, które idą między zbiorami niezależnymi. W ten sposób powstaje graf dwudzielny H . Każda krawędź będzie w nim istniała z prawdopodobieństwem $\frac{1}{2}$, bo wierzchołki które łączy są w różnych lub tych samych zbiorach. Zatem wartość oczekiwana liczby krawędzi to $\frac{k}{2}$, gdzie k to liczba krawędzi w G . Zatem istnieje taki podgraf dwudzielny, który zawiera co najmniej $\frac{k}{2}$ krawędzi. □

Twierdzenie 14. Oczekiwana liczba cykli w losowej permutacji długości n wynosi $H_n \approx \ln n$.

Dowód. Niech X – liczba cykli w permutacji. Rozbijamy na X_i – liczba cykli długości i . Wybierzmy dowolny element $[n]$. Ustawiamy go na początku permutacji, pozostałe $n - 1$ elementów dowolnie. Pierwsze i tworzy cykl długości i , a pozostałe są zwykłą permutacją. W ten sposób zliczamy cykle – mamy $n \cdot (n - 1)! = n!$, ale każdy cykl zliczamy i razy (bo nie ma znaczenia, który element postawimy jako pierwszy). Zatem $E[X_i] = \frac{1}{i} \cdot \frac{n!}{n!} = \frac{1}{i}$, zatem $E[X] = \sum_{i=1}^n \frac{1}{i} = H_n \approx \ln n$. \square

Twierdzenie 15. W talii $2n$ kart mamy n czerwonych i n czarnych kart. Talia jest potasowana i kolejno wszystkie karty są wykładane na stół. Dla każdej wyciągniętej czerwonej karty, jeśli po jej wyciągnięciu na stole jest więcej czerwonych niż czarnych kart to zdobywamy jeden punkt. Oczekiwana liczba zdobytych punktów to $\frac{n}{2}$.

Dowód. Będziemy przyznawać punkty za czarne karty analogicznie jak za czerwone. Pokażemy, że razem ich oczekiwana liczba to n , a więc z symetrii odpowiedź to $\frac{n}{2}$.

Niech X_i oznacza liczbę czerwonych, a Y_i czarnych kart po i -tym kroku. Niech $Z_i = \|X_i - Y_i\|$. Zmienna $R_i = Z_{i+1} - Z_i$ przyjmuje tylko wartości $\{-1, 1\}$. Dla 1 dostajemy punkt, dla -1 nie. Mamy $\sum_{i=0}^{2n-1} R_i = Z_{2n} - Z_0 = 0$, więc dokładnie połowa z przyjmowanych przez R_i wartości to 1. \square

Twierdzenie 16. Jeśli $\text{Var}(X) = 0$, to istnieje takie a , że $P(X = a) = 1$.

Dowód. Mamy $\text{Var}(X) = E[(X - E[X])^2] = 0$, a więc $X - E[X]$ przyjmuje zawsze wartość 0, co dowodzi, że $P(X = E[X]) = 1$. \square

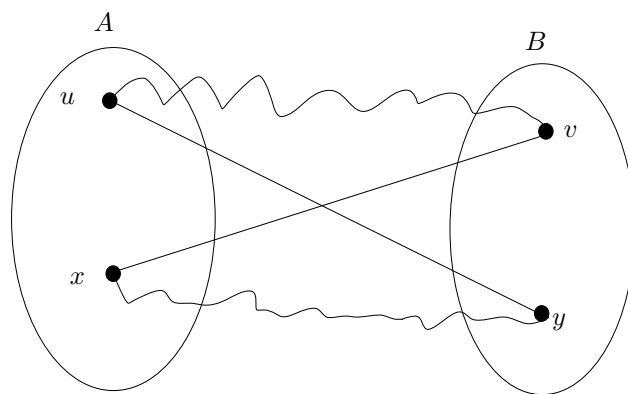
Twierdzenie 17. Każdy graf dwudzielny z m krawędziami zawiera podgraf na co najmniej $\frac{3}{4}m^{\frac{2}{3}}$ krawędziach, który nie zawiera cyklu na czterech wierzchołkach jako podgrafu.

Dowód. Dla dowolnego grafu G losujemy podgraf H – mamy wszystkie wierzchołki i losujemy każdą krawędź niezależnie z prawdopodobieństwem p na należenie do H . W tak otrzymanym grafie najprawdopodobniej będą cykle C_4 . Rozważamy każdy z nich i z każdego usuwamy po jednej krawędzi, w ten sposób otrzymując podgraf o żądanej własności. Niech X oznacza liczbę C_4 w wylosowanym H . Na początek wyznaczmy $E[X]$. Mamy

$$E[X] = \sum_{\substack{uv \in E(G) \\ xy \in E(G) \\ u, x \in A \\ v, y \in B \\ uv, xy \in E(G) \\ u < x \\ v < y}} P(uv, xy, uv, xv \in H) \leq \frac{1}{2} \binom{m}{2} p^4 \leq \frac{1}{2} \cdot \frac{m^2}{2} p^4,$$

gdzie warunki sumowania oznaczają, że rozważane krawędzie tworzą cykl C_4 jak na Rysunku 1 (nierówność na wierzchołkach to pewien porządek liniowy który służy temu, by nie zliczać par dwa razy). Nierówność wynika z tego, że każde dwie pary liczymy raz.

Daje nam to ostatecznie $E[E(H) - X] \geq mp - \frac{m^2}{4} p^4$, co po przyjęciu $p = m^{-\frac{1}{3}}$ (co oczywiście należy do $(0, 1)$) przyjmuje wartość $\frac{3}{4}m^{\frac{2}{3}}$. Zatem oczekiwana liczba krawędzi w tak wylosowanym grafie jest większa od $\frac{3}{4}m^{\frac{2}{3}}$, a więc istnieje graf, który przyjmuje co najmniej taką wartość. \square

Rysunek 1: C_4 w grafie H

Zajęcia 6: Miara i całka

2024-10-18

Przykład. Losujemy jednostajnie $x \in \Omega$. Mając pewien zbiór A jest intuicyjnie

$$P(x \in A) = \frac{\text{wielkość}(A)}{\text{wielkość}(\Omega)}.$$

W przypadku $\|\Omega\| < \infty$ mamy $P(x \in A) = \frac{\|A\|}{\|\Omega\|}$. W ogólności chcemy znaleźć jakiś formalizm, który oddaje pojęcie wielkości.

Twierdzenie 18 (Paradoks Banacha-Tarskiego). Mając kulę jednostkową $K \subseteq \mathbb{R}^3$ można ją podzielić na 5 rozłącznych części i za pomocą obrotów i przesunięć utworzyć dwie kopie K .

Wniosek. Nie dla wszystkich zbiorów da się zdefiniować pojęcie objętości (wielkości). Gdyby zbiory tworzące podział w paradoksie Banacha-Tarskiego były mierzalne, to kula musiałaby mieć objętość 0.

Definicja 15. Niech Ω będzie zbiorem (dowolnym). Niech $\Sigma \subseteq 2^\Omega$. Mamy warunki:

1. $\emptyset, \Omega \in \Sigma$
2. $E \in \Sigma \implies \Omega \setminus E \in \Sigma$
3. $E_1, E_2, \dots, E_n \in \Sigma \implies \bigcup_{i=1}^n E_i \in \Sigma$ (domknięcie na skończone sumy)
4. $E_1, E_2, E_3, \dots \in \Sigma \implies \bigcup_{i=1}^\infty E_i \in \Sigma$ (domknięcie na przeliczalne sumy)

Jeśli zachodzą pierwsze 3 warunki, to Σ jest algebrą. Z czwartym warunkiem jest σ -algebrą.

Przykład. Zbiory $\Sigma = 2^\Omega$ i $\{\emptyset, \Omega\}$ są σ -algebrami. Natomiast $\{\bigcup_{i=1}^n [a_i, b_i) \subseteq \mathbb{R} : a_i < b_i, n \in \mathbb{N}\}$ nie jest – mamy pierwsze dwa warunki, trzeci też (suma przedziałów jest takim samym przedziałem), ale przeliczalna suma może dać większy zbiór (np. całe \mathbb{R}).

Propozycja 9. Niech I będzie zbiorem indeksów. Jeśli Σ_α jest σ -algebrą nad pewnym Ω dla każdego $\alpha \in I$, to $\bigcap_{\alpha \in I} \Sigma_\alpha$ też jest σ -algebrą.

Dowód. Prosty dowód, jeśli element należy do każdego elementu przecięcia i każdy z nich jest σ -algebrą, to odpowiedni zbiór należy też do przecięcia. \square

Definicja 16. σ -algebra generowana przez zbiór $\mathcal{F} \subseteq 2^\Omega$ to

$$\sigma(\mathcal{F}) = \bigcap_{\substack{\Sigma \text{ jest } \sigma\text{-algebrą,} \\ \mathcal{F} \subseteq \Sigma}} \Sigma.$$

Przykład. Mamy $\sigma(\{[a, b] : a < b\}) = \mathcal{B}(\mathbb{R})$, gdzie $\mathcal{B}(\mathbb{R})$ to zbiory borelowskie w \mathbb{R} . To są "porządne zbiory" w \mathbb{R} .

Notacja. Rodzinę zbiorów otwartych nad \mathbb{R} oznaczamy przez A_o .

Definicja 17. Zbiory borelowskie nad liczbami rzeczywistymi $\mathcal{B}(\mathbb{R})$ definiujemy jako σ -algebrę generowaną przez zbiory otwarte, czyli $\mathcal{B}(\mathbb{R}) = \sigma(A_o)$.

Definicja 18. Dla Ω i $\Sigma \subseteq 2^\Omega$ będącego σ -algebrą parę (Ω, Σ) nazywamy przestrzenią mierzalną.

Definicja 19. Funkcja $\mu : \Sigma \rightarrow [0, +\infty]$ jest miarą, jeśli:

- $\mu(\emptyset) = 0$
- $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$, gdzie $(E_i)_{i=1}^{\infty}$ są parami rozłącznymi zbiorami z Σ (przeliczalna addytywność).

Definicja 20. Miara μ jest probabilistyczna, jeśli zachodzi $\mu(\Omega) = 1$.

Notacja. Jeśli zamiast (Ω, Σ) piszemy tylko Ω i $\Omega \subseteq \mathbb{R}$, to mamy na myśli przestrzeń mierzalną z $\Sigma = \mathcal{B}(\mathbb{R})$.

Definicja 21. Funkcja $f : (\Omega_1, \Sigma_1) \rightarrow (\Omega_2, \Sigma_2)$ jest mierzalna, jeśli $\forall E \in \Sigma_2 \ f^{-1}(E) \in \Sigma_1$.

Definicja 22. Funkcję $s : \Omega \rightarrow [0, \infty)$ nazywamy prostą (lub schodkową), jeśli istnieją takie parami rozłączne zbiory $\{A_i\}_{i=1}^n \subseteq \Sigma$ (gdzie Σ jest σ -algebrą nad Ω), że

$$s = \sum_{i=1}^n \alpha_i \chi_{A_i},$$

gdzie χ_{A_i} jest funkcją charakterystyczną danego zbioru a $\{\alpha_i\}_{i=1}^n \subseteq [0, \infty)$. Znaczy to, że funkcja s przyjmuje skończenie wiele wartości na pewnych (skończenie wielu) zbiorach.

Definicja 23 (Całka Lebesgue'a). Rozważmy przestrzeń mierzalną (Ω, Σ) . Dla $E \in \Sigma$ definiujemy całkę mierzalnej funkcji prostej s jako

$$\int_E s \, d\mu = \sum_{j=1}^n \alpha_j \cdot \mu(A_j \cap E).$$

Dla dowolnej funkcji mierzalnej $f : \Omega \rightarrow [0, \infty]$ definiujemy jej całkę jako

$$\int_E f \, d\mu = \sup \left\{ \int_E s \, d\mu : s \text{ jest prosta}, 0 \leq s \leq f \right\}.$$

Definicja 24. Funkcja $f : \Omega \rightarrow \overline{\mathbb{R}}$ jest całkowalna, jeśli $\int_{\Omega} \|f\| \, d\mu < \infty$.

Uwaga. Funkcja \sin nie jest całkowalna w sensie Lebesgue, mimo że jest w sensie Riemanna.

Definicja 25. Całkę całkowalnej funkcji mierzalnej $f : \Omega \rightarrow \overline{\mathbb{R}}$ definiujemy jako

$$\int_E f \, d\mu = \int_E f_+ \, d\mu - \int_E f_- \, d\mu,$$

gdzie $f_+ = \max\{0, f\}$ oraz $f_- = -\min\{0, f\}$.

Definicja 26. Całkę całkowalnej funkcji mierzalnej $F : \Omega \rightarrow \mathbb{C}$ definiujemy jako

$$\int_E f d\mu = \int_E u d\mu + i \int_E v d\mu,$$

gdzie $f = u + iv$. Funkcje u, v są mierzalnymi, całkowalnymi funkcjami rzeczywistymi.

Twierdzenie 19 (O aproksymacji). Jeśli $f : (\Omega, \Sigma) \rightarrow [0, \infty]$ jest mierzalna, to istnieje ciąg funkcji prostych $\{s_n\}_{n=1}^\infty$ taki, że $\forall_{n \in \mathbb{N}} s_n : (\Omega, \Sigma) \rightarrow [0, +\infty)$ oraz $s_1 \leq s_2 \leq \dots \leq f$, dla którego zachodzi

$$f(x) = \lim_{n \rightarrow \infty} s_n(x)$$

dla każdego $x \in \Omega$.

Twierdzenie 20. Jeśli $f : (\Omega, \Sigma) \rightarrow [0, \infty]$ jest mierzalna, to dla ciągu funkcji prostych $\{s_n\}_{n=1}^\infty$ takiego, że $\forall_{n \in \mathbb{N}} s_n : (\Omega, \Sigma) \rightarrow [0, +\infty)$ oraz $s_1 \leq s_2 \leq \dots \leq f$, dla którego dla każdego $x \in \Omega$ zachodzi

$$f(x) = \lim_{n \rightarrow \infty} s_n(x)$$

jest też

$$\forall_{E \in \Sigma} \int_E f d\mu = \lim_{n \rightarrow \infty} \int_E s_n d\mu.$$

Algorytm 1 (Liczenia całek). Rozbijamy funkcję na część rzeczywistą i urojoną, je rozbijamy na dodatnie i ujemne części, znajdujemy odpowiedni ciąg funkcji prostych i liczymy granicę całek funkcji prostych.

Tabela 1: Prawdopodobieństwo a całki

Prawdopodobieństwo	Miara i całka
funkcja P	miara μ
zmienna losowa X	funkcja mierzalna $X : (\Omega, \Sigma) \rightarrow \mathbb{R}$
$E[X]$	$\int_\Omega X d\mu$
$\text{Var}(X)$	$\int_\Omega (X - E[X])^2 d\mu$

Przykład. Dla $\|\Omega\| < \infty$ mamy $P(A) = \frac{\|A\|}{\|\Omega\|}$ oraz $\Sigma = 2^\Omega$.

