

METODY PROBABILISTYCZNE INFORMATYKI

NOTATKI

„Możemy cierpieć. Kocham cierpienie. Kto nie kocha cierpienia?”

MACIEJ MIKOŁAJCZAK
Kraków, 2024/2025

Spis treści

1. Przestrzeń probabilistyczna	2
2. Zmienna losowa	3
3. Własności wartości oczekiwanej	5
4. Miara i całka	8
5. Wariancja	11
6. Nierówności koncentrujące	14
7. Nierówności Czernowa	16
8. Rozkład Poissona	20
9. Łańcuchy Markowa	26
10. Spacery po grafach	32
11. Ciągłe zmienne losowe	34
12. Rozkład wykładniczy	37
13. Proces Poissona	41
14. Rozkład normalny	45
15. Sprzęganie łańcuchów Markowa	49
16. Metoda Monte Carlo	55

1. Przestrzeń probabilistyczna

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. $(\Omega, \mathcal{F}, P)$ 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:

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

Lemat 1. Dla przestrzeni probabilistycznej $(\Omega, \mathcal{F}, P)$ 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. $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$.

□

Propozycja 1 (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

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

□

2. Zmienna losowa

2024-10-04

Definicja 5. X jest zmienną losową na przestrzeni $(\Omega, \mathcal{F}, P)$, 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 2 (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 3 (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 $\binom{n}{2} \cdot \frac{1}{4}$.

3. Własności wartości oczekiwanej

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 2. 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 3. 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

$$\forall_{x_1, x_2, 0 \leq \lambda \leq 1} f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2).$$

Dla funkcji różniczkowalnej jest to równoważne dodatniości drugiej pochodnej.

Twierdzenie 4 (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 4. 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 5. 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 6. 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).

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

Propozycja 7. $E[Y] = E[E[Y | Z]]$.

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

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

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

□

Propozycja 9. Jeśli zmienna losowa X przyjmuje wartości nieujemne całkowite, to

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

Dowód.

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

□

4. 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 5 (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 10. 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. □

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 $f : \Omega \rightarrow \mathbb{R}$ 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 6 (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 7. 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

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

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 $P([a_1 \dots a_k]) = \frac{1}{2^k}$.

5. Wariancja

2024-10-24

Propozycja 11. 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 $P(Q \text{ wybrał za pivot } x_i \text{ lub } x_j \mid Q \text{ wybrał pivota spośród } \{x_i, \dots, x_j\}) = \frac{\frac{2}{j-i+1}}{\frac{Z}{j-i+1}} = \frac{2}{Z}$, 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 8 (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 indykátorem 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 2. 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 3. 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:

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

Twierdzenie 9 (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

$$\begin{aligned} E[X] &= \frac{nK}{N}, \\ \text{Var}(X) &= \frac{Kn}{N} + \frac{K(K-1)n(n-1)}{N(N-1)} - \frac{K^2n^2}{N^2}. \end{aligned}$$

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 indykátorem 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 $\text{Var}(X)$. $\square$

Twierdzenie 10 (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 odkrywamy 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

$$E[X] = \frac{Br}{C+1},$$

$$\text{Var} \left(\frac{rB(B+C+1)(C-r+1)}{(C+1)^2(C+2)} \right).$$

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. □

6. Nierówności koncentrujące

2024-10-25

Twierdzenie 11 (Nierówność Czebyszewa). Niech $a > 0$. Dla zmiennej losowej X mamy

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

Dowód.

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

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 12. 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 13. 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 $(M_X(t))^{(n)} = E[X^n e^{tX}]$. $\square$

Twierdzenie 14. 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 4. 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.

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

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

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

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

Twierdzenie 15 (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:

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

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

7. Nierówności Czernowa

2024-11-05

Twierdzenie 16. 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 17. 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 $\delta = \sqrt{6 \frac{\ln n}{n}}$ (dla odpowiednio dużego n należy do $(0, 1)$), ograniczenie $2 \cdot e^{-\frac{n}{2} \cdot 6 \frac{\ln n}{n} \cdot \frac{1}{3}} = \frac{2}{n}$.

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). Niech zmienna losowa $X = \hat{p}n$ oznacza liczbę wystąpień tej mutacji w naszej próbce. 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 18. Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi o rozkładzie prawdopodobieństwa $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. 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 odchyłeń multiplikatywnych, bo $E[X] = 0$.

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

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

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

Dowód. Bierzemy $Y_i = \frac{X_i+1}{2}$. Mamy $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Dla $X = \sum_{i=1}^n X_i = 2Y - 2\mu$ 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 $\{0, 1\}$.

$$\begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix}$$

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

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

Dowód. Niech i -ty wiersz $\bar{a}_i = a_{i1} \dots a_{im}$ ma w sobie k jedynek. $Z = \sum_{j=1}^m a_{ij} b_j$ jest sumą k zmiennych losowych, które z równym prawdopodobieństwem przyjmują 1 i -1 .

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

8. Rozkład Poissona

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?

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

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 20. Dla wystarczająco dużego n , gdy wrzucamy n kul do n urn, to prawdopodobieństwo, że maksymalne zapełnienie jest większe od $\frac{3 \ln n}{\ln \ln n}$ jest co najwyżej $\frac{1}{n}$.

Dowód. Niech $k > \frac{3 \ln n}{\ln \ln n}$. Prawdopodobieństwo, że konkretna urna ma co najmniej k kul jest ograniczone z góry przez

$$\binom{n}{k} \left(\frac{1}{n}\right)^k,$$

co wynika z union bounda – sumujemy się po wszystkich zbiorach k kul i mnożymy przez prawdopodobieństwo, że cały ten zbiór znalazł się w urnie. Dalej przekształcamy:

$$\begin{aligned} \binom{n}{k} \left(\frac{1}{n}\right)^k &\leq \frac{n^k}{k!} \left(\frac{1}{n}\right)^k = \frac{1}{k!} \leq \left(\frac{e}{k}\right)^k \leq \left(\frac{e \ln n}{3 \ln n}\right)^{\frac{3 \ln n}{\ln \ln n}} \leq \exp\left(\ln\left(\frac{\ln \ln n}{\ln n}\right) \cdot \frac{3 \ln n}{\ln \ln n}\right) \\ &= \exp\left(\frac{3 \ln n}{\ln \ln n} (\ln \ln \ln n - \ln \ln n)\right) = n^{-3 + \frac{3 \ln \ln \ln n}{\ln \ln n}} \leq n^{-2}. \end{aligned}$$

Najpierw 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 3 (bo nierówność) i przeliczamy, a potem wyrażenie w wykładniku zbiega do -3 , a więc dla dużych n można je ograniczyć przez -2 .

Kończymy za pomocą union bounda:

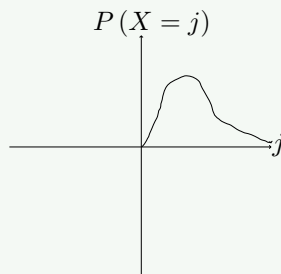
$$\begin{aligned} P\left(\text{maksymalne zapełnienie} > \frac{3 \ln n}{\ln \ln n}\right) &= P\left(\bigcup_{i \in [n]} \left\{i\text{-ta urna ma co najmniej } \frac{3 \ln n}{\ln \ln n} \text{ kul}\right\}\right) \\ &\leq \sum_{i \in [n]} P\left(i\text{-ta urna ma co najmniej } \frac{3 \ln n}{\ln \ln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

□

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 1: 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 21. 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 12. 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 13. 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 5. 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 22. 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:

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

□

Twierdzenie 23. Niech X będzie zmienną o rozkładzie Poissona z parametrem μ . Wtedy:

1. jeśli $x > \mu$, to $P(X \geq x) \leq \frac{e^{-\mu} (\mu)^x}{x^x}$
2. jeśli $x < \mu$, to $P(X \leq x) \leq \frac{e^{-\mu} (\mu)^x}{x^x}$
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 6. 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 7. 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 druga 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 $\ln 1 = 0$, 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. □

Twierdzenie 24. 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 indykátorem 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 25. 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 26. 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

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

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

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

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

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:

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

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. $\square$

9. Łańcuchy Markowa

2024-11-15

Definicja 34. 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 35. 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 36. 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 nie zgadzają się z S 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 $h_j = h_{j+1} + 2j + 1$: działa dla $j = 0$, a dalej mamy

$$h_{j+1} = 2h_j - h_{j-1} - 2 = 2h_j - (h_j + 2(j-1) + 1) - 1 = h_j - 2j - 1.$$

Z tego wynika, że $h_0 = h_1 + 1 = h_2 + 1 + 3 = \dots = \sum_{i=0}^{n-1} (2i+1) = n^2$.