Przykład. Mamy wagi o sumie 1: p_1, p_2, \dots, p_n . Można zdefiniować $P(A) = \sum_{i \in A} p_i$, $\Sigma = 2^{[n]}$.

Przykład (Losowanie z odcinka). Jest $P(x \in A) = \text{długość}(A) = \lambda(A)$ (miara Lebesgue'a). Co ciekawe biorąc $A = \mathbb{Q}$ dostajemy $P(x \in A) = 0$, mimo, że liczby wymierne są gęste w rzeczywistych.

Przykład. Jednostajnie rzucamy piłką na odcinek długości 1. Jeśli spadnie w odległości do $\frac{1}{3}$ od 0, to zostaje do niego przyciągnięta. Mamy $P(\{0\}) = \frac{1}{3}$, $P((0, \frac{1}{3})) = 0$, $P(A) = \lambda(A)$ o ile $A \subseteq [\frac{1}{3}, 1]$.

Przykład (Nieskończony ciąg rzutów monetą). Rozważamy $\Omega = \{(x_n)_{n=1}^\infty : x_n \in \{0, 1\}\}$ Niech $[a_1 a_2 \dots a_k] = \{(x_n)_{n=1}^\infty \in \Omega : a_i = x_i\}$ będzie zbiorem wszystkich ciągów zaczynających się w dany sposób. Przez cylinder $C = \{[a_1 a_2 \dots a_k] : a_i \in \{0, 1\}, k \in \mathbb{N}\}$ rozumiemy zbiór zawierający wszystkie takie rodziny ciągów o wspólnym początku. Rozważamy σ -algebrę $\Sigma = \sigma(C)$. Mamy $P([a_1 \dots a_k]) = \frac{1}{2^k}$.

Twierdzenie 21. Złożenie dwóch funkcji mierzalnych jest mierzalne.

Dowód. Rozważmy funkcje mierzalne $f : (\Omega, \Sigma) \rightarrow (\Omega', \Sigma')$, $g : (\Omega', \Sigma') \rightarrow (\Omega'', \Sigma'')$.

Dla $E'' \in \Sigma''$ mamy $g^{-1}(E'') = E' \in \Sigma'$ oraz $f^{-1}(E') = E \in \Sigma$. Zatem $(g \circ f)^{-1}(E'') = E \in \Sigma$, czyli $g \circ f$ jest mierzalna. \square

Twierdzenie 22. Każda funkcja ciągła $f : \mathbb{R} \rightarrow \mathbb{R}$ jest mierzalna.

Dowód. Rozważmy zbiór $S = \{A \subseteq \mathbb{R} : f^{-1}(A) \in \mathbb{B}(\mathbb{R})\}$. Mamy $\mathcal{A}_o \subseteq S$, bo funkcja jest ciągła i przeciwobraz zbioru otwartego jest otwarty (borelowski). S jest również σ -algebrą (\emptyset, \mathbb{R} są, zamknięcie na sumy wynika z tego, że przeciwobraz sumy to suma przeciwobrazów, a zamknięcie na dopełnienia wynika z tego, że $f^{-1}(\bar{A}) = \overline{f^{-1}(A)}$). Zatem S jest σ -algebrą zawierającą zbiory otwarte, czyli $\sigma(\mathcal{A}_o) \subseteq S$, ale $\sigma(\mathcal{A}_o) = \mathbb{B}(\mathbb{R})$, zatem przeciwobraz każdego zbioru borelowskiego w f jest borelowski, czyli funkcja jest mierzalna. \square

Twierdzenie 23. Dla mierzalnej funkcji $f : \Omega \rightarrow \mathbb{C}$ zachodzi

$$\int_E f \, d\mu = \int_{\Omega} \chi_E f \, d\mu.$$

Dowód. Dla funkcji prostej $f = \sum_{i=1}^n \alpha_i \cdot \chi_{A_i}$ mamy

$$\int_{\Omega} \chi_E \cdot f \, d\mu = \int_{\Omega} \chi_E \cdot \sum_{i=1}^n \alpha_i \chi_{A_i} \, d\mu = \int_{\Omega} \sum_{i=1}^n \alpha_i \chi_{A_i \cap E} \, d\mu = \sum_{i=1}^n \alpha_i \cdot \mu(A_i \cap E \cap \Omega) = \int_E f \, d\mu.$$

Dalej rozważmy funkcję $f : \mathbb{R} \rightarrow [0, +\infty)$. Istnieje ciąg funkcji prostych $(s_i)_{i=1}^{\infty}$ taki, że $s_1 \leq \dots \leq f$ oraz $\lim_{n \rightarrow \infty} s_n(x) = f(x)$ dla każdego $x \in \mathbb{R}$. Mamy

$$\int_E f \, d\mu = \lim_{n \rightarrow \infty} \int_E s_i \, d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} \chi_E \cdot s_i \, d\mu = \int_{\Omega} \chi_E \cdot f \, d\mu,$$

gdzie ostatnie przejście wynika z tego, że funkcje $\chi_E \cdot s_i$ są odpowiednim ciągiem dla funkcji $\chi_E \cdot f$.

Przejście z funkcji dodatnich na rzeczywiste: rozważamy $f_+ = \max(0, f)$ i $f_- = -\min(0, f)$, mamy $\int_E f \, d\mu = \int_E f_+ \, d\mu - \int_E f_- \, d\mu = \int_{\Omega} \chi_E f_+ \, d\mu - \int_{\Omega} \chi_E f_- \, d\mu = \int_{\Omega} \chi_E f \, d\mu$, gdzie pierwsze i ostatnie przejście wynika z definicji całki rzeczywistej.

Przejście na funkcje zespolone analogicznie z definicji. \square

Twierdzenie 24. Dla mierzalnych funkcji $f, g : \Omega \rightarrow \mathbb{C}$ zachodzi

$$\int_E f + g \, d\mu = \int_E f \, d\mu + \int_E g \, d\mu.$$

Dowód. Niech f, g będą funkcjami prostymi. $f = \sum_{i=1}^n \alpha_i \chi_{A_i}$, $g = \sum_{i=1}^m \beta_i \chi_{B_i}$.

Mamy

$$f + g = \sum_{i \in [n], j \in [m]} (\alpha_i + \beta_j) \cdot \chi_{A_i \cap B_j} + \sum_{i \in [n]} \alpha_i \cdot \chi_{A_i \setminus \bigcup_j B_j} + \sum_{j \in [m]} \beta_j \cdot \chi_{B_j \setminus \bigcup_i A_i}.$$

Teraz

$$\begin{aligned}
\int_E f + g \, d\mu &= \sum_{i \in [n], j \in [m]} (\alpha_i + \beta_j) \mu(A_i \cap B_j \cap E) + \sum_{i \in [n]} \alpha_i \mu\left((A_i \cap E) \setminus \bigcup_j B_j\right) + \\
&\quad \sum_{j \in [m]} \beta_j \mu\left((B_j \cap E) \setminus \bigcup_i A_i\right) = \\
&\quad \sum_{i \in [n]} \alpha_i \left(\left(\sum_{j \in [m]} \mu(A_i \cap B_j \cap E) \right) + \mu\left(\left(A_i \setminus \bigcup_j B_j\right) \cap E\right) \right) + \\
&\quad \sum_{j \in [m]} \beta_j \left(\left(\sum_{i \in [n]} \mu(A_i \cap B_j \cap E) \right) + \mu\left(\left(B_j \setminus \bigcup_i A_i\right) \cap E\right) \right) = \\
&\quad \sum_{i \in [n]} \alpha_i \mu(A_i \cap E) + \sum_{j \in [m]} \beta_j \mu(B_j \cap E) = \int_E f \, d\mu + \int_E g \, d\mu,
\end{aligned}$$

gdzie przedostatnia równość wynika z tego, że najpierw sumujemy wszystkie części zbioru A_i będące częścią kolejnych B_j , a później dodajemy to, co nie należy do żadnego B_j . Analogicznie zwiżamy drugą sumę.

Przejdźcie na funkcje dodatnie: niech $f, g : \Omega \rightarrow [0, +\infty)$. Istnieje ciąg $(s_i)_{i=1}^\infty$ funkcji prostych, który aproksymuje f i analogicznie $(t_i)_{i=1}^\infty$ dla g . Zauważmy, że $(s_i + t_i)_{i=1}^\infty$ aproksymuje $f + g$. Mamy

$$\begin{aligned}
\int_E f \, d\mu + \int_E g \, d\mu &= \lim_{n \rightarrow \infty} \int_E s_n \, d\mu + \lim_{n \rightarrow \infty} \int_E t_n \, d\mu = \lim_{n \rightarrow \infty} \left(\int_E s_n \, d\mu + \int_E t_n \, d\mu \right) \\
&= \lim_{n \rightarrow \infty} \int_E s_n + t_n \, d\mu = \int_E f + g \, d\mu.
\end{aligned}$$

Przejdźcie na funkcje rzeczywiste i zespolone prosto z definicji. \square

Twierdzenie 25. Niech h będzie mierzalną funkcją nieujemną na Ω . Jeśli dla zbioru mierzalnego E jest $\int_E h \, d\mu = 0$, to $\mu(\{x \in E : h(x) \neq 0\}) = 0$.

Dowód. Niech $B = \{x \in \mathbb{R} : h(x) \neq 0\}$. Nie wprost $\mu(B) > 0$. Niech $A_n = \{x \in \mathbb{R} : \frac{1}{n+1} \leq h(x) < \frac{1}{n}\}$ oraz $A_0 = \{x \in \mathbb{R} : 1 \leq h(x)\}$. Mamy $B = \bigcup_{n=1}^\infty A_n \cup A_0$. Zbiory $\{A_n\}_{n \in \mathbb{N}}$ są parami rozłączne, zatem istnieje takie $i \in \mathbb{N}$, że $\mu(A_i) > 0$ (bo inaczej $\mu(B) = 0$).

Teraz $0 = \int_E h \, d\mu \geq \int_{A_i} h \, d\mu \geq \int_E \frac{1}{n+1} \chi_{A_i} \, d\mu = \frac{1}{n+1} \mu(A_i) > 0$. Sprzeczność \nmid . \square

Twierdzenie 26. Każdy zbiór otwarty w \mathbb{R} jest co najwyżej przeliczalną sumą przedziałów otwartych.

Dowód. Niech $U \subseteq \mathbb{R}$ będzie otwarty. Każdemu $x \in \mathbb{R}$ przypisujemy I_x – maksymalny przedział otwarty zawierający x zawarty w U (istnieje jakiś z otwartości, a suma wszystkich też jest przedziałem). Dla $x \neq y$ mamy $I_x = I_y$ albo $I_x \cap I_y = \emptyset$, bo jak się trochę przecinają, to można rozszerzyć do całości. Zatem U jest sumą rozłącznych przedziałów, każdy zawiera liczbę wymierną, więc jest ich co najwyżej przeliczalnie wiele. \square

Twierdzenie 27. Liczba elementów skończonej algebry Σ jest potęgą dwójki.

Dowód. Niech Σ będzie algebrą nad zbiorem X . Dla każdego $x \in X$ definiujemy $M_x = \bigcap_{M \in \Sigma} M$. Mamy $M_x \in \Sigma$, bo algebry są zamknięte na przecięcia.

Zbiór $F = \{M_x\}_{x \in X}$ jest podziałem X , bo jeśli $M_x \cap M_y \neq \emptyset$ dla pewnych $x, y \in X$, to mamy $y \notin M_x$ lub $x \notin M_y$ (bo inaczej każdy zbiór zawierający x zawiera y i na odwrót, czyli ich przecięcia będą takie same). Teraz $M_x \setminus M_y = M_x \cap \overline{M_y}$ lub $M_y \setminus M_x$ daje sprzeczność z minimalnością tych zbiorów.

Dla dowolnego $M \in \Sigma$ mamy $M = \bigcup_{x \in M} M_x$, bo wszystkie elementy M są też w tej sumie, a każdy element sumy jest podzbiorem M . Zatem każde M można przedstawić jako sumę rozdzielných zbiorów, co daje nam $\|\Sigma\| = 2^{\|F\|}$. \square

Zajęcia 8: Wariancja

2024-10-24

Propozycja 10. Zmienna X o rozkładzie geometrycznym z parametrem p ma wartość oczekiwaną $E[X] = \frac{1}{p}$.

Dowód. Rozkład jest bez pamięci, a więc $E[X] = 1 \cdot p + (1-p)(1 + E[X])$, bo po nieudanej próbie wracamy do stanu początkowego. Z tego wynika, że $E[X] = \frac{1}{p}$. \square

Przykład (Problem kolekcjonera kuponów). Kupujemy paczki z kuponami. W każdej jest jeden z n kuponów. Prawdopodobieństwo, że i -ty kupon jest w kupionej paczce wynosi $\frac{1}{n}$. Jaka jest oczekiwana liczba paczek, które kupimy przed skolekcjonowaniem wszystkich kuponów?

Dowód. Niech X – liczba kupionych paczek do uzyskania n kuponów. Rozbijmy na $X = \sum_{i=1}^n X_i$ gdzie X_i – liczba kupionych paczek przy posiadaniu $i-1$ kuponów (czyli przed uzyskaniem i -tego kuponu). Zmienna X_i ma rozkład geometryczny, prawdopodobieństwo sukcesu $\frac{n-i+1}{n}$ (losujemy do wystąpienia sukcesu). Zatem mamy $E[X] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \frac{n}{n-i+1} = nH_n \approx n \log n$.

Przykład (Złożoność Quicksorta). Mamy algorytm sortujący Quicksort (losujemy pivota, dzielimy na mniejsze i większe i odpalamy rekurencyjnie). Jaka jest oczekiwana liczba wykonanych porównań na danych rozmiaru n ?

Dowód. Rozważamy już posortowane wartości x_1, \dots, x_n . Zmienna X – liczba wykonanych porównań. Rozbijamy na $X_{ij} = \begin{cases} 1 & x_i, x_j \text{ były porównane} \\ 0 & \text{wpp} \end{cases}$ Mamy $E[X] = \sum_{i,j} P(X_{ij} = 1)$.

Elementy x_i, x_j są porównywane, jeśli jako pivot zostanie wybrany jeden z nich. Nie zostaną porównane, gdy zostanie wybrany element pomiędzy nimi. Zanim to zostanie zdecydowane algorytm może wybrać inne pivoty, leżące na lewo lub na prawo od rozważanych elementów. Decyzja o porównaniu dzieje się, gdy jako pivot zostaje wybrany element z $\{x_i, x_{i+1}, \dots, x_j\}$. Mamy $P(Q \text{ wybrał za pivot } x_i \text{ lub } x_j \mid Q \text{ wybrał pivota spośród } \{x_i, \dots, x_j\}) = \frac{\frac{2}{j-i+1}}{\frac{\|Z\|}{\|Z\|}} = \frac{2}{j-i+1}$, gdzie Z jest zbiorem wartości, na których odpalony jest algorytm w momencie decyzji, czy elementy zostaną porównane. Jak widać wielkość Z nie ma znaczenia.

Pozostaje nam policzyć $E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} = \sum_{k=2}^n \sum_{i=1}^{n-k+1} \frac{2}{k} = \sum_{k=2}^n (n+1-k) \cdot \frac{2}{k} = 2n \ln n + \Theta(n)$. To daje nam szukaną złożoność.

Twierdzenie 28 (Nierówność Markowa). Niech X będzie zmienną losową przyjmującą wartości nieujemne. Dla każdej stałej $a > 0$ zachodzi

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Dowód. Ustalmy a . Indykator $I = \begin{cases} 1 & X \geq a \\ 0 & \text{wpp} \end{cases}$. Z $X \geq 0$ mamy $I \leq \frac{X}{a}$ (sprawdzamy przypadki dla obu wartości I). Mamy $P(X \geq a) = P(I = 1) = E[I] \leq E\left[\frac{X}{a}\right] = \frac{E[X]}{a}$. \square

Uwaga. Jest to przykład nierówności koncentrującej. Takie nierówności mówią o tym, jak bardzo wartość zmiennej może odbiegać od wartości oczekiwanej.

Przykład. Rzucamy monetą n razy. Zliczamy orły. Jak bardzo możemy ograniczyć prawdopodobieństwo otrzymania dużej liczby orłów?

Dowód. Niech X_i będzie indykatorem mówiącym, czy za i -tym razem wypadł orzeł. Szukana zmienna losowa to $X = \sum_i X_i$. Mamy $E[X] = \sum_{i=1}^n E[X_i] = \frac{n}{2}$. Zatem z nierówności

Markowa mamy $P(X \geq \frac{3}{4}n) \leq \frac{\frac{n}{2}}{\frac{3}{4}n} = \frac{2}{3}$.

Definicja 27. k -ty moment zmiennej losowej X to $E[X^k]$.

Definicja 28. Wariancja zmiennej X to wartość

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

Definicja 29. Odchylenie standardowe zmiennej X to wartość

$$\sigma = \sqrt{\text{Var}(X)}.$$

Przykład. Niech $X = c$ będzie zmienną stałą. Mamy $E[X] = c$ oraz $\text{Var}(X) = c^2 - c^2 = 0$.

Z drugiej strony jeśli $X = \begin{cases} k \cdot c & \frac{1}{k} \\ 0 & \frac{k-1}{k} \end{cases}$, to $E[X] = c$ oraz $\text{Var}(X) = (kc)^2 \cdot \frac{1}{k} - c^2 = c^2(k-1)$.
Widzimy zatem, że zmienne o tej samej wartości oczekiwanej mogą mieć bardzo różną wariancję.

Definicja 30. Kowariancja zmiennych losowych X, Y to wartość

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

Lemat 1. Dla dowolnych zmiennych losowych X, Y zachodzi

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

Dowód.

$$\begin{aligned} \text{Var}(X + Y) &= E[(X + Y - E[X + Y])^2] = E[((X - E[X]) + (Y - E[Y]))^2] \\ &= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2\text{Cov}(X, Y). \end{aligned}$$

□

Lemat 2. Dla dowolnych niezależnych zmiennych losowych X, Y zachodzi

$$E[XY] = E[X]E[Y].$$

Dowód.

$$\begin{aligned} E[XY] &= \sum_i \sum_j ijP(X = i \cap Y = j) = \sum_i \sum_j ijP(X = i)P(Y = j) \\ &= \sum_i iP(X = i) \sum_j jP(Y = j) = E[X]E[Y]. \end{aligned}$$

□

Wniosek. Dla niezależnych zmiennych losowych X, Y zachodzi

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

Wniosek. Dla niezależnych zmiennych losowych X, Y zachodzi

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Przykład. Zmienna X o rozkładzie dwumianowym z parametrami n, p ma wariancję

$$\text{Var}(X) = np(1-p).$$

Dowód. Mamy

$$\begin{aligned} E[X^2] &= \sum_{j=1}^n \binom{n}{j} \cdot p^j (1-p)^{n-j} j^2 = \sum_{j=1}^n \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} \cdot j(j-1) + \sum_{j=1}^n \frac{n!}{j!(n-j)!} p^j \cdot j \\ &= n(n-1)p^2 \sum_{j=2}^n \frac{(n-2)!}{(j-2)!(n-2-(j-2))!} p^{j-2} (1-p)^{n-2-(j-2)} \\ &\quad + np \sum_{j=1}^n \frac{(n-1)!}{(j-1)!(n-1-(j-1))!} p^{j-1} (1-p)^{n-1-(j-1)} \\ &= n(n-1)p^2 [p + (1-p)]^{n-2} + np [p + (1-p)]^{n-1} = n(n-1)p^2 + np \end{aligned}$$

Z tego mamy $\text{Var}(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p)$.

Można też prościej rozbić na pojedyncze próby:

$$\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n (p - p^2) = np(1-p).$$

Zajęcia 9: Nierówności koncentrujące

2024-10-25

Twierdzenie 29 (Nierówność Czebyszewa). Niech $a > 0$. Dla zmiennej losowej X mamy

$$P(\|X - E[X]\| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

Dowód.

$$P(\|X - E(X)\| \geq a) = P\left((X - E[X])^2 \geq a^2\right) \leq \frac{E[(X - E[X])^2]}{a^2} = \frac{\text{Var}(X)}{a^2},$$

gdzie nierówność to nierówność Markowa. □

Definicja 31. Funkcja tworząca/charakterystyczna momentów zmiennej losowej X to funkcja

$$M_X(t) = E[e^{tX}],$$

gdzie $e^{tX} = \sum_{n=0}^{\infty} \frac{(tX)^n}{n!} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \cdot X^n$

Twierdzenie 30. Wartość oczekiwaną i różniczkowanie można zamieniać miejscami, o ile $M_X(t)$ istnieje w otoczeniu 0.

Uwaga. Wszystkie zmienne, które rozważamy, spełniają to twierdzenie.

Twierdzenie 31. Dla $n > 0$ mamy $E[X^n] = M_X^{(n)}(0)$.

Dowód.

$$(M_X(t))' = (E[e^{tX}])' = E[(e^{tX})'] = E[Xe^{tX}]$$

$$(M_X(0))' = E[X]$$

Analogicznie $(M_X(t))^{(n)} = E[X^n e^{tX}]$. □

Twierdzenie 32. Niech X, Y będą zmiennymi losowymi. Jeśli $\exists_{\delta > 0} \forall_{t \in (\delta; \delta)} M_X(t) = M_Y(t)$, to $X \sim Y$ (mają ten sam rozkład).

Lemat 3. Jeśli zmienne X, Y są niezależne, to zachodzi

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t).$$

Dowód.

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}] = E[e^{tX}] \cdot E[e^{tY}] = M_X(t) \cdot M_Y(t),$$

gdzie przedostatnie przejście wynika z tego, że jeśli X, Y są niezależne, to ich funkcje też. \square

Przykład. Niech $Y \sim \text{Geo}(p)$. Wyznamy momenty Y korzystając z funkcji tworzącej.

$$M_Y(t) = E[e^{tY}] = \sum_{k=1}^{\infty} P(Y=k) \cdot e^{tk} = \sum_{k=1}^{\infty} p(1-p)^{k-1} e^{tk} = \frac{p}{1-p} \sum_{k=1}^{\infty} ((1-p)e^t)^k = \frac{p}{1-p} \left(\frac{1}{1-(1-p)e^t} - 1 \right)$$

$$\text{Mamy } (M_Y(t))' = \frac{p}{1-p} \cdot - \left(\frac{1}{1-(1-p)e^t} \right)^2 \cdot e^t(-(1-p)) = pe^t \frac{1}{(1-(1-p)e^t)^2}.$$

$$\text{Tak samo } (M_Y(t))'' = 2p(1-p)(1-(1-p)e^t)^{-3}e^{2t} + p(1-(1-p)e^t)^{-2}e^t.$$

$$\text{Z tego wynika, że } E[Y] = \frac{1}{p} \text{ oraz } E[Y^2] = 2p^{-2}(1-p) + p^{-1}.$$

Twierdzenie 33 (Nierówność Czernowa). Dla zmiennej losowej X i dowolnego a zachodzi

$$P(X \geq a) \leq \inf_{t>0} \frac{M_X(t)}{e^{ta}}$$

oraz

$$P(X \leq a) \leq \inf_{t<0} \frac{M_X(t)}{e^{ta}}.$$

Dowód. Dla $t > 0$ zachodzi $P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} = \frac{M_X(t)}{e^{ta}}$, gdzie pierwsza równość wynika z monotoniczności e^z a nierówność jest zastosowaniem nierówności Markowa. Taka nierówność zachodzi dla dowolnego wyboru t , więc można wziąć infimum.

Analogicznie dla $t < 0$ funkcja e^{-z} jest antymonotoniczna i z tego wynika drugie ograniczenie. \square

Przykład. Niech X będzie liczbą orłów w n rzutach monetą. Ograniczymy $P(X \geq \frac{3}{4}n)$. Mamy $E[X] = \frac{n}{2}$. Rozbijając na pojedyncze rzuty dostajemy $\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = \sum_{i=1}^n E[X_i^2] - E[X_i]^2 = n(\frac{1}{2} - \frac{1}{4}) = \frac{n}{4}$.

Z nierówności Markowa mamy $P(X \geq \frac{3}{4}n) \leq \frac{2}{3}$.

Z nierówności Czebyszewa $P(X \geq \frac{3}{4}n) \leq P(\|X - \frac{n}{2}\| \geq \frac{n}{4}) \leq \frac{4}{n}$.

Funkcję tworzącą możemy ograniczyć przez $M_{X_i}(t) = E[e^{tX_i}] = e^t \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2} \cdot (1 + e^t) = 1 + \frac{1}{2}(e^t - 1) \leq e^{\frac{1}{2}(e^t - 1)}$, gdzie pod koniec korzystamy z nierówności $1 + x \leq e^x$.

$$\text{Teraz } M_X(t) = \prod_{i=1}^n M_{X_i}(t) \leq \prod_{i=1}^n e^{\frac{1}{2}(e^t - 1)} = \left(e^{\frac{1}{2}(e^t - 1)} \right)^n = e^{\frac{1}{2}n(e^t - 1)}.$$

Z nierówności Czernowa $P(X \geq \frac{3}{4}n) \leq \frac{e^{\frac{1}{2}n(e^t - 1)}}{e^{\frac{3}{4}nt}} = e^{\frac{1}{2}ne^t - \frac{1}{2}n - \frac{3}{4}nt} = e^{\frac{1}{2}nf(t)}$, gdzie $f(t) = e^t - 1 - \frac{3}{2}t$.

$f(t)$ przyjmuje najmniejszą wartość dla $t = \ln \frac{3}{2} \approx 0,4$, co ostatecznie daje ograniczenie

$$P\left(X \geq \frac{3}{4}n\right) \leq e^{-\frac{n}{20}}.$$

Przykład. Niech Y będzie liczbą kupionych kuponów w problemie kolekcjonera kuponów dla n

kuponów. Ograniczymy $P(Y \geq 2nH_n)$. Wiemy, że $E[Y] = nH_n$. Przeliczamy

$$E[Y_i^2] = \sum_{i=1}^{\infty} p(1-p)^{i-1} i^2 = \frac{p}{1-p} \sum_{i=1}^{\infty} (1-p)^i i^2 = \frac{2-p}{p^2},$$

gdzie sumę zwiniliśmy stosując funkcje tworzące. Mamy $\text{Var}(Y_i) \leq \frac{1}{p^2}$, czyli

$$\text{Var}(Y) \leq \sum_{i=1}^n \left(\frac{n}{n-i+1} \right)^2 = n^2 \sum_{i=1}^n \frac{1}{i^2} \leq n^2 \frac{\pi^2}{6}.$$

Z nierówności Markowa mamy $P(Y \geq 2nH_n) \leq \frac{1}{2}$.

Z nierówności Czebyszewa $P(Y \geq 2nH_n) \leq P(\|Y - nH_n\| \geq nH_n) \leq \frac{n^2 \frac{\pi^2}{6}}{(nH_n)^2} = O\left(\frac{1}{(\ln n)^2}\right)$.

Nierówność Czernowa jest trudna, ale daje $O(e^{-n})$.

Stosując union bound można dostać sensowne ograniczenie, mamy

$$P(Y_i \geq 2nH_n) = \left(1 - \frac{1}{n}\right)^{2nH_n} \approx \left(1 - \frac{1}{n}\right)^{2n \ln n} < e^{-2 \ln n} = \frac{1}{n^2},$$

gdzie pierwsza równość wynika z tego, że przez pierwsze $2nH_n$ kroków musimy dostać inny kupon niż i .

Teraz mamy $P(Y \geq 2nH_n) = P(\bigcup_{i=1}^n \{Y_i \geq 2nH_n\}) \leq \sum_{i=1}^n P(Y_i \geq 2nH_n) \leq n \cdot \frac{1}{n^2} = \frac{1}{n}$, gdzie pierwsza nierówność wynika z tego, że jeśli całość została osiągnięta po takiej liczbie kroków, to któryś z pojedynczych kuponów też został po niej osiągnięty.

Tabela 2: Ograniczenia przez różne nierówności

Model	Rzut monetą	Kolekcjoner kuponów
Rozkład	Uni($\{0, 1\}$)	Geo($\frac{n+i-1}{n}$)
Nierówność Markowa	$\frac{2}{3}$	$\frac{1}{2}$
Nierówność Czebyszewa	$\frac{4}{n}$	$O\left(\frac{1}{(\ln n)^2}\right)$
Nierówność Czernowa	$e^{-\frac{n}{20}}$	$O(e^{-n})$
Union bound	–	$\frac{1}{n}$

Przykład (Próby Poissona). Pojedyncze rzuty różnymi niesymetrycznymi monetami. Mamy wartości $p_1, \dots, p_n \in (0, 1)$, zmienne losowe X_i takie, że $P(X_i = 1) = p_i$, $X_i \in \{0, 1\}$.