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

Definicja 38. 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 39. Stany i, j się wzajemnie komunikują, jeśli $i \rightarrow j$ oraz $j \rightarrow i$ (co oznaczamy $i \leftrightarrow j$).

Propozycja 14. Relacja wzajemnego skomunikowania jest relacją równoważności.

Definicja 40. Łańcuch Markowa jest nieprzywiedlny (nierozkładalny, irreducible), jeśli wszystkie stany się wzajemnie komunikują.

Definicja 41. 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 15. 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 16. 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 42. 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 17. 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 43. 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 8 (Lemat Schura). Jeśli $S \subseteq \mathbb{N}$ oraz $\gcd S = g_S$, to istnieje $m_S \in \mathbb{N}$ takie, że dla każdego $m \geq m_S$ iloczyn mg_S da się przedstawić jako (skończoną) kombinację liniową elementów z S o nieujemnych współczynnikach.

Dowód. Niech d_S będzie najmniejszą nieujemną liczbą, którą da się przedstawić jako kombinację liniową elementów S . d_S musi dzielić każdą liczbę w S , bo inaczej reszta z dzielenia przez d_S jest mniejszą od d_S kombinacją liniową elementów S . d_S jest wspólnym dzielnikiem elementów S , a więc $d_S \mid g_S$. Oczywiście jest $g_S \mid d_S$, a więc $g_S = d_S$.

Zauważmy, że wartość $\gcd(S \cap [1, n])$ dla rosnącego n jest malejąca, ale może zmniejszyć się tylko skończenie wiele razy. Zatem dla pewnego n_0 zmniejszy się po raz ostatni i dla $F = S \cap [1, n_0]$ mamy

$\gcd(S) = \gcd(F)$. Zatem wystarczy udowodnić tezę dla skończonych F . Zrobimy to indukcyjnie, zaczynając od bazy $F = \{a, b\}$.

Niech $\gcd F = g$. Dla $m > 0$ zapiszmy $mg = ca + db$ dla pewnych c, d . Zauważmy, że mamy $mg = (c + kb)a + (d - ka)b$, a więc można założyć $0 \leq c < b$. Jeśli $mg > (b - 1)a - b$, to musi być $d \geq 0$, a więc można położyć $m_F = \frac{ab-a-b}{g} + 1$.

W kroku indukcyjnym rozważmy zbiór $F \cup \{a\}$. Niech $g = \gcd(F \cup \{a\}) = \gcd(g_F, a)$. Weźmy n takie, że $ng > m_{\{a, g_F\}}g + m_F g_F$. Mamy $ng - m_F g_F = ca + dg_F$ dla pewnych $c, d \geq 0$. Zatem $ng = ca + (d + m_F)g_F = ca + \sum_{f \in F} c_f f$ dla pewnych $c_f \geq 0$, co wynika z określenia m_F . Teraz wystarczy położyć $m_{F \cup \{a\}} = m_{\{a, g_F\}} + \frac{m_F g_F}{g}$. $\square$

Propozycja 18. 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

$$\forall m \geq m_0 \exists \begin{matrix} r \geq 1, \\ n_1, \dots, n_r \in \mathcal{T}(j), \\ \ell_1, \dots, \ell_r \in \mathbb{N} \end{matrix} m = \sum_{i \in [r]} \ell_i n_i.$$

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 jest $j' \rightarrow j$, 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 44. Definiujemy zmienną losową

$$\tau_{i,j} = \min\{n \in \mathbb{N}_1 : X_n = j \mid X_0 = i\},$$

oznaczającą pierwsze odwiedzenie stanu j przez instancję łańcucha Markowa, który zaczyna w i . Przypomnijmy, że minimum ze zbioru pustego to ∞ . Definiujemy odpowiednie prawdopodobieństwa

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

$$f_{i,j} = P(\tau_{i,j} < \infty) = \sum_{n=1}^{\infty} f_{i,j}(n).$$

Definicja 45. Stan i jest powracający, jeśli $f_{i,i} = 1$, a chwilowy, jeśli $f_{i,i} < 1$. Powracający stan to taki, do którego na pewno wrócimy.

Uwaga. Stan powracający na pewno jest istotny (bo nieistotne to takie, do których jest szansa, że nie wrócimy).

Definicja 46. Stan powracający i jest dodatni, jeśli $E[\tau_{i,i}] = \sum_{n=1}^{\infty} n f_{i,i}(n) < \infty$. W przeciwnym wypadku mówimy, że 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 $f_{i,i} = \lim_{t \rightarrow \infty} \left(1 - \frac{1}{t+1}\right) = 1$, a więc stan 1 jest powracający.

Liczmy wartość oczekiwaną:

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

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

Lemat 9. W skończonym, nieprzywiedlnym łańcuchu Markowa każdy stan jest powracający dodatni.

Dowód. Z nieprzywiedlności wiemy, że dla każdych stanów i, j istnieje takie n , że $p_{i,j}(n) > 0$. Biorąc maksimum po wartościach n dla różnych par stanów i minimum po wartościach prawdopodobieństw dostajemy, że istnieje liczba naturalna $r > 0$ i rzeczywista $\varepsilon > 0$ takie, że dla każdych dwóch stanów i, j istnieje $j \leq r$ takie, że $p_{i,j}(j) > \varepsilon$.

Z tego wynika, że dla ustalonego stanu y prawdopodobieństwo na wejście do y pomiędzy czasem t a $t+r$ wynosi co najmniej ε , niezależnie od stanu łańcucha w czasie t . Zatem

$$P(\tau_{y,y} > kr) \leq (1 - \varepsilon) P(\tau_{y,y} > (k-1)r) \leq \dots \leq (1 - \varepsilon)^k.$$

Teraz możemy przeliczyć

$$E[\tau_{y,y}] = \sum_{t \in \mathbb{N}} P(\tau_{y,y} > t) \leq \sum_{k \geq 0} r P(\tau_{y,y} > kr) \leq r \sum_{k \geq 0} (1 - \varepsilon)^k < \infty,$$

gdzie pierwsza nierówność to skorzystanie z faktu, że $P(\tau_{y,y} > t)$ jest nierosnące względem t . Wiadomo, że stan y jest dodatni, a więc w szczególności jest też powracający (skończona wartość oczekiwana daje prawdopodobieństwo 1, że zmienna ma skończoną wartość). $\square$

Definicja 47. Rozkład π łańcucha Markowa w danym momencie to rozkład zmiennej losowej odpowiadającej aktualnemu stanowi. 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 48. Rozkład łańcucha Markowa π o macierzy przejścia P jest stacjonarny, jeśli $\pi = \pi P$.