Definiujemy $X = \sum_{i=1}^n X_i$, mamy $E[X] = \sum_{i=1}^n p_i = \mu$. Ograniczamy funkcję tworzącą:

$$M_{X_i}(t) = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t - 1)}.$$

$$M_X(t) = \prod_{i=1}^n M_{X_i}(t) \leq \prod_{i=1}^n e^{p_i(e^t - 1)} = e^{\mu(e^t - 1)}.$$

Zajęcia 10: Zestaw 4

2024-10-29

Twierdzenie 34. Niech X będzie zmienną losową liczącą liczbę kopii kliku K_4 na czterech wierzchołkach w grafie wylosowanym z modelu $G_{n,p}$ (to znaczy, mając $n > 4$ wierzchołków dla każdej pary wierzchołków losujemy niezależnie z prawdopodobieństwem $p \in (0, 1)$, czy są połączone krawędzią, czy nie). Wariancja zmiennej losowej X wynosi

$$\text{Var}(X) = \binom{n}{4} \left[p^6 + 4(n-4)p^9 + 6\binom{n-4}{2}p^{11} + \left[4\binom{n-4}{3} + \binom{n-4}{4} - 1 \right] p^{12} \right].$$

Dowód. Rozbijamy naszą zmienną X na sumę $X = \sum_{\{i,j,k,l\} \in \binom{[n]}{4}} X_{ijkl}$, gdzie X_{ijkl} jest indykato-rem mówiącym, czy na wybranych wierzchołkach jest klika. Mamy

$$E[X] = \sum_{\{i,j,k,l\} \in \binom{[n]}{4}} E[X_{ijkl}] = \binom{n}{4} \cdot p^6.$$

Liczymy $E[X^2]$ rozbijając na indykatory:

$$E[X^2] = \sum_{\substack{\{ijkl\} \in \binom{[n]}{4} \\ \{efds\} \in \binom{[n]}{4}}} E[X_{ijkl} X_{efds}] = \binom{n}{4} \left[\binom{4}{4} \binom{n-4}{0} p^6 + \binom{4}{3} \binom{n-4}{1} p^9 + \binom{4}{2} \binom{n-4}{2} p^{11} + \binom{4}{1} \binom{n-4}{3} p^{12} + \binom{4}{0} \binom{n-4}{4} p^{12} \right]$$

To rozbiecie otrzymujemy w taki sposób, że najpierw wyznaczamy pierwszą czwórkę, potem ile druga ma z nią wspólnych wierzchołków, a na koniec wybieramy pozostałe.

Po przeliczeniu dostajemy

$$\text{Var}(X) = \binom{n}{4} \left[p^6 + 4(n-4)p^9 + 6\binom{n-4}{2}p^{11} + \left[4\binom{n-4}{3} + \binom{n-4}{4} - 1 \right] p^{12} \right].$$

□

Twierdzenie 35 (Rozkład hipergeometryczny). W urnie znajduje się N kul, z czego K jest zielonych. Losujemy kule bez zwracania. Niech X będzie liczbą zielonych kul, które wylosujemy, jeśli wyciągnęliśmy n kul. Zachodzi

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}},$$

a do tego

$$E[X] = \frac{nK}{N},$$

$$\text{Var}(X) = \frac{Kn}{N} + \frac{K(K-1)n(n-1)}{N(N-1)} - \frac{K^2n^2}{N^2}.$$

Dowód. Prawdopodobieństwo wynika z tego, że aby mieć k zielonych kul musimy najpierw z K zielonych kul wybrać k , a następnie dopełnić je $n-k$ kulami z reszty. Wszystkich możliwych wyborów kul jest $\binom{N}{n}$.

Niech X_i będzie indykato-rem tego, czy i -ta zielona kula została wzięta. Mamy

$$E[X] = \sum_{i=1}^K E[X_i] = \sum_{i=1}^K \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{Kn}{N}.$$

Podobnie

$$E[X^2] = \sum_{i,j \in [K]} E[X_i X_j] = K \frac{\binom{N-1}{n-1}}{\binom{N}{n}} + K(K-1) \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{Kn}{N} + \frac{K(K-1)n(n-1)}{N(N-1)},$$

co daje odpowiednią wartość $\text{Var}(X)$. □

Twierdzenie 36 (Słabe prawo wielkich liczb). Niech X_1, X_2, \dots będą niezależnymi zmiennymi losowymi z tego samego rozkładu o skończonej wartości oczekiwanej μ oraz skończonym odchyleniu standardowym σ . Dla każdego $\varepsilon > 0$ zachodzi

$$\lim_{n \rightarrow \infty} P\left(\left\| \frac{X_1 + \dots + X_n}{n} - \mu \right\| \geq \varepsilon\right) = 0.$$

Dowód. Zachodzi

$$\text{Var} \left(\frac{X_1 + \dots + X_n}{n} \right) = \frac{1}{n^2} \text{Var} (X_1 + \dots + X_n) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n},$$

gdzie pierwszej przejście jest prostym przeliczeniem z definicji wariancji i liniowości wartości oczekiwanej, a drugie wynika z niezależności rozważanych zmiennych. Na mocy nierówności Czebyszewa mamy

$$P \left(\left\| \frac{X_1 + \dots + X_n}{n} - \mu \right\| \geq \varepsilon \right) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0.$$

□

Zajęcia 11: Nierówności Czernowa

2024-11-05

Twierdzenie 37. Niech X_1, \dots, X_n będą niezależnymi zmiennymi takimi, że $P(X_i = 1) = p_i$ oraz $P(X_i = 0) = 1 - p_i$. Niech $X = \sum_{i=1}^n X_i$ oraz $\mu = E[X]$. Wtedy:

1. $\forall_{\delta > 0} P(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu$
2. $\forall_{\delta \in (0, 1]} P(X \geq (1 + \delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}}$
3. $\forall_{R \geq 6\mu} P(X \geq R) \leq 2^{-R}$.

Dowód. Liczymy funkcję tworzącą

$$M_{X_i}(t) = E[e^{tX_i}] = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t - 1)}.$$

Zatem

$$M_X(t) = \prod_{i=1}^n M_{X_i}(t) \leq \prod_{i=1}^n e^{p_i(e^t - 1)} = e^{(e^t - 1)\mu}.$$

Ustalmy $t > 0$, mamy

$$P(X \geq (1 + \delta)\mu) = P(e^{tX} \geq e^{t(1 + \delta)\mu}) \leq \frac{E[e^{tX}]}{e^{t(1 + \delta)\mu}} \leq \frac{e^{(e^t - 1)\mu}}{e^{t(1 + \delta)\mu}}.$$

Niech $t = \ln(1 + \delta) > 0$. Wychodzi nam $P(X \geq (1 + \delta)\mu) \leq \left(\frac{e^{1 + \delta - 1}}{(1 + \delta)^{1 + \delta}} \right)^\mu$, co kończy dowód pierwszej części.

Punkt drugi dowodzimy korzystając z pierwszego, wystarczy pokazać, że dla $\delta \in (0, 1]$ jest

$$\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \leq e^{-\frac{\delta^2}{3}}.$$

Logarytmujemy stronami, chcemy pokazać, że $\delta - (1 + \delta) \ln(1 + \delta) + \frac{\delta^2}{3} \leq 0$. Oznaczmy lewą stronę przez $f(\delta)$. Liczymy pochodne:

$$f'(\delta) = 1 - 1 \cdot \ln(1 + \delta) - \frac{1 + \delta}{1 + \delta} + \frac{2}{3}\delta = -\ln(1 + \delta) + \frac{2}{3}\delta,$$

$$f''(\delta) = -\frac{1}{1 + \delta} + \frac{2}{3}.$$

$f'(0) = 0$, a potem maleje do $\delta = \frac{1}{2}$ (tam druga pochodna się zeruje, przedtem ujemna), potem rośnie, ale $f'(1) < 0$, więc jest ujemna na całym $(0, 1]$.

$f(\delta)$ tylko maleje na $(0, 1]$, więc nierówność działa, bo $f(0) = 0$.

Dowodząc punkt trzeci zakładamy $R \geq 6\mu$. Niech $R = (1 + \delta)\mu$, czyli $\delta = \frac{R}{\mu} - 1 \geq 5$.

$$P(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu \leq \left(\frac{e}{1 + \delta} \right)^{(1+\delta)\mu} \leq \left(\frac{1}{2} \right)^R = 2^{-R}.$$

□

Twierdzenie 38. Niech X_1, \dots, X_n będą niezależnymi zmiennymi takimi, że $P(X_i = 1) = p_i$ oraz $P(X_i = 0) = 1 - p_i$. Niech $X = \sum_{i=1}^n X_i$, $\mu = E[X]$. Wtedy dla każdego $\delta \in (0, 1)$ zachodzi:

1. $P(X \leq (1 - \delta)\mu) \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}} \right)^\mu$
2. $P(X \leq (1 - \delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}.$

Dowód. Dowód identyczny jak w poprzednim twierdzeniu, wybieramy $t = \ln(1 - \delta) < 0$ i korzystamy z tego, że e^{-z} jest antymonotoniczne. Drugiego punktu ponownie dowodzimy licząc pochodne i na ich podstawie dowodząc odpowiedniej nierówności. □

Wniosek. Dla $\delta \in (0, 1)$ mamy $P(\|X - \mu\| \geq \delta\mu) \leq 2 \cdot e^{-\frac{\mu\delta^2}{3}}$ (ograniczamy przez większą z wartości z poprzednich twierdzeń).

Przykład. Rozważamy n niezależnych rzutów monetą. Mamy $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$. Zliczamy orły, z nierówności Czernowa dostajemy

$$P\left(\left\|X - \frac{n}{2}\right\| \geq \text{val}\right) \leq \frac{c}{n}.$$

Wybieramy val tak, by było jak najmniejsze i mimo to dalej dawało liniowe ograniczenie. Bierzemy $\frac{1}{2}\sqrt{6n \ln n}$, co daje

$$P\left(\left\|X - \frac{n}{2}\right\| \geq \frac{1}{2}\sqrt{6n \ln n}\right) \leq \frac{2}{n},$$

bo podstawiając $\frac{1}{2}\sqrt{6n \ln n} = \delta \frac{n}{2}$ dostajemy $\delta = \sqrt{\frac{6 \ln n}{n}}$ (dla odpowiednio dużego n należy do $(0, 1)$), ograniczenie $2 \cdot e^{-\frac{n}{2} \cdot 6 \frac{\ln n}{n} \cdot \frac{1}{3}} = \frac{2}{n}$.

Wzięliśmy ograniczenie pierwiastkowe, bo liniowe dałoby za dużą wartość (wolną koncentrację zmiennej).

Przykład. Z nierówności Czebyszewa dostaliśmy $P\left(\left\|X - \frac{n}{2}\right\| \geq \frac{n}{4}\right) \leq \frac{4}{n}$. Tutaj mamy dużo większe odchylenie (liniowe), które ma małe prawdopodobieństwo. Czernow pokazuje, że dużo mniejsze odchylenie jest równie mało prawdopodobne.

Dla $P\left(\left\|X - \frac{n}{2}\right\| \geq \frac{n}{4}\right)$ mamy ograniczenie z Czernowa $2 \cdot e^{-\frac{n}{2} \cdot \left(\frac{1}{2}\right)^2 \cdot \frac{1}{3}} \leq 2 \cdot e^{-\frac{n}{24}}$.

Przykład (Estymacja parametrów). Bierzemy z dużej populacji próbkę wielkości n wybraną w sposób jednostajny. p to nieznaną wartość – szukane prawdopodobieństwo, które chcemy szacować (np. prawdopodobieństwo jakiejś mutacji genetycznej) przez \hat{p} . Zmienna losowa $X = \hat{p}n$. Spodziewamy się, że jak n rośnie, to $\hat{p} \rightarrow p$.

Definicja 32. Mówimy, że $[\hat{p} - \delta, \hat{p} + \delta]$ jest $(1 - \gamma)$ przedziałem ufności dla parametru p , jeśli $P(p \in [\hat{p} - \delta, \hat{p} + \delta]) \geq 1 - \gamma$. Chcemy, żeby n, γ, δ były małe, ale musi być między nimi jakiś balans.

Przykład. Jeśli $p < \hat{p} - \delta$, to $X = n\hat{p} > n(p + \delta) = np \cdot \left(1 + \frac{\delta}{p}\right)$.

Jeśli $p > \hat{p} + \delta$, to $X = n\hat{p} < n(p - \delta) = np \cdot \left(1 - \frac{\delta}{p}\right)$.

Mamy $E[X] = np$, a więc Czernow daje

$$\begin{aligned} P(p \notin [\hat{p} - \delta, \hat{p} + \delta]) &= P\left(X < np\left(1 - \frac{\delta}{p}\right)\right) + P\left(X > np\left(1 + \frac{\delta}{p}\right)\right) \\ &\leq 2 \cdot e^{-np \cdot \left(\frac{\delta}{p}\right)^2 \cdot \frac{1}{3}} = 2 \cdot e^{-n \frac{\delta^2}{p} \cdot \frac{1}{3}} \leq 2 \cdot e^{-n \frac{\delta^2}{3}} = \gamma. \end{aligned}$$

Pod koniec wzięliśmy $p = 1$, bo daje najgorsze ograniczenie. W ten sposób związaliśmy ze sobą wartości n, γ, δ .

Twierdzenie 39. Niech X_1, \dots, X_n będą niezależnymi zmiennymi losowymi, dla których $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Niech $X = \sum_{i=1}^n X_i$. Dla każdego $a > 0$ mamy $P(X \geq a) \leq e^{-\frac{a^2}{2n}}$. Zauważmy, że nie ma sensu rozważać tu odchyleń multiplikatywnych, bo $E[X] = 0$.

Dowód. Mamy $E[e^{tX}] = -\frac{1}{2}e^t + \frac{1}{2}e^{-t}$.

Rozwijamy w szereg Taylora:

$$\begin{aligned} e^t &= 1 + t + \frac{t^2}{2} + \dots + \frac{t^i}{i!} + \dots \\ e^{-t} &= 1 - t + \frac{t^2}{2} + \dots + (-1)^i \frac{t^i}{i!} + \dots \end{aligned}$$

Z tego wynika

$$E[e^{tX_i}] = \sum_{i \geq 0} \frac{t^{2i}}{(2i)!} \leq \sum_{i \geq 0} \frac{\left(\frac{t^2}{2}\right)^i}{i!} = e^{\frac{t^2}{2}},$$

gdzie w nierówności wyciągnęliśmy 2 z dwukrotności każdej liczby od 1 do i , a pozostałe składniki zignorowaliśmy.

Zatem $E[e^{tX}] = \prod_{i=1}^n e^{tX_i} \leq e^{\frac{t^2 n}{2}}$.

Dostajemy $P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} \leq e^{t^2 n \cdot \frac{1}{2} - ta} = e^{a^2 \frac{1}{n} \cdot \frac{1}{2} - \frac{a^2}{n}} = e^{-a^2 \cdot \frac{1}{2n}}$ gdzie podstawiliśmy $t = \frac{a}{n} > 0$. \square

Uwaga. Symetrycznie dowodzimy $P(X \leq -a) \leq e^{-\frac{a^2}{2n}}$ (bo $-X = X$), więc zachodzi również nierówność $P(\|X\| \geq a) \leq 2 \cdot e^{-\frac{a^2}{2n}}$.

Wniosek. Niech Y_1, \dots, Y_n będą niezależnymi indykatorami $P(Y_i = 0) = P(Y_i = 1) = \frac{1}{2}$. Niech $Y = \sum_{i=1}^n Y_i$, $\mu = E[Y] = \frac{n}{2}$. Wtedy

1. $\forall a > 0 \quad P(Y \geq \mu + a) \leq e^{-\frac{2a^2}{n}}$
2. $\forall \delta > 0 \quad P(Y \geq (1 + \delta)\mu) \leq e^{-\delta^2 \mu}$

Dowód. Bierzemy $Y_i = \frac{X_i + 1}{2}$. Mamy $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Dla $X = \sum_{i=1}^n X_i = 2Y - \mu$ mamy

$$P(Y \geq \mu + a) = P(X \geq 2a) \leq e^{-\frac{2a^2}{n}}$$

oraz

$$P(Y \geq (1 + \delta)\mu) = P(X \geq 2\delta\mu) \leq e^{-\frac{2\delta^2 \mu^2}{n}} = e^{-\delta^2 \mu}.$$

\square

Przykład (Set balancing). Mamy macierz $n \times m$ wypełnioną wartościami z $\{0, 1\}$.

$$\begin{bmatrix} a_1 & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix}$$

Każdy wiersz to jakaś cecha osoby, kolumna to osoba. Chcemy przeprowadzić jakieś badanie, a do tego potrzebujemy zrobić grupę badawczą i kontrolną, które będą możliwie identyczne (to znaczy o podobnym zagęszczeniu wszystkich cech).

Mnożymy tę macierz A przez wektor $\bar{b} \in \{-1, 1\}^m$ (umieszczenie kolejnych osób w jednej lub drugiej grupie) i dostajemy wektor \bar{c} , w którym będą różnice między ilością osób z daną cechą między grupami. Chcemy, żeby norma $\|\bar{c}\|_\infty$ była jak najmniejsza.

Wektor \bar{b} wyznaczamy, losując.

Twierdzenie 40. Dla losowego \bar{b} (każda współrzędna niezależnie, jednostajnie z $\{-1, 1\}$) zachodzi

$$P\left(\|A\bar{b}\|_\infty \geq \sqrt{4m \ln n}\right) \leq \frac{2}{n}.$$

Dowód. Ważne są tylko te wiersze A , gdzie jest więcej jedynek niż nasze ograniczenie (bo jak jest mniej, to wzięcie wszystkich z tym samym znakiem nic nam nie zepsuje).

Niech i -ty wiersz $\bar{a}_i = a_{i1} \dots a_{im}$ ma w sobie k jedynek.

Jeśli $k < \sqrt{4m \ln n}$, to $\|A\bar{b}\| \leq \sqrt{4m \ln n}$

Jeśli $k > \sqrt{4m \ln n}$, to $Z = \sum_{j=1}^m a_{ij}b_j$ jest sumą k zmiennych losowych, które z równym prawdopodobieństwem przyjmują 1 i -1 .

Mamy zatem $P\left(\|Z_i\| > \sqrt{4m \ln n}\right) \leq 2e^{-\frac{4m \ln n}{2k}} \leq \frac{2}{n^2}$, ostatnia nierówność wynika z $m \geq k$. Jest to ograniczenie dla jednego wiersza, dla wszystkich wierszy dostajemy z union bounda ograniczenie $\frac{2}{n}$. \square

Zajęcia 12: Rozkład Poissona

2024-11-08

Przykład. Mamy u urn i wrzucamy do nich niezależnie i jednostajnie k kul. Możliwe pytania:

- Ile jest pustych urn?
- Jakie jest maksymalne zapełnienie urny?
- Ile kul powinniśmy wrzucić aby wszystkie urny były pełne?

Przykład (Paradoks dnia urodzin). Mamy u dni w roku, k osób. Jakie jest prawdopodobieństwo, że żadne dwie osoby nie urodziły się w tym samym dniu?

$$P(X) = 1 \cdot \frac{u-1}{u} \cdot \frac{u-2}{u} \cdot \dots \cdot \frac{u-k+1}{u} = \frac{(u-1)!}{k!u^{k-1}}.$$

Wstawiając $u = 365$, $k = 23$ mamy prawdopodobieństwo około 0,4927.

Przykład. Niech $u = k = n$. Wtedy średnie zapełnienie urny to $\frac{k}{u} = 1$.

Jakie jest maksymalne zapełnienie urny?

Uwaga. Poniżej korzystamy z rozwinięcia e^x w szereg Taylora:

$$e^x = \sum_{i \in \mathbb{N}} \frac{x^i}{i!}.$$

Twierdzenie 41. Dla wystarczająco dużego n , gdy wrzucamy n kul do n urn, to prawdopodobieństwo, że maksymalne zapełnienie jest większe od $\frac{4 \ln n}{\ln \ln n}$ jest co najwyżej $\frac{1}{n}$.

Dowód. Niech $k > \frac{4 \ln n}{\ln \ln n}$. Dowiedzimy, że

$$\begin{aligned} P(\text{konkretna urna ma dokładnie } k \text{ kul}) &= \binom{n}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{n-k} \leq \frac{n^k}{k!} \left(\frac{1}{n}\right)^k = \frac{1}{k!} \leq \left(\frac{e}{k}\right)^k \\ &\leq \left(\frac{e \ln \ln n}{4 \ln n}\right)^{\frac{4 \ln n}{\ln \ln n}} \leq \exp\left(\ln\left(\frac{e \ln \ln n}{4 \ln n}\right) \cdot \frac{4 \ln n}{\ln \ln n}\right) = \exp\left(\frac{4 \ln n}{\ln \ln n} (\ln \ln \ln n - \ln \ln n)\right) \\ &= n^{-4 + \frac{4 \ln \ln \ln n}{\ln \ln n}} \leq n^{-3}. \end{aligned}$$

Najpierw wybieramy kule wrzucane do urny i mnożymy przez odpowiednie prawdopodobieństwa, potem pozbywamy się wielu czynników i korzystamy z $e^k \geq \frac{k^k}{k!}$ wynikającego z rozwinięcia Taylora. Potem korzystamy z określenia k i faktu, że k^k rośnie szybciej niż e^k , skracamy e z 4 (bo nierówność) i przeliczamy, a potem wyrażenie w wykładniku zbiega do -4 , a więc dla dużych n można je ograniczyć przez -3 .

Kończymy za pomocą union bounda:

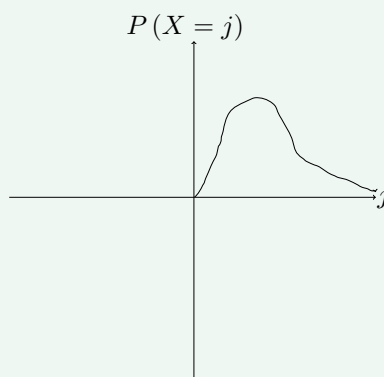
$$\begin{aligned} P\left(\text{maksymalne zapełnienie} > \frac{4 \ln n}{\ln \ln n}\right) &= P\left(\bigcup_{i \in [n]} \bigcup_{k > \frac{4 \ln n}{\ln \ln n}} \{i\text{-ta urna ma dokładnie } k \text{ kul}\}\right) \\ &\leq \sum_{\substack{i \in [n] \\ k > \frac{4 \ln n}{\ln \ln n}}} P(i\text{-ta urna ma dokładnie } k \text{ kul}) \leq \sum_{\substack{i \in [n] \\ k > \frac{4 \ln n}{\ln \ln n}}} n^{-3} \leq n^{-3} \cdot n^2 = \frac{1}{n}, \end{aligned}$$

gdzie ostatnia nierówność wynika z faktu, że każda z n urn może mieć w sobie co najwyżej n kul. \square

Definicja 33. Zmienna X ma rozkład Poissona z parametrem $\mu > 0$, jeśli

$$P(X = j) = \frac{e^{-\mu} \cdot \mu^j}{j!}$$

dla każdego $j \in \mathbb{N}$.



Rysunek 2: Rozkład Poissona. Przypomina krzywą rozkładu dwumianowego, bo jest w pewnym sensie granicą rozkładów dwumianowych.

Mamy

$$\sum_{j \in \mathbb{N}} P(X = j) = \sum_{j \in \mathbb{N}} \frac{e^{-\mu} \mu^j}{j!} = e^{-\mu} \sum_{j \in \mathbb{N}} \frac{\mu^j}{j!} = e^{-\mu} \cdot e^{\mu} = 1,$$

czyli faktycznie jest to rozkład.

Twierdzenie 42. Dla $n > \lambda$ niech X_n będzie zmienną o rozkładzie dwumianowym z parametrami $(n, \frac{\lambda}{n})$. Wtedy dla każdego $k \in \mathbb{N}$ mamy

$$\lim_{n \rightarrow \infty} P(X_n = k) = \frac{e^{-\lambda} \lambda^k}{k!}.$$

Czyli zwiększając liczbę prób w nieskończoność dostajemy rozkład Poissona.

Propozycja 11. Niech X ma rozkład Poissona z parametrem μ . Mamy

$$E[X] = \mu,$$

$$\text{Var}(X) = \mu.$$

Dowód.

$$E[X] = \sum_{j \in \mathbb{N}} j P(X = j) = \sum_{j=1}^{\infty} j \frac{e^{-\mu} \mu^j}{j!} = e^{-\mu} \cdot \mu \sum_{j=1}^{\infty} \frac{\mu^{j-1}}{(j-1)!} = e^{-\mu} \cdot \mu \sum_{j=0}^{\infty} \frac{\mu^j}{j!} = e^{-\mu} \mu e^{\mu} = \mu.$$

Podobnie

$$\begin{aligned} E[X^2] &= \sum_{j \in \mathbb{N}} j^2 \frac{e^{-\mu} \mu^j}{j!} = e^{-\mu} \mu \sum_{j=1}^{\infty} j \frac{\mu^{j-1}}{(j-1)!} = e^{-\mu} \mu \left(\sum_{j=1}^{\infty} \frac{\mu^{j-1}}{(j-1)!} + \sum_{j=1}^{\infty} (j-1) \frac{\mu^{j-1}}{(j-1)!} \right) \\ &= \mu + \mu^2. \end{aligned}$$

□

Propozycja 12. Niech X ma rozkład Poissona z parametrem μ . Mamy $M_X(t) = e^{\mu(e^t-1)}$.

Dowód.

$$M_X(t) = \sum_{k \in \mathbb{N}} \frac{e^{-\mu} \mu^k}{k!} e^{tk} = e^{-\mu} \sum_{k \in \mathbb{N}} \frac{(\mu e^t)^k}{k!} = e^{-\mu} \cdot e^{\mu e^t}.$$

□

Lemat 4. Jeśli X ma rozkład Poissona z parametrem μ_1 , a Y z μ_2 i te zmienne są niezależne, to $X + Y$ ma rozkład Poissona z parametrem $\mu_1 + \mu_2$.

Dowód.

$$\begin{aligned} P(X + Y = j) &= \sum_{k=0}^j P(X = k \cap Y = j - k) = \sum_{k=0}^j P(X = k) P(Y = j - k) \\ &= \sum_{k=0}^j \frac{e^{-\mu_1} \mu_1^k}{k!} \cdot \frac{e^{-\mu_2} \mu_2^{j-k}}{(j-k)!} = e^{-(\mu_1 + \mu_2)} \cdot \frac{1}{j!} \sum_{k=0}^j \frac{j!}{k! (j-k)!} \mu_1^k \mu_2^{j-k} = e^{-(\mu_1 + \mu_2)} \cdot \frac{1}{j!} (\mu_1 + \mu_2)^j. \end{aligned}$$

Można też rozważyć funkcję tworzącą sumy:

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = e^{\mu_1(e^t-1)} e^{\mu_2(e^t-1)} = e^{(\mu_1 + \mu_2)(e^t-1)},$$

która jest funkcją tworzącą rozkładu Poissona z odpowiednim parametrem.

□

Twierdzenie 43. Niech X będzie zmienną o rozkładzie Poissona z parametrem μ . Wtedy:

1. jeśli $x > \mu$, to $P(X \geq x) \leq \frac{e^{-\mu} (\mu e)^x}{x^x}$
2. jeśli $x < \mu$, to $P(X \leq x) \leq \frac{e^{-\mu} (\mu e)^x}{x^x}$
3. jeśli $\delta > 0$, to $P(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu$

4. jeśli $0 < \delta < 1$, to $P(X \leq (1 - \delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu$

Dowód. Niech $t > 0, x > \mu$. Mamy

$$P(X \geq x) \leq \frac{E[e^{tX}]}{e^{tx}} = e^{\mu(e^t - 1) - tx} \leq e^{\mu \frac{x}{\mu} - \mu - \ln(\frac{x}{\mu})x} = e^{-\mu} \cdot \left(\frac{e\mu}{x}\right)^x,$$

gdzie podstawiliśmy $t = \ln\left(\frac{x}{\mu}\right) > 0$. Drugi punkt robi się identycznie, wtedy mamy $\ln\left(\frac{x}{\mu}\right) < 0$.

Trzeci i czwarty punkt są po prostu podstawieniem do poprzednich. \square

Pomysł (Aproksymacja Poissona (Poissonization)). Mamy k kul, które wrzucamy jednostajnie do u urn. Robimy to z jakimiś założeniami (np. pewne urny są puste, w jakichś jest konkretna liczba kul), których czasem nie da się uwzględnić podczas liczenia prawdopodobieństwa.

Niech $X_1^{(k)}, \dots, X_u^{(k)}$ będą liczbami kul w kolejnych urnach w rozważanym rozkładzie.

Będziemy chcieli ograniczyć ich wartość przez ciąg niezależnych zmiennych o rozkładzie Poissona $Y_1^{(k)}, \dots, Y_u^{(k)}$ z parametrem $\frac{k}{u}$.

Lemat 5. Mamy ciąg niezależnych zmiennych o rozkładzie Poissona $(Y_1^{(k)}, \dots, Y_u^{(k)})$ z parametrem $\frac{k}{u}$. Przy założeniu, że $\sum_{i=1}^u Y_i^{(k)} = m$, rozkład tych zmiennych jest taki sam jak rozkład zmiennych $(X_1^{(m)}, \dots, X_u^{(m)})$ modelujących wrzucanie m kul do u urn (niezależnie od wartości k).