Lemat 10. Dla każdego skończonego, nieprzywiedlnego, nieokresowego łańcucha Markowa i jego stanu i zachodzi

$$\lim_{t \rightarrow \infty} p_{ii}(t) = \frac{1}{E[\tau_{i,i}]}.$$

Twierdzenie 27. Każdy skończony, nieprzywiedlny, nieokresowy łańcuch Markowa posiada (jedyny) rozkład stacjonarny $\pi = (\pi_i)_{i \in S}$. Dla każdego $i, j \in S$ granica $\lim_{t \rightarrow \infty} p_{ji}(t)$ istnieje, a jej wartość jest niezależna od j . Do tego zachodzi

$$\pi_i = \lim_{t \rightarrow \infty} p_{ji}(t) = \frac{1}{E[\tau_{i,i}]}.$$

Dowód. Ustalmy stany i i j . Mamy $\sum_{t=1}^{\infty} f_{j,i}(t) = 1$, bo w skończonym, nieprzywiedlnym łańcuchu każdy stan jest powracający. Dla każdego $\varepsilon > 0$ istnieje t_1 takie, że $\sum_{t=1}^{t_1} f_{j,i}(t) \geq 1 - \varepsilon$. Dla $t \geq t_1$

mamy

$$p_{j,i}(t) = \sum_{k=1}^t f_{j,i}(k) p_{i,i}(t-k) \geq \sum_{k=1}^{t_1} f_{j,i}(k) p_{i,i}(t-k).$$

Korzystamy z przejścia granicznego:

$$\lim_{t \rightarrow \infty} p_{j,i}(t) \geq \lim_{t \rightarrow \infty} \sum_{k=1}^{t_1} f_{j,i}(k) p_{i,i}(t-k) = \sum_{k=1}^{t_1} f_{j,i}(k) \lim_{t \rightarrow \infty} p_{i,i}(t) \geq (1-\varepsilon) \lim_{t \rightarrow \infty} p_{i,i}(t).$$

Teraz szacujemy z drugiej strony:

$$p_{j,i}(t) = \sum_{k=1}^t f_{j,i}(k) p_{i,i}(t-k) \leq \sum_{k=1}^{t_1} f_{j,i}(k) p_{i,i}(t-k) + \sum_{k=t_1+1}^t f_{j,i}(k) \leq \sum_{k=1}^{t_1} f_{j,i}(k) p_{i,i}(t-k) + \varepsilon,$$

$$\lim_{t \rightarrow \infty} p_{j,i}(t) \leq \lim_{t \rightarrow \infty} \left(\sum_{k=1}^{t_1} f_{j,i}(k) p_{i,i}(t-k) + \varepsilon \right) = \sum_{k=1}^{t_1} f_{j,i}(k) \lim_{t \rightarrow \infty} p_{i,i}(t) + \varepsilon \leq \lim_{t \rightarrow \infty} p_{i,i}(t) + \varepsilon.$$

Wobec dowolności ε mamy

$$\lim_{t \rightarrow \infty} p_{j,i}(t) = \lim_{t \rightarrow \infty} p_{i,i}(t) = \frac{1}{E[\tau_{i,i}]}.$$

Weźmy π jak w wypowiedzi. Suma wyrazów w wierszu P^t wynosi 1, a więc

$$1 = \lim_{t \rightarrow \infty} \sum_{k \in S} p_{j,k}(t) = \sum_{k \in S} \lim_{t \rightarrow \infty} p_{j,k}(t) = \sum_{k \in S} \pi_k.$$

Do tego $p_{j,i}(t+1) = \sum_{k \in S} p_{j,k}(t) P_{ki}$, co w przejściu granicznym daje

$$\pi_i = \sum_{k \in S} \pi_k P_{ki}.$$

Te dwie równości pokazują, że π jest rozkładem i to stacjonarny. Teraz pokażemy jedyność. Niech ϕ będzie drugim rozkładem stacjonarnym. Rozważmy $x \in S$ taki, że $\frac{\pi(x)}{\phi(x)}$ jest minimalne. Zachodzi

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

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 28. 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 (i zauważając, że z sum po obu stronach powinny zniknąć składniki odpowiadające nowym elementom zbioru, a one się skrócą) dostajemy odpowiednią równość dla dowolnego zbioru S . $\square$

Twierdzenie 29. Rozważmy skończony, nieprzywiedlny, nieokresowy łańcuch Markowa. Niech $\pi = (\pi_1, \dots, \pi_n)$ będzie wektorem nieujemnych liczb takich, że $\sum_{i=1}^n \pi_i = 1$. Jeśli dla każdego dwóch stanów i, j zachodzi

$$\pi_i P_{ij} = \pi_j P_{ji},$$

to π jest rozkładem stacjonarnym tego łańcucha Markowa. Łańcuchy, w których zachodzi ten warunek, nazywamy łańcuchami z odwracalnym czasem.

Dowód. Dla każdego j mamy

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

a więc $\pi P = \pi$. W połączeniu z $\sum_{i=1}^n \pi_i = 1$ dostajemy, że π jest rozkładem stacjonarnym. $\square$

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.

Można też zgadnąć wynik i pokazać jego poprawność – to działa ładnie zwłaszcza w łańcuchach z odwracalnym czasem.

10. Spacery po grafach

2024-11-24

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

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

($\Leftarrow$) 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 nieprzywiedlny i nieokresowe. W szczególności mają one rozkład stacjonarny.

Twierdzenie 30. 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. $\square$

Definicja 49. Czas pokrycia grafu G to chwila (indeks łańcucha Markowa), w której spacer odwiedził już każdy wierzchołek. Taką zmienną losową oznaczamy C_G .

Lemat 12. Dla każdej krawędzi uv w grafie G zachodzi $E[\tau_{uv}] + E[\tau_{vu}] \leq 2E(G)$.

Dowód. Mając graf G będziemy tworzyć łańcuch Markowa na krawędziach skierowanych. Rozważamy skierowany graf D , który jest grafem G , w którym każda krawędź została przedstawiona jako dwie krawędzie skierowane. Stanem łańcucha będą krawędzie, a z zadanej krawędzi będzie można przejść do krawędzi wychodzących z jej końca (z równym prawdopodobieństwem).

W takim łańcuchu rozkład jednostajny $\pi_{uv} = \frac{1}{2E(G)}$ jest stacjonarny. Po pierwsze $\sum_{uv \in E(D)} \pi_{uv} = \sum_{uv \in E(D)} \frac{1}{2E(G)} = 1$, więc jest to rozkład. Mamy też

$$\sum_{w \in N(u)} \pi_{uw} \frac{1}{d(u)} = \frac{1}{d(u)} \cdot \frac{d(u)}{2E(G)} = \frac{1}{2E(G)} = \pi_{uv},$$

gdzie uv jest pewną krawędzią w D . Z tego wynika, że rozkład jest stacjonarny.

Ograniczana wartość $E[\tau_{uv}] + E[\tau_{vu}]$ jest oczekiwaną liczbą kroków w spacerze $u \rightarrow v \rightarrow u$. W grafie D można patrzeć na spacer z krawędzi vu do vu . Idzie on tak samo jak przejście z u do v i z powrotem do u , ale ma ustaloną krawędź, którą trzeba wrócić do u . Zatem będzie dłuższy od zwykłego spaceru po wierzchołkach i mamy

$$E[\tau_{uv}] + E[\tau_{vu}] \leq E[\tau_{(vu)(vu)}] = \frac{1}{\pi_{vu}} = 2E(G).$$

□

Lemat 13. Oczekiwany czas pokrycia grafu jest ograniczony przez

$$E[C_G] \leq 2|E| \cdot (|V| - 1),$$

gdzie E, V to zbiory krawędzi i wierzchołków tego grafu.

Dowód. Niech T będzie drzewem rozpinającym G . Przejdziemy po jego wierzchołkach w kolejności DFSa. Niech $v_0, v_1, \dots, v_{2|V|-2}$ będą kolejnymi wierzchołkami odwiedzionymi przez DFSa. Oczekiwany czas pokrycia grafu jest ograniczony przez oczekiwany czas kolejnego odwiedzania wierzchołków wypisanych w takiej kolejności. Zatem

$$E[C_G] \leq \sum_{i=0}^{2|V|-3} E[\tau_{v_i v_{i+1}}] = \sum_{xy \in T} E[\tau_{xy}] + E[\tau_{yx}] \leq 2|E| \cdot (|V| - 1).$$

□

Lemat 14. Zachodzi

$$E[C_G] \leq H(n-1) \cdot \max_{\substack{u, v \in V, \\ u \neq v}} E[\tau_{uv}],$$

gdzie $G = (V, E)$ jest grafem, $|V| = n$ a $H(n)$ to liczba harmoniczna.

Dowód. Oznaczmy $B = \max_{\substack{u, v \in V, \\ u \neq v}} E[\tau_{uv}]$. Wybierzmy losową permutację wierzchołków $Z_1, \dots, Z_n$ i pewien konkretny wierzchołek startowy u . Niech $T_1, \dots, T_n$ będą zmiennymi losowymi takimi, że T_j oznacza pierwszy moment przejścia przez wierzchołki $Z_1, \dots, Z_j$. Dla $j \geq 2$ definiujemy

$$Y_j = E[T_j - T_{j-1} \mid Z_1, \dots, Z_{j-1}, X_1, \dots, X_{T_{j-1}}],$$

czyli oczekiwany czas między odwiedzeniem odpowiednich wierzchołków warunkowany stanem na-

szego eksperymentu – przez X_r oznaczamy stan naszego spaceru w momencie r . Zachodzi

$$E[C_G] \leq \sum_{j=2}^n Y_j + E[T_1],$$

bo jest to wartość oczekiwana przejścia po wierzchołkach w niedowolnej kolejności.

Jeśli $Z_1 = u$ (co ma prawdopodobieństwo $\frac{1}{n}$), to jest $T_1 = 0$. Inaczej $E[T_1 | Z_1] = E[\tau_{u, Z_1}] \leq B$. Zatem $E[T_1] \leq (1 - \frac{1}{n})B$.

Jeśli Z_j nie jest ostatnim odwiedzionym wierzchołkiem spośród $Z_1, \dots, Z_j$, to $Y_j = 0$, bo $T_j = T_{j-1}$. Jeśli jest, to mamy $Y_j = E[\tau_{Z_k, Z_j}] \leq B$, gdzie Z_k jest ostatnim wierzchołkiem odwiedzionym spośród $Z_1, \dots, Z_{j-1}$. Z_j jest ostatnim odwiedzionym wierzchołkiem z prawdopodobieństwem $\frac{1}{j}$, a więc $Y_j \leq \frac{B}{j}$. Zatem mamy

$$E[C_G] \leq \sum_{j=2}^n \frac{B}{j} + \left(1 - \frac{1}{n}\right)B = \left(1 + \sum_{j=2}^n \frac{1}{j}\right)B - \frac{1}{n}B = H(n-1)B.$$

□

11. Ciągłe zmienne losowe

2024-11-29

Definicja 50. Zmienna losowa X jest ciągła, jeśli istnieje funkcja $f: \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ taka, że

$$\forall B \in \mathcal{B}(\mathbb{R}) \quad P(X \in B) = \int_B f(x) dx.$$

Taką funkcję nazywamy funkcją gęstości lub gęstością zmiennej X .

Uwaga. Własności funkcji gęstości:

- $\forall x \in \mathbb{R} \quad f(x) \geq 0$
- $\int_{\mathbb{R}} f(x) dx = 1$
- $P(y < X < y + \delta) = \int_y^{y+\delta} f(x) dx \approx \delta \cdot f(y)$. Czyli $f(x)$ mówi, jak szybko rośnie prawdopodobieństwo przyjmowania wartości z przedziału, gdy przedział zaczyna się od x .

Uwaga. Mamy $\forall x \in \mathbb{R} \quad P(X = x) = 0$. W szczególności daje to $P(X \leq x) = P(X < x)$.

Definicja 51. Dystrybuenta zmiennej losowej X to funkcja $F(x) = P(X \leq x)$. Dla zmiennej ciągłej jest $F(x) = \int_{-\infty}^x f(y) dy$, a więc $f(x) = F'(x)$.

Definicja 52. Wartość oczekiwana ciągłej zmiennej X to

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

Przykład. Mamy

$$\begin{aligned} \text{Var}(X) &= E[(X - E[X])^2] = \int_{\mathbb{R}} (x - E[X])^2 f(x) dx \\ &= \int_{\mathbb{R}} x^2 f(x) dx - 2E[X] \int_{\mathbb{R}} x f(x) dx + \int_{\mathbb{R}} E[X]^2 f(x) dx \\ &= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2. \end{aligned}$$

Propozycja 19. Dla ciągłej zmiennej losowej X przyjmującej wartości nieujemne jest

$$E[X] = \int_0^{\infty} P(X \geq x) dx.$$

Dowód. Niech f będzie gęstością X . Mamy

$$\begin{aligned} \int_0^{\infty} P(X \geq x) dx &= \int_{x=0}^{\infty} \int_{y=x}^{\infty} f(y) dy dx = \int_{y=0}^{\infty} \int_{x=0}^y f(y) dx dy \\ &= \int_{y=0}^{\infty} f(y) \int_{x=0}^y dx dy = \int_{y=0}^{\infty} y f(y) dy = \int_{-\infty}^{\infty} y f(y) dy, \end{aligned}$$

gdzie ostatnie przejście wynika z nieujemności X . □

Definicja 53. Wspólna dystrybuanta zmiennych losowych X, Y to

$$F(x, y) = P(X \leq x, Y \leq y).$$

Definicja 54. Wspólna gęstość ciągłych zmiennych losowych X, Y to funkcja f taka, że

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du,$$

a więc

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

Definicja 55. Dla dwóch zmiennych X, Y o zadanej wspólnej dystrybuancie $F(x, y)$ brzegowa dystrybuanta zmiennej X to funkcja

$$F_X(x) = \lim_{y \rightarrow \infty} F(x, y) = P(X \leq x),$$

której odpowiada brzegowa gęstość $f_X(x)$. Analogiczne pojęcia definiujemy dla zmiennej Y .

Definicja 56. Zmienne X, Y są niezależne, jeśli

$$\forall x, y \in \mathbb{R} \quad P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y).$$

Uwaga. Niezależność jest równoważna odpowiednim równościom dystrybant i gęstości:

$$F(x, y) = F_X(x) F_Y(y),$$

$$f(x, y) = f_X(x) f_Y(y).$$

Definicja 57. Prawdopodobieństwo warunkowe definiujemy jako całkę

$$P(X \leq x \mid Y = y) = \int_{u=-\infty}^x \frac{f(u, y)}{f_Y(y)} du,$$

gdzie funkcję $f_{X|Y} = \frac{f(x, y)}{f_Y(y)}$ nazywamy warunkową gęstością.

Definicja 58. Warunkowa wartość oczekiwana to całka

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx.$$

Definicja 59. Rozkład ciągłej zmiennej losowej X nazywamy jednostajnym na $[a, b]$ (krańce nie

mają znaczenia, mogą być też otwarte), jeśli jej dystrybuenta przyjmuje wartości

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & x \in [a, b] \\ 1 & x > b \end{cases}.$$

Gęstość jest wtedy funkcją

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & \text{wpp} \end{cases}.$$

Przykład. Wartość oczekiwana zmiennej ciągłej X jednostajnej na $[a, b]$ to

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b x \cdot \frac{1}{b-a} dx = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

Wariancję dostaniemy z podobnych przeliczeń dla drugiego momentu:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{b^2 + ab + a^2}{3}.$$

Propozycja 20. Funkcja tworząca momentów zmiennej losowej X o rozkładzie jednostajnym na $[a, b]$ to

$$M_X(t) = \begin{cases} \frac{e^{tb} - e^{ta}}{t(b-a)} & t \neq 0 \\ 1 & t = 0 \end{cases}.$$

Dowód.

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} f(x) e^{tx} dx = \frac{1}{b-a} \int_a^b e^{tx} dx = \begin{cases} \frac{e^{tb} - e^{ta}}{t(b-a)} & t \neq 0 \\ 1 & t = 0 \end{cases}.$$

□

Propozycja 21. Niech X ma rozkład jednostajny na $[a, b]$. Jeśli $a \leq c \leq d \leq b$, to

$$P(X \leq c \mid X \leq d) = \frac{c-a}{d-a},$$

czyli wynikowa zmienna ma rozkład jednostajny na krótszym przedziale.