Dowód. Dla $\sum_{i=1}^u m_i = m$ mamy

$$P\left(\left(X_1^{(m)}, \dots, X_u^{(m)}\right) = (m_1, \dots, m_u)\right) = \frac{m!}{m_1! \cdot \dots \cdot m_u! \cdot u^m},$$

bo szansa na konkretną konfigurację to $\left(\frac{1}{u}\right)^m$, a wybierając odpowiednio kule w każdej urnie dostajemy liczbę konfiguracji spełniających nasz warunek:

$$\binom{m}{m_1} \cdot \binom{m_2 + \dots + m_u}{m_2} \cdot \binom{m_3 + \dots + m_u}{m_3} \cdot \dots \cdot \binom{m_{u-1} + m_u}{m_{u-1}} \cdot \binom{m_u}{m_u},$$

co równa się odpowiedniej wartości.

W drugim modelu jest

$$\begin{aligned} P\left(\left(Y_1^{(k)}, \dots, Y_u^{(k)}\right) = (m_1, \dots, m_u) \mid \sum_{i=1}^u Y_i^{(k)} = m\right) &= \frac{P\left(Y_1^{(k)} = m_1 \cap \dots \cap Y_u^{(k)} = m_u\right)}{P\left(\sum_{i=1}^u Y_i^{(k)} = m\right)} \\ &= \frac{\prod_{i=1}^u e^{-\frac{k}{u}} \left(\frac{k}{u}\right)^{m_i} \frac{1}{m_i!}}{e^{-k} \cdot k^m \frac{1}{m!}} = \frac{m!}{m_1! \cdot \dots \cdot m_u! \cdot u^m}, \end{aligned}$$

gdzie pierwsze przejście wynika z tego, że warunek zakładany zawiera się w tym, którego prawdopodobieństwo liczymy. Drugie przejście to zastosowanie niezależności i tego, że suma niezależnych zmiennych Poissona jest zmienną Poissona. Później tylko skracamy. \square

Lemat 6. Zachodzi

$$n! \leq e\sqrt{n} \left(\frac{n}{e}\right)^n.$$

Dowód. Korzystamy z

$$\ln(n!) = \sum_{i=1}^n \ln i.$$

Mamy

$$\sum_{i=1}^n \ln i - \frac{1}{2} (\ln 1 + \ln n) = \sum_{i=1}^{n-1} \frac{1}{2} (\ln i + \ln (i+1)) (i+1-i) \leq \int_1^n \ln x \, dx,$$

gdzie nierówność wynika z tego, że druga suma jest przybliżeniem pola pod $\ln x$ za pomocą trapezów, a logarytm jest funkcją wklęsłą (bo pochodna jest ujemna).

Dla pełności policzymy tę całkę:

$$\int_1^n \ln x \, dx = x \ln x - \int_1^n x \cdot \frac{1}{x} \, dx = n \ln n - n + 1.$$

Przypominając, że $\ln 1 = 0^a$, dostajemy z poprzednich:

$$n \ln n - n + 1 \geq \ln(n!) - \frac{\ln n}{2},$$

czyli

$$n! \leq e^{n \ln n - n + 1 - \frac{\ln n}{2}},$$

a więc dokładnie to co chcieliśmy. □

^aaa

Twierdzenie 44. Niech $f(x_1, \dots, x_u)$ będzie dowolną nieujemną funkcją. Mamy ciąg niezależnych zmiennych o rozkładzie Poissona $(Y_1^{(k)}, \dots, Y_u^{(k)})$ z parametrem $\frac{k}{u}$ oraz ciąg zmiennych modelujących wrzucanie k kul do u urn $(X_1^{(k)}, \dots, X_u^{(k)})$. Zachodzi dla nich nierówność

$$E[f(X_1^{(k)}, \dots, X_u^{(k)})] \leq e\sqrt{k} E[f(Y_1^{(k)}, \dots, Y_u^{(k)})].$$

Funkcja f może być na przykład indykatorem zdarzenia które nas interesuje. Model jest skomplikowany, ale ograniczenie wynikające z rozkładu Poissona też może być ciekawe.

Dowód.

$$\begin{aligned} E[f(Y_1^{(k)}, \dots, Y_u^{(k)})] &= \sum_{j=0}^{\infty} E\left[f(Y_1^{(k)}, \dots, Y_u^{(k)}) \mid \sum_{i=1}^u Y_i^{(k)} = j\right] \cdot P\left(\sum_{i=1}^u Y_i^{(k)} = j\right) \\ &\geq E\left[f(Y_1^{(k)}, \dots, Y_u^{(k)}) \mid \sum_{i=1}^u Y_i^{(k)} = k\right] \cdot P\left(\sum_{i=1}^u Y_i^{(k)} = k\right) \\ &= E[f(X_1^{(k)}, \dots, X_u^{(k)})] \cdot P\left(\sum_{i=1}^u Y_i^{(k)} = k\right) = E[f(X_1^{(k)}, \dots, X_u^{(k)})] \cdot e^{-k} k^k \cdot \frac{1}{k!} \\ &\geq E[f(X_1^{(k)}, \dots, X_u^{(k)})] \cdot \frac{1}{e\sqrt{k}}. \end{aligned}$$

Pierwsza nierówność to ograniczenie się do $j = k$, a druga to ograniczenie górne na $k!$. □

Twierdzenie 45. Dla wystarczająco dużego n , gdy rzucamy n kul do n urn (w sposób niezależny i jednostajny), to prawdopodobieństwo, że maksymalne wypełnienie urny jest mniejsze od $\frac{\ln n}{\ln \ln n}$ jest co najwyżej $\frac{1}{n}$.

Dowód. Dokonamy aproksymacji przez zmienne Poissona. Każda urna dostaje zmienną losową Y_1, \dots, Y_n – niezależne zmienne Poissona z parametrem $\frac{n}{n} = 1$. Niech $M = \frac{\ln n}{\ln \ln n}$. Dla każdego i jest

$$P(Y_i \geq M) \geq \frac{e^{-1} 1^M}{M!},$$

bo takie jest prawdopodobieństwo równości. Zatem

$$P\left(\bigcap_{j \in [n]} Y_j < M\right) < \left(1 - \frac{e^{-1} 1^M}{M!}\right)^n = \left(1 - \frac{1}{eM!}\right)^n \leq e^{-\frac{n}{eM!}} < n^{-2},$$

gdzie przedostatnie przejście to skorzystanie z $1 + x \leq e^x$, a ostatnie to dość długie przekształcenia:

Chcemy $\frac{-n}{eM!} < -2 \ln n$, czyli $\frac{n}{2e \ln n} > M!$. Logarytmując dostajemy, że jest to równoważne

$$\ln n - \ln \ln n - \ln(2e) > \ln M!.$$

$M! \leq e\sqrt{M} \left(\frac{M}{e}\right)^M \leq M \left(\frac{M}{e}\right)^M$, gdzie w ostatnim przejściu zastąpiliśmy e przez \sqrt{M} (możemy, bo dla dużego n mamy duże M). Zatem $\ln M! \leq \ln M + M \ln M - M \ln e$. Podstawiając definicję M :

$$\ln M! \leq \ln \ln n - \ln \ln \ln n + \frac{\ln n}{\ln \ln n} \cdot (\ln \ln n - \ln \ln \ln n) - \frac{\ln n}{\ln \ln n}.$$

Teraz korzystamy z $\ln \ln n = o\left(\frac{\ln n}{\ln \ln n}\right)$. Dla dużych n wartość $\ln \ln n$ jest mniejsza niż $\frac{\ln n}{\ln \ln n} \cdot \ln \ln \ln n$, więc je ze sobą skracamy. Do tego pozbywamy się pierwszego wyrazu z minusem. Dostajemy

$$\ln M! < \ln n - \frac{\ln n}{\ln \ln n} - \ln(2e),$$

gdzie ostatni wyraz możemy dopisać, bo nierówność jest bardzo ostra (skracaliśmy ze sobą funkcje różne asymptotycznie). Mamy więc to, co chcieliśmy. Kończymy stosując aproksymację Poissona:

$$P\left(\bigcap_{j \in [n]} X_j < M\right) \leq e\sqrt{n}P\left(\bigcap_{j \in [n]} Y_j < M\right) \leq \frac{e}{n\sqrt{n}} < \frac{1}{n},$$

gdzie przez X_i rozumiemy liczbę kul w i -tej urnie, a ostatnie przejście to zastosowanie $n \gg e$. \square

Twierdzenie 46. Niech X oznacza zmienną losową z problemu kolekcjonera kuponów – liczbę kupionych paczek przed zebraniem wszystkich n kuponów. Wiemy, że jest $E[X] \approx n \ln n$. Zachodzi

$$\forall_c \lim_{n \rightarrow \infty} P(X > n \ln n + cn) = 1 - e^{-e^{-c}}.$$

Dowód. Będziemy modelować sytuację przez wrzucanie kul do urn, gdzie kule to paczki, a urny to kupony. Interesuje nas ograniczenie zdarzenia, że jest jakaś pusta urna przy $m = n \ln n + nc$ rzutach.

Rozważamy model Poissona Y_1, \dots, Y_n z parametrem $\frac{m}{n} = \ln n + c$. Dla każdego i mamy

$$P(Y_i = 0) = e^{-(\ln n + c)} \cdot (\ln n + c)^0 \cdot \frac{1}{0!} = \frac{e^{-c}}{n}.$$

Niech E będzie zdarzeniem $\bigcap_{j \in [n]} Y_j > 0$, czyli, że każda urna jest niepusta (w modelu Poissona). Mamy

$$P(E) = \left(1 - \frac{e^{-c}}{n}\right)^n \rightarrow e^{-e^{-c}}.$$

Z twierdzenia o prawdopodobieństwie całkowitym mamy

$$P(E) = P\left(E \mid \|Y - m\| \leq \sqrt{2m \ln m}\right) P\left(\|Y - m\| \leq \sqrt{2m \ln m}\right) + P\left(E \mid \|Y - m\| > \sqrt{2m \ln m}\right) P\left(\|Y - m\| > \sqrt{2m \ln m}\right),$$

gdzie $Y = \sum_{j \in [n]} Y_j$. Udowodnimy teraz dwa fakty:

$$1. P\left(\|Y - m\| \geq \sqrt{2m \ln m}\right) \rightarrow 0$$

$$2. \left\| P(E \mid \|Y - m\| \leq \sqrt{2m \ln m}) - P(E \mid Y = m) \right\| = o(1).$$

Z tego będzie wynikać nasza teza, bo we wzorze z prawdopodobieństwa całkowitego drugi składnik znika, a drugi czynnik w pierwszym dąży do 1 (oba na mocy punktu 1.), czyli na mocy punktu 2. jest

$$e^{-e^{-c}} = \lim_{n \rightarrow \infty} P(E) = \lim_{n \rightarrow \infty} P(E \mid Y = m),$$

a to jest równe granicy w prawdziwym modelu, bo dla ustalonej sumy zmienne Poissona zachowują się jak kule w urnach.

Punkt pierwszy pokazujemy za pomocą nierówności Czebyszewa:

$$P(\|Y - m\| \geq \sqrt{2m \ln m}) \leq \frac{m}{2m \ln m} \rightarrow 0,$$

bo Y to zmienna Poissona z parametrem m , a więc taką właśnie wartością oczekiwaną i wariancją.

Punktu drugiego dowodzimy zauważając, że $P(E \mid Y = k)$ jest rosnące względem k (im więcej kul, tym większa szansa, że każda urna zostanie zapełniona – tu też korzystamy z tego, że dla ustalonego k model Poissona jest równy prawdziwemu). Z tego wynika

$$P(E \mid Y = m - \sqrt{2m \ln m}) \leq P(E \mid \|Y - m\| \leq \sqrt{2m \ln m}) \leq P(E \mid Y = m + \sqrt{2m \ln m}),$$

bo są to odpowiednio najmniejsza i największa wartość Y , która spełnia zadaną nierówność. Zatem

$$\begin{aligned} & \left\| P(E \mid \|Y - m\| \leq \sqrt{2m \ln m}) - P(E \mid Y = m) \right\| \\ & \leq P(E \mid Y = m + \sqrt{2m \ln m}) - P(E \mid Y = m - \sqrt{2m \ln m}), \end{aligned}$$

bo zdarzenia których prawdopodobieństwa odejmujemy leżą bliżej siebie niż te, przez które szacujemy. Prawa strona jest równoważna zdarzeniu: była jakaś pusta urna po $m - \sqrt{2m \ln m}$ rzutach, a po kolejnych $2\sqrt{2m \ln m}$ już nie. Prawdopodobieństwo, że pusta urna zostanie zapełniona w ciągu tylu rzutów jest ograniczone z union bounda przez $\frac{2\sqrt{2m \ln m}}{n} = o(1)$, czyli cała różnica też. To ostatecznie kończy dowód. \square

Zajęcia 13: Zestaw 5

2024-11-12

Twierdzenie 47 (Negatywny rozkład hipergeometryczny). Tasujemy talię rozróżnialnych kart. W talii mamy B białych kart i C czarnych kart. Niech r będzie ustaloną liczbą taką, że $1 \leq r \leq C$. Wykładamy karty dopóki nie odkryjemy r czarnych kart. Niech X będzie liczbą odkrytych białych kart. Zachodzi

$$P(X = k) = \frac{\binom{k+r-1}{k} \binom{B+C-k-r}{B-k}}{\binom{B+C}{B}},$$

a do tego

$$\begin{aligned} E[X] &= \frac{Br}{C+1}, \\ \text{Var} &\left(\frac{rB(B+C+1)(C-r+1)}{(C+1)^2(C+2)} \right). \end{aligned}$$

Dowód. Rozważmy wszystkie możliwe wyłożenia kart (bez zatrzymywania się po r czarnych). Prawdopodobieństwo wynika z tego, że aby mieć k białych kart przed r czarnymi musimy najpierw wybrać k białych kart na pierwsze $k+r-1$ pozycji, na $(k+r)$ -tej jest czarna, a potem dopełniamy do pełnego wyłożenia, czyli wybieramy na ilu z pozostałych $B+C-k-r$ pozycji są białe karty (których zostało $B-k$). Całość dzielimy przez wszystkie możliwe wybory.

Niech X_i będzie indykátorem tego, czy i -ta biała karta jest wyłożona. Mamy

$$E[X_i] = \frac{r}{C+1}.$$

Rozważmy ułożenie wszystkich kart w ciąg. Najpierw ustawiamy wszystkie czarne. i -ta biała karta zostanie wyłożona, jeśli leży przed r -tą czarną, więc jest r możliwości na dobre wsadzenie jej przy $C + 1$ wszystkich.

W analogiczny sposób (drugą kartę wsadzamy już w ciąg $C + 2$ kart) dostajemy dla $i \neq j$

$$E[X_i X_j] = \frac{r(r+1)}{(C+1)(C+2)},$$

co pozwala nam policzyć wariancję. □

Definicja 34. Zmienna losowa X przyjmująca wartości w \mathbb{N}_1 jest bez pamięci, jeśli dla każdego $n, m \in \mathbb{N}$ mamy

$$P(X > n) > 0 \text{ oraz } P(X = m) = P(X = m + n \mid X > n).$$

Twierdzenie 48. Jeśli dyskretna zmienna losowa jest bez pamięci, to ma rozkład geometryczny.