Dowód.

$$\frac{P(X \leq c \cap X \leq d)}{P(X \leq d)} = \frac{P(X \leq c)}{P(X \leq d)} = \frac{c-a}{b-a} \cdot \frac{b-a}{d-a} = \frac{c-a}{d-a}.$$

□

Twierdzenie 31. Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi o rozkładzie jednostajnym na $[0, 1]$. Niech $(Y_1, \dots, Y_n) = \text{sort}(X_1, \dots, X_n)$. Wtedy

$$\forall_{k \in [n]} E[Y_k] = \frac{k}{n+1}.$$

Dowód. Pokażemy to wprost dla $k = 1$. Mamy $Y_1 = \min(X_1, \dots, X_n)$, zatem

$$P(Y_1 \geq y) = P(X_1 \geq y \cap \dots \cap X_n \geq y) = \prod_{i \in [n]} P(X_i \geq y) = (1-y)^n.$$

Z tego mamy dystrybuantę i gęstość

$$\begin{aligned} F(y) &= 1 - (1 - y)^n, \\ f(y) &= n(1 - y)^{n-1}. \end{aligned}$$

Teraz możemy policzyć wartość oczekiwaną wprost z definicji:

$$E[Y_1] = \int_{\mathbb{R}} y f(y) dy = \int_0^1 y n (1 - y)^{n-1} dy.$$

To liczymy przez części:¹

$$E[Y_1] = -y(1 - y)^n \Big|_0^1 - \int_0^1 -(1 - y)^n dy = \frac{-1}{n+1} (1 - y)^{n+1} \Big|_0^1 = \frac{1}{n+1}.$$

Lub stosując przejście

$$E[Y_1] = \int_0^\infty P(Y_1 \geq y) dy = \int_0^\infty (1 - y)^n = \frac{1}{n+1}.$$

Dla większych k takie przeliczenia są jeszcze trudniejsze, można zastosować bardziej pomysłowy argument: zamiast losować na odcinku losujemy jednostajnie punkty $P_0, \dots, P_n$ na okręgu o obwodzie 1. Definiujemy $X_i = P_i - P_0$ (odległość na łuku idącym zgodnie ze wskazówkami zegara). Rozkład dalej jest jednostajny, a zmienne są niezależne.

Jest pełna symetria, a więc odcinki $P_i P_{i+1}$ są w oczekiwaniu tej samej długości (można wybrać za P_0 dowolny z wylosowanych punktów, więc długość każdego odcinka ma taki sam rozkład jak pierwszego). Zatem mają długość $\frac{1}{n+1}$. A teraz nasze zmienne to sumy długości odcinków. $\square$

¹Warto zaznaczyć, iż prowadzący wykład policzył tę całkę samodzielnie.

12. Rozkład wykładniczy

2024-12-10

Pomysł. Mamy kilka okienek w urzędzie, czekamy w kolejce na zwolnienie któregoś. Czas oczekiwania będziemy mogli modelować właśnie rozkładem wykładniczym.

Definicja 60. Dla parametru $\lambda > 0$ rozkład o funkcji gęstości $f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$ nazywamy rozkładem wykładniczym. Prawdopodobieństwo spada wykładniczo wraz ze wzrostem x . Mamy

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \lim_{R \rightarrow \infty} \int_0^R \lambda e^{-\lambda x} dx = \lim_{R \rightarrow \infty} -e^{-\lambda x} \Big|_0^R = \lim_{R \rightarrow \infty} -e^{-\lambda R} + 1 = 1,$$

wiec jest to poprawna gęstość.

Propozycja 22. Dystrybuenta, wartość oczekiwana i wariancja zmiennej X o rozkładzie wykładniczym z parametrem λ to

$$\begin{aligned} F(t) &= \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 & t \leq 0 \end{cases}, \\ E[X] &= \frac{1}{\lambda}, \\ \text{Var}(X) &= \frac{1}{\lambda^2}. \end{aligned}$$

Dowód.

$$F(t) = \int_{-\infty}^t f(x) dx = -e^{-\lambda x} \Big|_0^t = 1 - e^{-\lambda t},$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} \lambda x e^{-\lambda x} dx = -x e^{-\lambda x} \Big|_0^{\infty} - \int_0^{\infty} -e^{-\lambda x} dx = -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} = \frac{1}{\lambda},$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \dots^1 = \frac{2}{\lambda^2},$$

a więc $\text{Var}(X) = \frac{1}{\lambda^2}$. □

¹Ćwiczenie dla czytelnika, wymaga podwójnego całkowania przez części (adnotacja powstała na prośbę pewnej osoby z miasta, gdzie są tylko lasy i tramwaje).

Propozycja 23. Funkcja tworząca momentów zmiennej losowej X o rozkładzie wykładniczym z parametrem λ to

$$M_X(t) = \begin{cases} \frac{\lambda}{\lambda - t} & t < \lambda \\ +\infty & t \geq \lambda \end{cases}.$$

Dowód.

$$M_X(t) = \int_0^{\infty} \lambda e^{-\lambda x} e^{tx} dx = \frac{\lambda}{t - \lambda} e^{x(t-\lambda)} \Big|_0^{\infty} = \begin{cases} \frac{\lambda}{\lambda - t} & t < \lambda \\ +\infty & t \geq \lambda \end{cases}.$$

□

Lemat 15. Rozkład wykładniczy ma własność „bez pamięci”, czyli dla zmiennej $X \sim \text{Exp}(\lambda)$ i $s, t > 0$ mamy

$$P(X > s + t \mid X > t) = P(X > s).$$

Dowód.

$$\begin{aligned} P(X > s + t \mid X > t) &= \frac{P(X > s + t)}{P(X > t)} = \frac{1 - F_X(s + t)}{1 - F_X(t)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} = 1 - F_X(s) = P(X > s). \end{aligned}$$

□

Lemat 16. Rozkład wykładniczy jest jedynym ciągłym rozkładem bez pamięci, czyli jeśli X jest ciągłą zmienną losową i zachodzi

$$\forall s, t > 0 \quad P(X > s + t \mid X > t) = P(X > s),$$

to istnieje takie $\lambda > 0$, że $X \sim \text{Exp}(\lambda)$.

Dowód. Niech $S(x) = 1 - F_X(x)$. Mamy

$$\frac{S(s+t)}{S(t)} = S(s) \implies S(s+t) = S(s)S(t).$$

Z tego widzimy $S(2t) = S(t)^2$ i indukując mamy $S(pt) = S(t)^p$, $S\left(\frac{1}{q}t\right) = S(t)^{\frac{1}{q}}$, czyli $S\left(\frac{p}{q}t\right) = S(t)^{\frac{p}{q}}$, a więc $\forall r \in \mathbb{Q}_{\geq 0} \quad S(rt) = S(t)^r$, teraz dla $a \in \mathbb{R}_{\geq 0}$ znajdujemy ciąg $\{r_n\}$ zbiegający do a i mamy $S(r_n t) = S(t)^{r_n}$, a więc z ciągłości $S(at) = S(t)^a$.

Wstawiając $t = 1$ mamy $S(a) = S(1)^a$. Definiujemy $\lambda = -\ln S(1)$. Dla $x \in \mathbb{R}_{\geq 0}$ mamy $S(x) = S(1)^x = e^{\ln S(1) \cdot x} = e^{-\lambda x}$, czyli rozkład wykładniczy.

$S(x)$ jest nierosnąca i mamy $\lim_{x \rightarrow 0^+} S(x) = 1$, więc $P(x < 0) = 0$, czyli dla liczb ujemnych też się zgadza. □

Lemat 17. Niech X, Y będą niezależnymi zmiennymi o rozkładzie wykładniczym z parametrami λ i μ . Zachodzi $\min(X, Y) \sim \text{Exp}(\lambda + \mu)$. Dodatkowo $P(X < Y) = \frac{\lambda}{\lambda + \mu}$.

Dowód.

$$P(\min(X, Y) > t) = P(X > t \cap Y > t) = P(X > t)P(Y > t) = e^{-\lambda t}e^{-\mu t} = e^{-(\lambda+\mu)t}.$$

$$\begin{aligned} P(X < Y) &= P((X, Y) \in \{(x, y) \in \mathbb{R}^2 : x < y\}) = \int_0^\infty \int_0^y f_{X,Y}(x, y) dx dy \\ &= \int_0^\infty \int_0^y \lambda e^{-\lambda x} \mu e^{-\mu y} dx dy = \int_0^\infty \mu e^{-\mu y} (1 - e^{-\lambda y}) dy = \int_0^\infty \mu e^{-\mu y} dy - \int_0^\infty \mu e^{-(\lambda+\mu)y} dy \\ &= 1 - \left[-\frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)y} \right]_0^\infty = 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}. \end{aligned}$$

□

Uwaga. Żeby znaleźć $P((X, Y) \in A)$ dla $A \in \mathbb{R}^2$ liczymy $\int_A f_{X,Y}(x, y) dx dy$.