Dowód. Oznaczmy $P(X = 1) = p$. Mamy $P(X > 1) = P(X \neq 1) = 1 - p$. Pokażemy indukcyjnie, że $P(X = k) = (1 - p)^{k-1} p$. Mamy

$$P(X = k) = P(X = k + 1 \mid X > 1) = \frac{P(X = k + 1 \cap X > 1)}{P(X > 1)} = \frac{P(X = k + 1)}{P(X > 1)},$$

gdzie ostatnie przejście wynika z tego, że jedno zdarzenie zawiera się w drugim. Mamy zatem

$$P(X = k + 1) = P(X = k) P(X > 1) = (1 - p)^k p,$$

gdzie skorzystaliśmy z założenia indukcyjnego. □

Twierdzenie 49. Niech X_1, \dots, X_n będą niezależnymi zmiennymi losowymi o rozkładzie geometrycznym z parametrem p . Niech $X = \sum_{i=1}^n X_i$ oraz $\delta > 0$. Zachodzi

$$P(X \geq (1 + \delta) 2n) \leq \left(\frac{(1 + \delta) 2p}{e^{(1 + \delta) 2p - 1}} \right)^n.$$

Dowód. Zauważmy, że zdarzenie $X \geq (1 + \delta) 2n$ ma mniejsze prawdopodobieństwo od zdarzenia: mając zmienną Y będącą sumą $(1 + \delta) 2n$ zmiennych Bernoulliego (indykatorowych) z parametrem p , zachodzi $Y \leq n$. Wynika to z tego, że pierwsze $(1 + \delta) 2n$ prób w naszych kolejnych zmiennych geometrycznych możemy modelować indykatorami. Jeśli n -ty sukces pojawi się po $(1 + \delta) 2n$ próbach, to w zmiennych indykatorowych będzie co najwyżej n jedynek. Mamy zatem

$$\begin{aligned} P(X \geq (1 + \delta) 2n) &\leq P(Y \leq n) = P\left(Y \leq \left(1 - \left(1 - \frac{1}{(1 + \delta) 2p}\right)\right) (1 + \delta) 2np\right) \\ &= \left(\frac{e^{-(1 - \frac{1}{(1 + \delta) 2p})}}{\left(\frac{1}{(1 + \delta) 2p}\right)^{\frac{1}{(1 + \delta) 2p}}} \right)^{(1 + \delta) 2np} = \left(\frac{(1 + \delta) 2p}{e^{(1 + \delta) 2p - 1}} \right)^n, \end{aligned}$$

gdzie skorzystaliśmy z ograniczenia Czernowa dla niezależnych indykatorów. □

Twierdzenie 50. Niech X będzie sumą oczek, które wypadły w ciągu n rzutów sześcienną kostką. Zachodzi

$$P\left(X - \frac{7}{2}n \geq 3\sqrt{2n \ln n}\right) \leq \frac{6}{n}.$$

Dowód. Niech Z_j będzie zmienną oznaczającą, ile razy wypadło j . Mamy $X = \sum_{j=1}^6 jZ_j$. Zachodzi

$$\begin{aligned} P\left(\sum_{j=1}^6 jZ_j - \frac{7}{2}n \geq 3\sqrt{2n \ln n}\right) &\leq P\left(\bigcup_{j \in [6]} jZ_j - \frac{j}{6}n \geq \frac{1}{2}\sqrt{2n \ln n}\right) \\ &\leq \sum_{j=1}^6 P\left(Z_j - \frac{1}{6}n \geq \frac{1}{2j}\sqrt{2n \ln n}\right) \leq 6P\left(Z_1 - \frac{1}{6}n \geq \frac{1}{2}\sqrt{2n \ln n}\right) \\ &= 6P\left(Z_1 - \frac{1}{6}n \geq \frac{1}{6}n \cdot 3\sqrt{\frac{2 \ln n}{n}}\right) \leq 6e^{-\left(3\sqrt{\frac{2 \ln n}{n}}\right)^2 \cdot \frac{1}{6}n \cdot \frac{1}{3}} = \frac{6}{n}, \end{aligned}$$

gdzie pierwsza nierówność wynika z tego, że gdyby każda zmienna Z_j odchyliła się od wartości oczekiwanej o mniej, to X też odchyliłby się o mniej. Potem korzystamy z union bounda i szacujemy przez $j = 1$. Na koniec korzystamy z nierówności Czernowa dla zmiennych indykatorowych. \square

Zajęcia 14: Łańcuchy Markowa

2024-11-15

Definicja 35. Proces stochastyczny to rodzina zmiennych losowych $\{X_t : t \in T\}$ indeksowana zbiorem T (zazwyczaj interpretowanym jako czas). X_t nazywamy stanem procesu w chwili t . Dla co najwyżej przeliczalnego T (zwykle $T = \mathbb{N}$) mówimy, że proces jest z czasem dyskretnym.

Definicja 36. Proces stochastyczny z czasem dyskretnym (i $T = \mathbb{N}$) jest łańcuchem Markowa, jeśli dla każdego $t \in \mathbb{N}$ i ciągu $(a_0, a_1, \dots, a_{t-1}, x)$ takiego, że

$$P\left(\bigcap_{i=0}^{t-1} X_i = a_i \wedge X_t = x\right) > 0$$

zachodzi

$$P\left(X_{t+1} = y \mid \bigcap_{i=0}^{t-1} X_i = a_i \wedge X_t = x\right) = P(X_{t+1} = y \mid X_t = x)$$

dla każdego y . Jest to automat, który zmienia stany z danym prawdopodobieństwem, ale zależnie tylko od poprzedniego stanu. Zbiór stanów będziemy oznaczać przez S .

Definicja 37. Macierz przejścia łańcucha Markowa o liczbie stanów s to macierz $s \times s$

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

gdzie $p_{ij} = P(X_t = j \mid X_{t-1} = i)$. Oznaczamy $p_{ij}(n) = P(X_n = j \mid X_0 = i) = P_{ij}^n$.

Przykład. Żaba skacze między dwoma liśćmi 0, 1. Z prawdopodobieństwem p skacze $0 \rightarrow 1$, a z $1 - p$ zostaje w 0. Analogicznie definiujemy q . Jest to łańcuch Markowa o macierzy przejścia

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$

Wraz z zadaniem stanem początkowym determinuje ona łańcuch.

Przykład (Randomizowany 2-SAT). Rozważmy formułę postaci $C = \bigwedge_i x_i \vee y_i$, gdzie x, y są jednymi z n zmiennych (lub ich zaprzeczeniami). Zastosujemy następujący algorytm: wybieramy losowe wartościowanie, dopóki formuła jest niespełniona wybieramy losową niespełnioną klauzulę i jeden z jej literalów, a następnie odwracamy jego wartościowanie. Powtarzamy taką operację maksymalnie $m \cdot 2n^2$, gdzie m jest pewnym parametrem. Jeśli formuła jest spełniona zwracamy wartościowanie,

inaczej stwierdzamy niespełnialność.

Taki algorytm zwraca poprawną odpowiedź z prawdopodobieństwem co najmniej $1 - \frac{1}{2^m}$.

Dowód. Algorytm może się pomylić tylko, jeśli formuła jest spełnialna, a on stwierdzi niespełnialność. Niech więc będzie spełnialna, a S będzie pewnym spełniającym ją wartościowaniem. Definiujemy X_i jako liczbę zmiennych wartościowanych tak samo jak w S w i -tym kroku algorytmu.

Zachodzi $P(X_{i+1} = 1 \mid X_i = 0) = 1$, a dla $j > 0$ jest

$$P(X_{i+1} = j + 1 \mid X_i = j) \geq \frac{1}{2},$$

$$P(X_{i+1} = j - 1 \mid X_i = j) \leq \frac{1}{2},$$

bo w poprawianej klauzuli albo obie zmienne się nie zgadzają i na pewno jedną zmienimy poprawnie, albo jedna się zgadza i mamy $\frac{1}{2}$ na zmianę poprawnej. Taki proces stochastyczny nie jest łańcuchem Markowa. Zamiast niego analizujemy proces o stanach Y_i takich, że $P(Y_{i+1} = j + 1 \mid Y_i = j) = \frac{1}{2}$ oraz $Y_0 = X_0$. To już jest łańcuch Markowa.

Niech h_j oznacza oczekiwaną liczbę kroków potrzebnych do osiągnięcia wartości n zaczynając od j . Mamy $h_n = 0$, $h_0 - h_1 + 1$ oraz $h_j = \frac{h_{j+1}}{2} + \frac{h_{j-1}}{2} + 1$. Można pokazać indukcyjnie, że $h_{j+1} = h_j - 2j - 1$, z czego wynika, że $h_0 = n^2$.

Podzielmy wykonanie algorytmu na segmenty rozmiaru $2n^2$. Niech Z będzie liczbą kroków potrzebną do otrzymania rozwiązania (licząc od początku segmentu). Mamy $E[Z] \leq n^2$. Zachodzi

$$P(Z > 2n^2) \leq \frac{n^2}{2n^2} = \frac{1}{2}.$$

Zatem prawdopodobieństwo, że algorytm nie zwróci poprawnego wartościowania po m segmentach jest ograniczone przez $\frac{1}{2^m}$.

Definicja 38. Stan i jest pochłaniający, jeśli $p_{ii} = 1$. Jest to stan, z którego już się nie wyjdzie.

Definicja 39. Stan j jest osiągalny ze stanu i , jeśli istnieje taka liczba kroków $n \in \mathbb{N}$ (może być 0), że $p_{ij}(n) > 0$. To znaczy, że da się osiągnąć ten stan w skończonej liczbie kroków.

Notacja. Jeśli j jest osiągalny z i , to piszemy $i \rightarrow j$.

Definicja 40. Stany i, j się wzajemnie komunikują, jeśli $i \rightarrow j$ oraz $j \rightarrow i$ (co oznaczamy $i \leftrightarrow j$).

Propozycja 13. Relacja wzajemnego skomunikowania jest relacją równoważności.

Definicja 41. Łańcuch Markowa jest nieprzywiedlny (nierozkładalny, irreducible), jeśli wszystkie stany się wzajemnie komunikują.

Definicja 42. Stan i jest nieistotny, jeśli istnieje j taki, że $i \rightarrow j$ oraz $j \not\rightarrow i$. Stan jest istotny, jeśli nie jest nieistotny.

Propozycja 14. Jeśli stan i jest nieistotny, to wszystkie stany w jego klasie skomunikowania są nieistotne. Podobnie jeśli jest istotny, to wszystkie są istotne.

Dowód. Ze stanów w jednej klasie da się dojść do dokładnie tych samych stanów. \square

Propozycja 15. Skończony łańcuch Markowa ma co najmniej jeden stan istotny.

Dowód. Klasy skomunikowania można modelować skierowanym, acyklicznym grafem. W takim grafie jest wierzchołek o stopniu wychodzącym 0. Elementy tej klasy są istotne. \square

Definicja 43. Dla nieprzywiedlnego łańcucha Markowa i stanu i definiujemy zbiór

$$\mathcal{T}(i) = \{t \geq 1 : p_{ii}(t) > 0\}.$$

Okres stanu i definiujemy jako $\gcd(\mathcal{T}(i))$. W szczególności zauważmy, że okres nie zawsze jest w zbiorze $\mathcal{T}(i)$.

Propozycja 16. W nieprzywiedlnym łańcuchu Markowa wszystkie stany mają ten sam okres.

Dowód. Rozważmy dwa stany $i, j \in S$. Z nieprzywiedlności mamy $i \leftrightarrow j$, a więc istnieją takie m i ℓ , że $p_{ij}(m) > 0$ oraz $p_{ji}(\ell) > 0$. Rozważmy dowolne $n \in \mathcal{T}(j)$ i oznaczmy przez $o(i)$ okres i . Mamy

$$p_{ii}(m+n+\ell) \geq p_{ij}(m) \cdot p_{jj}(n) \cdot p_{ji}(\ell) > 0,$$

bo chcąc przejść z i do i w $m+n+\ell$ krokach można przejść z i do j w m krokach, potem wykonać n kroków, po których dalej będzie się w j , a na koniec wrócić do i w ℓ krokach.

Podobnie $p_{ii}(m+\ell) \geq p_{ij}(m)p_{ji}(\ell) > 0$. Zatem $m+\ell, m+\ell+n \in \mathcal{T}(i)$, więc $o(i) \mid m+\ell$ i $o(i) \mid m+\ell+n$, czyli $o(i) \mid n$. Z tego mamy $o(i) \leq o(j)$, bo $o(i)$ też okazał się być wspólnym dzielnikiem elementów $\mathcal{T}(j)$. Nierówności w drugą stronę dowodzimy symetrycznie. \square

Definicja 44. Okres nieprzywiedlnego łańcucha Markowa definiujemy jako okres dowolnego jego stanu. Łańcuch jest nieokresowy, jeśli jego okres wynosi 1, jeśli jest większy, to łańcuch jest okresowy.

Lemat 7 (Lemat Schura). Zachodzi

$$\forall_{X \subseteq \mathbb{N}_1} \exists_{m_0} \forall_{m \geq m_0} \exists_{\substack{r \geq 1, \\ x_1, \dots, x_r \in X, \\ \ell_1, \dots, \ell_r \in \mathbb{Z}}} m = \sum_{i \in [r]} \ell_i x_i.$$

Propozycja 17. Jeśli $\{X_n : n \in \mathbb{N}\}$ jest nieprzywiedlnym nieokresowym łańcuchem Markowa, to

$$\forall_{j, j' \in S} \exists_{n_0} \forall_{n \geq n_0} p_{jj'}(n) > 0.$$

Znaczy to, że im dalej jesteśmy w łańcuchu, tym bardziej możemy znaleźć się w dowolnym stanie.

Dowód. Niech $j, j' \in S$. Z nieokresowości mamy $\gcd\{n : p_{jj}(n) > 0\} = 1$. Na mocy lematu Schura istnieje takie m_0 , że

$$\forall_{m \geq m_0} \exists_{\substack{r \geq 1, \\ n_1, \dots, n_r \in \mathcal{T}(j), \\ \ell_1, \dots, \ell_r \in \mathbb{Z}_+}} m = \sum_{i \in [r]} \ell_i n_i.$$

Teraz $p_{jj}(m) \geq \prod_{i \in [r]} p_{jj}(\ell_i n_i) \geq \prod_{i \in [r]} p_{jj}(n_i)^{\ell_i} > 0$.

Z nieprzywiedlności $j' \rightarrow j$, a więc $\exists_{m_1} p_{j'j}(m_1) > 0$.