Wniosek. Dla niezależnych zmiennych $X_1 \sim \text{Exp}(\lambda_1), \dots, X_k \sim \text{Exp}(\lambda_k)$ mamy

$$P(X_j = \min(X_1, \dots, X_k)) = \frac{\lambda_j}{\sum_{i=1}^k \lambda_i}.$$

Definicja 61. Definiujemy funkcję gamma jako

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$$

dla każdego $a > 0$.

Propozycja 24. Zachodzi $\Gamma(a+1) = a\Gamma(a)$.

Dowód.

$$\Gamma(a+1) = \int_0^\infty x^a e^{-x} dx = -x^a e^{-x} \Big|_0^\infty + a \int_0^\infty x^{a-1} e^{-x} dx = 0 + a\Gamma(a).$$

□

Uwaga. Mamy $\Gamma(1) = 1$, a więc $\Gamma(n) = (n-1)!$. Zatem funkcja gamma jest w pewnym sensie uogólnieniem pojęcia silni na liczby rzeczywiste.

Definicja 62. Zmienna losowa X ma rozkład $\Gamma(a, 1)$ jeśli jej gęstość jest zadana wzorem

$$f(x) = \begin{cases} \frac{1}{\Gamma(a)} x^{a-1} e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}.$$

Propozycja 25. Jeśli $X \sim \Gamma(a, 1)$, to $E[X^c] = \frac{\Gamma(a+c)}{\Gamma(a)}$.

Dowód.

$$E[X^c] = \frac{1}{\Gamma(a)} \int_0^\infty x^{a-1+c} e^{-x} dx = \frac{\Gamma(a+c)}{\Gamma(a)}.$$

□

Wniosek. Dla $X \sim \Gamma(a, 1)$ zachodzi $E[X] = \frac{\Gamma(a+1)}{\Gamma(a)} = a$ i $\text{Var}(X) = (a+1)a - a^2 = a$.

Definicja 63. Zmienna losowa X ma rozkład $\Gamma(a, \lambda)$, jeśli istnieje takie $Y \sim \Gamma(a, 1)$, że $X = \frac{Y}{\lambda}$.

Taka zmienna losowa ma gęstość

$$f_X(x) = f_Y(y) \left| \frac{dy}{dx} \right| = \frac{1}{\Gamma(a)} (\lambda x)^{a-1} e^{-\lambda x} \lambda = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}.$$

Uwaga. Dla $X \sim \Gamma(a, \lambda)$ zachodzi $E[X] = \frac{a}{\lambda}$ i $\text{Var}(X) = \frac{a}{\lambda^2}$.

Przykład (Problem kul i urn ze wzmocnionym feedbackiem). Rozważamy problem kul i urn, w którym im więcej kul jest w urnie, tym większa szansa na wpadnięcie tam. Mamy dwie urny (np. czerwona i niebieska), zastanawiamy się, w której będzie więcej kul (po długim czasie).

Lemat 18 (Feedback bez wzmocnienia). Niech X_n, Y_n będą liczbą kul wrzuconych odpowiednio do czerwonej i niebieskiej urny w momencie czasu n . Zaczynamy z $X_0 = Y_0 = 1$. Jeśli $X = x$ i $Y = y$, to niech prawdopodobieństwo wybrania kolejnej kuli do czerwonej urny wynosi $\frac{x}{x+y}$.

Dla $x \in [n+1]$ mamy $P(X_n = x) = \frac{1}{n+1}$ (i analogicznie dla Y).

Dowód. Dowód indukcyjny, baza oczywista. Dla $n > 0$ jest

$$P(X_n = x) = \frac{x-1}{n+1} P(X_{n-1} = x-1) + \frac{n+1-x}{n+1} P(Y_{n-1} = n+1-x) = \frac{n}{n+1} \cdot \frac{1}{n} = \frac{1}{n+1},$$

gdzie pierwsze przejście to rozważenie przypadków osiągnięcia x kul w czerwonej urnie, a drugie to założenie indukcyjne. $\square$

Twierdzenie 32 (Feedback ze wzmocnieniem). Niech X_n, Y_n będą liczbą kul wrzuconych odpowiednio do czerwonej i niebieskiej urny w momencie czasu n . Zaczynamy z $X_0 = x_0, Y_0 = y_0$. Jeśli $X = x$ i $Y = y$, to niech prawdopodobieństwo wybrania kolejnej kuli do czerwonej urny wynosi $\frac{x^p}{x^p + y^p}$ dla ustalonego parametru p .

Dla każdych $x_0, y_0 \in \mathbb{N}_1$ i $p > 1$ z prawdopodobieństwem 1 istnieje $c \in \mathbb{N}$ takie, że jedna z urn otrzyma mniej niż c kul (przy rzucaniu w nieskończoność). Innymi słowy

$$P\left(\lim_{n \rightarrow \infty} X_n = \infty \cap \lim_{n \rightarrow \infty} Y_n = \infty\right) = 0.$$

Dowód. Niech T_x będzie zmienną modelującą czas oczekiwania na $(x+1)$ -szą kulę w urnie czerwonej mając w niej x kul. W naszym modelu $T_x \sim \text{Exp}(x^p)$, czyli zmienna T_x formalnie nie ma nic wspólnego z kulami i urnami. Podobnie definiujemy U_y . Wcześniej widzieliśmy, że zachodzi

$$P(T_x < U_y) = \frac{x^p}{x^p + y^p},$$

a więc te zmienne faktycznie dobrze modelują nasz eksperyment – szansa na „skończenie” T_x przed U_y jest taka, jak wrzucenia kuli do urny czerwonej. W momencie skończenia T_x możemy „zrestartować” U_y (czyli skorzystać z własności bez pamięci rozkładu wykładniczego) i zacząć T_{x+1} – odpowiednia nierówność będzie dalej zachodzić. Oczywiście wszystko działa tak samo, gdy skończy się U_y .

Definiujemy

$$F = \sum_{j=x_0}^{\infty} T_j, \quad G = \sum_{j=y_0}^{\infty} U_j.$$

Te wartości oznaczają momenty, w których liczba kul wrzuconych do danej urny staje się nieograniczona. Mamy

$$E[F] = \sum_{j=x_0}^{\infty} E[T_j] = \sum_{j=x_0}^{\infty} \frac{1}{j^p} < \infty.$$

To samo zachodzi dla G . Zatem $P(F < \infty) = P(G < \infty) = 1$ i te liczby są dobrze zdefiniowane z prawdopodobieństwem 1 (ograniczona wartość oczekiwana daje zerowe prawdopodobieństwo tego, że zmienna jest nieograniczona). Mamy $P(F \neq G) = 1$ (tak zachowują się każde niezależne zmienne

ciągłe), czyli $P(F < G \cup F > G) = 1$. Załóżmy, że $F < G$. Mamy zatem (dla pewnego t)

$$\sum_{j=1}^t U_j < F < \sum_{j=1}^{t+1} U_j \implies \sum_{j=1}^t U_j < \sum_{i=1}^m T_i < \sum_{j=1}^{t+1} U_j$$

dla wszystkich odpowiednio dużych m . To znaczy, że do niebieskiej urny wpadnie t kul i zanim wpadnie kolejna, do czerwonej urny wpadnie nieograniczona liczba kul – czyli w granicy jest $\lim_{n \rightarrow \infty} Y_n = t$.

Analogiczny argument działa, gdy $G > F$. Zatem z prawdopodobieństwem 1 zachodzi przedstawiona sytuacja, co kończy dowód. $\square$

13. Proces Poissona

2024-12-13

Definicja 64. Rozważmy ciąg zdarzeń losowych. Niech $N(t)$ zlicza liczbę zdarzeń w przedziale czasu $[0, t]$. Proces $\{N(t) : t \geq 0\}$ nazywamy stochastycznym procesem zliczającym.

Definicja 65. Proces Poissona z parametrem $\lambda > 0$ to stochastyczny proces zliczający taki, że:

1. $N(0) = 0$
2. proces ma stacjonarne i niezależne przyrosty, czyli
 - (a) $\forall s, t \geq 0$ $N(t)$ i $N(s+t) - N(s)$ mają ten sam rozkład – proces nie zmienia się w czasie.
 - (b) $\forall t_1 < t_2 \leq t_3 < t_4$ $N(t_2) - N(t_1)$ i $N(t_4) - N(t_3)$ są niezależne, czyli zdarzenia na niezależnych przedziałach są niezależne.
3. $\lim_{t \rightarrow 0} \frac{P(N(t)=1)}{t} = \lambda$, czyli prędkość pojawiania się zdarzeń wynosi λ .
4. $\lim_{t \rightarrow 0} \frac{P(N(t) \geq 2)}{t} = 0$, czyli wystąpienie więcej niż jednego wydarzenia na raz jest nieprawdopodobne.

Twierdzenie 33. Niech $\{N(t) : t \geq 0\}$ będzie procesem Poissona z parametrem λ . Wtedy

$$\forall t, s \geq 0, n \in \mathbb{N} \quad P(N(t+s) - N(s) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

czyli taka zmienna ma rozkład Poissona z parametrem λt .

Dowód. Indukcja po n . Niech $P_n(t) := P(N(t+s) - N(s) = n)$. Dowodzimy bazy $n = 0$. Mamy

$$\begin{aligned} P_0(t+h) &= P(N(t+h) = 0) = P(N(t) = 0 \cap N(t+h) - N(t) = 0) \\ &= P(N(t) = 0) \cdot P(N(t+h) - N(t) = 0) = P(N(t) = 0) \cdot P(N(h) = 0) = P_0(t) \cdot P_0(h), \end{aligned}$$

gdzie najpierw podstawiamy $s = 0$, potem korzystamy z niezależności i stacjonarności. Mamy

$$\begin{aligned} P'_0(t) &= \lim_{h \rightarrow 0} \frac{P_0(t+h) - P_0(t)}{h} = \lim_{h \rightarrow 0} \frac{P_0(t) \cdot P_0(h) - P_0(t)}{h} = P_0(t) \lim_{h \rightarrow 0} \frac{P_0(h) - 1}{h} \\ &= P_0(t) \lim_{h \rightarrow 0} \frac{1 - P(N(h) = 1) - P(N(h) \geq 2) - 1}{h} = P_0(t) (-\lambda - 0) = -\lambda P_0(t), \end{aligned}$$

gdzie w przedostatnim przejściu skorzystaliśmy z obu granic z definicji procesu Poissona. Mamy $\frac{P'_0(t)}{P_0(t)} = -\lambda$, więc całkując obustronnie dostajemy $\ln(P_0(t)) = -\lambda t + C$, a więc

$$P_0(t) = e^{-\lambda t + C},$$

a podstawiając $P_0(0) = 1$ mamy $C = 0$, więc jest tak jak chcemy.

Teraz robimy krok indukcyjny dla $n \geq 1$. Mamy

$$P_n(t+h) = \sum_{k=0}^n P_{n-k}(t) \cdot P_k(h),$$

bo jeśli $n-k$ zdarzeń będzie miało miejsce w ciągu t jednostek czasu, to pozostałe k musi się wydarzyć w kolejnych h . Mamy

$$\begin{aligned} P'_n(t) &= \lim_{h \rightarrow 0} \frac{P_n(t+h) - P_n(t)}{h} = \lim_{h \rightarrow 0} \frac{\sum_{k=0}^n P_{n-k}(t) \cdot P_k(h) - P_n(t)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \left(P_n(t)(P_0(h) - 1) + P_{n-1}(t)P_1(h) + \sum_{k=2}^n P_{n-k}(t)P_k(h) \right) \\ &= P_n(t) \lim_{h \rightarrow 0} \frac{P_0(h) - 1}{h} + P_{n-1}(t) \lim_{h \rightarrow 0} \frac{P_1(h)}{h} = P_n(t) \cdot (-\lambda) + P_{n-1}(t) \cdot \lambda, \end{aligned}$$

gdzie ostatnie przejście to zastosowanie granicy z poprzednich przeliczeń i definicji procesu Poissona, a przedostatnie wynika z szacowania

$$\frac{1}{h} \sum_{k=2}^n P_{n-k}(t) P_k(h) \leq \frac{1}{h} P(N(h) \geq 2) \rightarrow 0.$$

Rozwiązujemy równanie różniczkowe $P'_n(t) + \lambda P_n(t) = \lambda P_{n-1}(t)$. Przekształcamy domnażając przez $e^{\lambda t}$: $e^{\lambda t}(P'_n(t) + \lambda P_n(t)) = e^{\lambda t} \lambda P_{n-1}(t)$, czyli (wzór na pochodną mnożenia):

$$\frac{d}{dt} e^{\lambda t} P_n(t) = e^{\lambda t} \cdot \lambda P_{n-1}(t) = e^{\lambda t} \cdot \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} = \lambda \frac{(\lambda t)^{n-1}}{(n-1)!}.$$

Teraz całkujemy i mamy $e^{\lambda t} P_n(t) = \lambda^n \cdot \frac{1}{n!} \cdot t^n + C$, licząc w $t = 0$ dostajemy $C = 0$. $\square$

Twierdzenie 34. Niech $\{N(t) : t \geq 0\}$ będzie procesem zliczającym takim, że

1. $N(0) = 0$
2. proces ma niezależne przyrosty
3. $\forall t, s \geq 0$ $N(s+t) - N(s)$ jest zmienną o rozkładzie Poissona z parametrem λt .

Taki proces jest procesem Poissona z parametrem λ .

Dowód. Stacjonarność przyrostów wynika z tego, że zawsze mają zadany rozkład Poissona, więc w szczególności zawsze mają taki sam rozkład. Pozostaje nam policzyć odpowiednie granice:

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{P(N(t) = 1)}{t} &= \lim_{t \rightarrow 0} \frac{e^{-\lambda t} \lambda t}{t} = \lambda, \\ \lim_{t \rightarrow 0} \frac{P(N(t) \geq 2)}{t} &= \lim_{t \rightarrow 0} \frac{1}{t} (1 - e^{-\lambda t} - e^{-\lambda t} \lambda t) = \lim_{t \rightarrow 0} \frac{\lambda e^{-\lambda t}}{1} - \lambda = 0, \end{aligned}$$

gdzie w ostatnim przejściu skorzystaliśmy z reguły de l'Hospitala i poprzedniej granicy. $\square$

Definicja 66 (Międzyzasy). Niech $\{N(t) : t \geq 0\}$ będzie procesem zliczającym. Definiujemy x_1 jako czas pierwszego zdarzenia, a x_i jako czas pomiędzy $(i-1)$ -szym a i -tym zdarzeniem (dla $i \geq 2$.)

Twierdzenie 35. Niech $\{N(t) : t \geq 0\}$ będzie procesem Poissona z parametrem λ . Wtedy międzyzasy procesu to niezależne zmienne o rozkładzie wykładniczym z parametrem λ .

Dowód.

$$P(x_1 > t) = P(N(t) = 0) = e^{-\lambda t}.$$

Mamy $P(x_1 \leq t) = 1 - e^{-\lambda t}$, a to jest dystrybucja rozkład wykładniczego.

$$\begin{aligned} P(x_i > t_i \mid (x_1, \dots, x_{i-1}) = (t_1, \dots, t_{i-1})) &= P\left(N\left(\sum_{j=1}^i t_j\right) - N\left(\sum_{j=1}^{i-1} t_j\right) = 0\right) \\ &= P(N(t_i) = 0) = e^{-\lambda t_i}. \end{aligned}$$

To prawdopodobieństwo jest niezależne od warunku, więc rozważane zmienne są niezależne. Wiadomo też, że x_i jest zadane odpowiednim rozkładem wykładniczym. $\square$

Twierdzenie 36. Niech $\{N(t) : t \geq 0\}$ będzie procesem zliczającym takim, że

1. $N(0) = 0$
2. międzyczasy procesu są niezależne i mają rozkład wykładniczy z parametrem λ .

Taki proces jest procesem Poissona z parametrem λ .