Kładziemy $n_0 = m_0 + m_1$. Dla dowolnego $n \geq n_0$ mamy

$$p_{j'j}(n) \geq p_{j'j}(m_1) \cdot p_{jj}(n - m_1) > 0,$$

bo $n - m_1 \geq m_0$. \square

Definicja 45. Definiujemy zmienne losowe

$$\tau_j = \min\{n \in \mathbb{N} : X_n = j\},$$

$$\tau_j^+ = \min\{n \in \mathbb{N}_1 : X_n = j\}.$$

Oznaczają one pierwsze odwiedzenie stanu j przez instancję łańcucha Markowa (opcjonalnie nie licząc zaczątku w nim). Przypomnijmy, że minimum ze zbioru pustego to ∞ . Definiujemy odpowiednie

prawdopodobieństwa

$$f_{ij}(n) = P(\tau_j^+ = n \mid X_0 = i) = P(X_n = j \cap X_{n-1} \neq j \cap \dots \cap X_1 \neq j \mid X_0 = i),$$

$$f_{ij} = P(\tau_j^+ < \infty \mid X_0 = i) = \sum_{n=1}^{\infty} f_{ij}(n).$$

Definicja 46. Stan i jest powracający, jeśli $f_{ii} = 1$, a chwilowy, jeśli $f_{ii} < 1$. Powracający stan to taki, do którego na pewno wrócimy.

Uwaga. Stan powracający na pewno jest istotny (bo nieistotne to takie, z których nie da się wrócić).

Definicja 47. Stan powracający i jest dodatni, jeśli $E[\tau_i^+] = \sum_{n=1}^{\infty} n f_{ii}(n) < \infty$. W przeciwnym wypadku jest zerowy.

Przykład. Dla każdego stanu i mamy prawdopodobieństwo $\frac{i}{i+1}$ na pójście do $i+1$ oraz $\frac{1}{i+1}$ na pójście do 1. Zaczynając z 1 prawdopodobieństwo niewrócenia w pierwszych t krokach to

$$\prod_{j=1}^t \frac{j}{j+1} = \frac{1}{t+1}.$$

Zatem $f_{ii} = \lim_{t \rightarrow \infty} \left(1 - \frac{1}{t+1}\right) = 1$, a więc stan 1 jest powracający.

Liczymy wartość oczekiwaną:

$$E[\tau_1^+] = \sum_{t=1}^{\infty} t \cdot \frac{1}{t} \frac{1}{t+1} = \sum_{t=1}^{\infty} \frac{1}{t+1} = \infty,$$

bo najpierw musimy $t-1$ razy nie wrócić do 1, a potem wybrać powrót. Zatem stan 1 jest zerowy.

Definicja 48. Rozkład π łańcucha Markowa w danym momencie to rozkład zmiennej losowej odpowiadającej następnemu stanowi warunkowanej aktualnym stanem. Mamy $\pi = (\pi_i)_{i \in S}$ dla π_i będącego prawdopodobieństwem przejścia do stanu i .

Uwaga. Rozkład π_0 zmiennej X_0 można reprezentować poziomym wektorem. Mnożąc go przez macierz przejścia dostajemy kolejne rozkłady: $\pi_t = \pi_0 \cdot P^t$.

Definicja 49. Rozkład łańcucha Markowa π o macierzy przejścia P jest stacjonarny, jeśli $\pi = \pi P$.

Twierdzenie 51. Każdy skończony nieprzywiedlny nieokresowy łańcuch Markowa posiada (jeden) rozkład stacjonarny $\pi = (\pi_i)_{i \in S}$. Dla każdego $i, j \in S$ granica $\lim_{t \rightarrow \infty} p_{ij}(t)$ istnieje, a jej wartość jest niezależna od i . Do tego zachodzi

$$\pi_i = \lim_{t \rightarrow \infty} p_{ii}(t) = \frac{1}{E[\tau_i^+]}$$

Dowód. Najpierw pokażemy jedyność. Niech ϕ, π będą dwoma rozkładami stacjonarnymi. Rozważmy $x \in S$ taki, że $\frac{\pi(x)}{\phi(x)}$ jest minimalne. Zachodzi

$$\pi(x) = \sum_{y \in S} \pi(y) \cdot P_{y,x} = \sum_{y \in S} \frac{\pi(y)}{\phi(y)} \phi(y) P_{y,x} \geq \frac{\pi(x)}{\phi(x)} \sum_{y \in S} \phi(y) P(y, x) = \frac{\pi(x)}{\phi(x)} \phi(x) = \pi(x),$$

gdzie pierwsze i przedostatnie przejście to skorzystanie ze stacjonarności, a nierówność to założona minimalność. Wykazaliśmy, że tak naprawdę jest to równość, a więc π jest jakąś krotnością ϕ . Oba te wektory mają taką samą normę, a więc są po prostu równe. \square

Twierdzenie 52. Niech S będzie (dowolnym) zbiorem stanów w skończonym nieprzywiedlnym nieokresowym łańcuchu Markowa. W rozkładzie stacjonarnym π prawdopodobieństwo opuszczenia zbioru S jest równe prawdopodobieństwu wejścia do niego.

Dowód. Rozważmy na początek $S = \{i\}$. Mamy

$$\sum_{j=1}^n \pi_j P_{ji} = \pi_i = \pi_i \sum_{j=1}^n P_{ij},$$

a więc można skasować wyraz $\pi_i P_{ii}$, czyli mamy

$$\sum_{j \neq i} \pi_j P_{ji} = \sum_{j \neq i} \pi_i P_{ij}.$$

Sumując takie równości dla różnych i dostajemy odpowiednią równość dla dowolnego zbioru S . \square

Algorytm 2 (Znajdowanie rozkładu stacjonarnego). Mamy $\pi = \pi P$ oraz ustaloną normę π , a więc można po prostu rozwiązać układ równań. Jest to najbardziej brutalne rozwiązanie.

Przyjemniejsza metoda (trochę mniej równań) to skorzystanie z twierdzenia, że prawdopodobieństwo wyjścia ze zbioru jest równe prawdopodobieństwu wejścia do niego. Dostajemy wtedy równania powstałe dla wybranych zbiorów.

Trzecim sposobem jest liczenie wartości własnych.

Pomysł. Mając zadany graf G możemy zdefiniować łańcuch Markowa, w którym stany to wierzchołki, a z każdego wierzchołka można przejść do każdego jego sąsiada z równym prawdopodobieństwem. Taki łańcuch nazywamy (losowym) spacerem po grafie. Zauważmy, że dla spójnego grafu taki łańcuch jest nieprzywiedlny.

Lemat 8. Niech graf G będzie spójny. Łańcuch spaceru po nim jest nieokresowy wtedy i tylko wtedy, gdy graf nie jest dwudzielny.

Dowód. (\implies) Jeśli G jest dwudzielny, to do każdego wierzchołka v można wrócić tylko po parzystej liczbie kroków, bo co krok zmieniamy stronę, po której jesteśmy.

(\impliedby) W niedwudzielnym G musi istnieć nieparzysty cykl. Niech v leży na tym cyklu. Z jednej strony można wyjść do dowolnego sąsiada v i wrócić, co da $p_{vv}(2) > 0$, a z drugiej można przejść całym cyklem, czyli $p_{vv}(2k+1) > 0$. Zatem okres v (czyli całego spaceru) to 1. \square

Uwaga. Dalej będziemy rozważać grafy spójne i niedwudzielne, a więc takie, których spacery są nieprzywiedlne i nieokresowe. W szczególności mają one rozkład stacjonarny.

Twierdzenie 53. Rozkład stacjonarny $\pi = (\pi_v)_{v \in V(G)}$ spaceru po grafie G ma postać $\pi_v = \frac{d(v)}{2E(G)}$.

Dowód. Pokażemy, że tak zadane π faktycznie jest rozkładem stacjonarnym. Mamy

$$\sum_{v \in V(G)} \pi_v = \sum_{v \in V(G)} \frac{d(v)}{2E(G)} = \frac{1}{2E(G)} \sum_{v \in V(G)} d(v) = 1,$$

a więc faktycznie jest to rozkład. Mamy też

$$(\pi P)_v = \sum_{u \in V(G)} \pi_u \cdot P_{uv} = \sum_{u \in N(v)} \frac{d(u)}{2E(G)} \cdot \frac{1}{d(u)} = \frac{d(v)}{2E(G)} = \pi_v,$$

gdzie drugie przejście to zastosowanie określenia macierzy P dla spaceru. \square

Twierdzenie 54. Niech X będzie zmienną Poissona z parametrem $\mu \in \mathbb{N}_1$. Zachodzi $P(X \geq \mu) \geq \frac{1}{2}$.

Dowód. Pokażemy, że dla $h \in \{0, \dots, \mu - 1\}$ zachodzi $P(X = \mu + h) \geq P(X = \mu - h - 1)$. Chcemy pokazać

$$\frac{e^{-\mu} \mu^{\mu+h}}{(\mu+h)!} \geq \frac{e^{-\mu} \mu^{\mu-h-1}}{(\mu-h-1)!} \iff \mu^{2h+1} \geq (\mu+h) \dots (\mu-h),$$

ale grupując skrajne wyrazy dostajemy $(\mu+h) \dots (\mu-h) = \mu \prod_{i=1}^h (\mu^2 - i^2) \leq \mu^{2h+1}$.

Teraz wystarczy przeliczyć

$$P(X < \mu) = \sum_{h=0}^{\mu-1} P(X = \mu - h - 1) \leq \sum_{h=0}^{\mu-1} P(X = \mu + h) \leq P(X \geq \mu).$$

Zatem $1 = P(X < \mu) + P(X \geq \mu) \leq 2P(X \geq \mu)$, co daje to co chcemy. \square

Twierdzenie 55. Niech zmienna $X_i^{(m)}$ oznacza liczbę kul w i -tym kubku, przy założeniu, że rzucamy m kul do n kubków w sposób jednostajny i niezależny. Niech zmienne $Y_1^{(m)}, \dots, Y_n^{(m)}$ będą niezależnymi zmiennymi z rozkładu Poissona o parametrze $\frac{m}{n}$. Niech f będzie nieujemną, rosnącą funkcją. Zachodzi

$$E[f(X_1^{(m)}, \dots, X_n^{(m)})] \leq 2 \cdot E[f(Y_1^{(m)}, \dots, Y_n^{(m)})].$$

Dowód.

$$\begin{aligned} 2 \cdot E[f(Y_1^{(m)}, \dots, Y_n^{(m)})] &= 2 \cdot \sum_{k=0}^{\infty} E[f(Y_1^{(m)}, \dots, Y_n^{(m)}) \mid \sum_{i=1}^n Y_i^{(m)} = k] P\left(\sum_{i=1}^n Y_i^{(m)} = k\right) \geq \\ &2 \sum_{k=m}^{\infty} E[f(X_1^{(k)}, \dots, X_n^{(k)})] P\left(\sum_{i=1}^n Y_i^{(m)} = k\right) \geq E[f(X_1^{(m)}, \dots, X_n^{(m)})] \cdot 2P\left(\sum_{i=1}^n Y_i^{(m)} \geq m\right) \\ &\geq E[f(X_1^{(m)}, \dots, X_n^{(m)})], \end{aligned}$$

gdzie drugie przejście to skorzystanie z tego, że zmienne Poissona z ustaloną sumą mają taki rozkład jak prawdziwy model. W ostatnim przejściu korzystamy z tego, że dla zmiennej Poissona A z parametrem μ mamy $P(A \geq \mu) \geq \frac{1}{2}$. \square

Twierdzenie 56. Niech X będzie zmienną losową z rozkładem Poissona o parametrze μ . Niech $X = Y + Z$, gdzie wartości tych zmiennych definiujemy w następujący sposób: mając X obiektów każdy wybieramy z prawdopodobieństwem p do jednego zbioru lub z prawdopodobieństwem $1 - p$ do drugiego. Moc jednego zbioru to Y , a drugiego Z . Zmienne Y, Z mają rozkład Poissona z parametrami odpowiednio $p\mu$ i $(1-p)\mu$ oraz są niezależne.

Dowód.

$$\begin{aligned} P(Y = k) &= \sum_{m=0}^{\infty} P(Y = k \mid X = m) P(X = m) = \sum_{m=k}^{\infty} \binom{m}{k} p^k (1-p)^{m-k} \cdot \frac{e^{-\mu} \mu^m}{m!} = \\ &\frac{e^{-\mu p} \mu^k}{k!} p^k \sum_{m=k}^{\infty} \frac{e^{-\mu(1-p)} \mu^{m-k} k!}{m!} (1-p)^{m-k} \binom{m}{k} = \\ &\frac{e^{-\mu p} (\mu p)^k}{k!} \sum_{m=0}^{\infty} \frac{e^{-\mu(1-p)} (\mu(1-p))^m k!}{(m+k)!} \binom{m+k}{k} = \frac{e^{-\mu p} (\mu p)^k}{k!} \sum_{m=0}^{\infty} \frac{e^{-\mu(1-p)} (\mu(1-p))^m}{m!} = \\ &\frac{e^{-\mu p} (\mu p)^k}{k!} e^{-\mu(1-p)} e^{\mu(1-p)} = \frac{e^{-\mu p} (\mu p)^k}{k!}, \end{aligned}$$

zatem Y ma rozkład Poissona z odpowiednim parametrem. Podobne przekształcenia działają dla

Z. Niezależności dowodzimy z definicji:

$$P(Y = y \cap Z = z) = P(X = y + z) p^y (1 - p)^z \binom{y + z}{y} = \frac{e^{-\mu} \mu^{y+z}}{(y + z)!} p^y (1 - p)^z \frac{(y + z)!}{y! z!} =$$

$$\frac{e^{-\mu p} (\mu p)^y}{y!} \cdot \frac{e^{-\mu(1-p)} (\mu(1-p))^z}{z!} = P(Y = y) \cdot P(Z = z).$$

□

Twierdzenie 57. Dla niezależnych zmiennych Poissona X, Y o parametrze μ zachodzi

$$P(X > Y) = \frac{1}{2} - \frac{1}{2} e^{-2\mu} \sum_{t=0}^{\infty} \frac{\mu^{2t}}{(t!)^2}.$$

Dowód. Mamy $1 = P(X > Y) + P(X = Y) + P(X < Y)$, ale z symetrii $P(X > Y) = P(X < Y)$, więc mamy

$$P(X > Y) = \frac{1 - P(X = Y)}{2}.$$

Szukamy teraz $P(X = Y)$ z definicji:

$$P(X = Y) = \sum_{t=0}^{\infty} P(X = Y = t) = \sum_{t=0}^{\infty} \left(\frac{e^{-\mu} \mu^t}{t!} \right)^2 = e^{-2\mu} \sum_{t=0}^{\infty} \frac{\mu^{2t}}{(t!)^2}.$$

□