Dowód. Stacjonarność przyrostów wynika z tego, że można spojrzeć na aktualny międzyczas w czasie s i skorzystać z własności bez pamięci. Zaczynając zliczać zdarzenia od tego momentu dostajemy proces taki sam jak początkowy proces.

Niezależność przyrostów wygląda tak samo: restartujemy międzyczasy na początkach przedziałów i korzystamy z niezależności różnych międzyczasów oraz własności bez pamięci.

Zostały nam granice:

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{P(N(t) = 1)}{t} &= \lim_{t \rightarrow 0} \frac{P(x_1 \leq t \cap x_1 + x_2 > t)}{t} = \lim_{t \rightarrow 0} \frac{1}{t} \cdot \int_0^t \int_{t-x}^{\infty} \lambda e^{-\lambda x} \lambda e^{-\lambda y} dy dx = \lambda, \\ \lim_{t \rightarrow 0} \frac{P(N(t) \geq 2)}{t} &= \lim_{t \rightarrow 0} \frac{P(x_1 + x_2 \leq t)}{t} = \lim_{t \rightarrow 0} \frac{1}{t} \int_0^t \int_0^{t-x} \lambda e^{-\lambda x} \lambda e^{-\lambda y} dy dx = 0. \end{aligned}$$

$\square$

Definicja 67. Dwa procesy zliczające $\{N_1(t) : t \geq 0\}$ i $\{N_2(t) : t \geq 0\}$ są niezależne, jeśli

$$\forall x, y \geq 0 \quad N_1(x) \text{ i } N_2(y) \text{ są niezależne.}$$

Twierdzenie 37. Jeśli $\{N_1(t) : t \geq 0\}, \{N_2(t) : t \geq 0\}$ są niezależnymi procesami Poissona z parametrami λ_1 i λ_2 , to $\{N_1(t) + N_2(t) : t \geq 0\}$ jest procesem Poissona z parametrem $\lambda_1 + \lambda_2$. Do tego każde zdarzenie pochodzi z N_1 z prawdopodobieństwem $\frac{\lambda_1}{\lambda_1 + \lambda_2}$.

Dowód. Dowodzimy w oparciu o charakteryzację procesów Poissona przez zmienne Poissona. Mamy $N_1(0) + N_2(0) = 0 + 0 = 0$. Niech $N(x) = N_1(x) + N_2(x)$. Dla pokazania niezależności przyrostów bierzemy dowolne $t_1 < t_2 \leq t_3 < t_4$ i rozpisujemy

$$\begin{aligned} N(t_2) - N(t_1) &= N_1(t_2) - N_1(t_1) + N_2(t_2) - N_2(t_1) \\ N(t_4) - N(t_3) &= N_1(t_4) - N_1(t_3) + N_2(t_4) - N_2(t_3). \end{aligned}$$

Wiemy, że składniki związane z tymi samymi procesami są niezależne, bo N_1 i N_2 to procesy Poissona. Niezależność „na krzyż” wynika z założonej niezależności procesów.

$N(t)$ ma rozkład Poissona z parametrem $t(\lambda_1 + \lambda_2)$, bo to suma niezależnych Poissonów z odpowiednimi parametrami.

Druga ostatnia część twierdzenia wynika z charakteryzacji przez międzyczasy. W dowolnym momencie czasu prawdopodobieństwo, że międzyczas pierwszego procesu skończy się przed międzyczasem drugiego wynosi właśnie $\frac{\lambda_1}{\lambda_1 + \lambda_2}$ – niezależnie od stanu procesu, bo zmienne są bez pamięci. $\square$

Twierdzenie 38. Niech $\{N(t) : t \geq 0\}$ będzie procesem Poissona z parametrem λ . Każde zdarzenie tego procesu będziemy niezależnie etykietować jako 1 z prawdopodobieństwem p lub 2 z prawdopo-

dobieństwem $1-p$. Niech $\{N_1(t) : t \geq 0\}$ i $\{N_2(t) : t \geq 0\}$ będą procesami zliczającymi odpowiednio zdarzenia etykietowane 1 i 2. Wtedy są one niezależnymi procesami Poissona z parametrami odpowiednio $p\lambda$ i $(1-p)\lambda$.

Dowód. Pokażemy tezę dla N_1 , dla N_2 symetrycznie. Mamy $N_1(0) = 0$, niezależność przyrostów wynika z niezależności w oryginalnym procesie i niezależności wyborów. Wystarczy przeliczyć wartość

$$\begin{aligned} P(N_1(t) = k) &= \sum_{j=k}^{\infty} P(N_1(t) = k \mid N(t) = j) \cdot P(N(t) = j) = \sum_{j=k}^{\infty} \binom{j}{k} p^k (1-p)^{j-k} \frac{e^{-\lambda t} (\lambda t)^j}{j!} \\ &= \frac{e^{-\lambda t} (p\lambda t)^k}{k!} \sum_{j=k}^{\infty} \frac{((1-p)\lambda t)^{j-k}}{(j-k)!} = \frac{e^{-\lambda t} (p\lambda t)^k}{k!} e^{(1-p)\lambda t} = \frac{e^{-p\lambda t} (p\lambda t)^k}{k!}. \end{aligned}$$

Dla pokazania niezależności przeliczamy dla ustalonego czasu t :

$$\begin{aligned} P(N_1(t) = m \cap N_2(t) = n) &= P(N(t) = m+n \cap N_2(t) = n) \\ &= e^{-\lambda t} \frac{(\lambda t)^{m+n}}{(m+n)!} \cdot \binom{m+n}{n} p^m (1-p)^n = \frac{e^{-p\lambda t} (p\lambda t)^m}{m!} \cdot \frac{e^{-(1-p)\lambda t} ((1-p)\lambda t)^n}{n!} \\ &= P(N_1(t) = m) \cdot P(N_2(t) = n). \end{aligned}$$

Pozostaje nam zauważyć, że dla $y > x$ zmienne $N_1(x)$ i $N_2(y)$ też są niezależne, bo na odcinku $[0, x]$ są niezależne, a na $[x, y]$ niezależność wynika z niezależności przyrostów wyjściowego procesu i niezależności etykietowania. $\square$

Lemat 19. Niech X_1 będzie pierwszym międzyczasem procesu Poissona N z parametrem λ . Zmienna $X_1 \mid N(t) = 1$ ma rozkład jednostajny na $[0, t]$.

Dowód.

$$\begin{aligned} P(X_1 < s \mid N(t) = 1) &= \frac{P(X_1 < s \cap N(t) = 1)}{P(N(t) = 1)} = \frac{P(N(s) = 1) \cdot P(N(t) - N(s) = 0)}{P(N(t) = 1)} \\ &= \frac{e^{-\lambda s} \lambda s \cdot e^{-\lambda(t-s)}}{e^{-\lambda t} \lambda t} = \frac{s}{t}. \end{aligned}$$

$\square$

Twierdzenie 39. Niech $\{N(t) : t \geq 0\}$ będzie procesem Poissona z parametrem λ . Niech T_i będzie czasem przyjścia i -tego zdarzenia. Przy warunku $N(t) = n$ rozkład $(T_1, \dots, T_n)$ jest taki sam jak sort $(X_1, \dots, X_n)$, gdzie zmienne $X_1, \dots, X_n$ mają rozkład jednostajny na $[0, t]$ i są niezależne.

Dowód. Oznaczmy $(Y_1, \dots, Y_n) = \text{sort}(X_1, \dots, X_n)$. Niech $(i_1, \dots, i_n)$ będzie permutacją $[n]$. Zauważmy, że zdarzenia postaci

$$X_{i_1} \leq X_{i_2} \leq \dots \leq X_{i_n} \cap X_{i_1} \leq s_1 \cap \dots \cap X_{i_n} \leq s_n$$

są rozłączne dla różnych permutacji (z dokładnością do zbioru miary 0 – może być tak, że dwie permutacje pasują do naszej sytuacji, gdy dwie zmienne przyjęły tę samą wartość). Do tego wszystkie są równie prawdopodobne. Zatem mamy

$$\begin{aligned} P((Y_1, \dots, Y_n) \leq (s_1, \dots, s_n)) &= \sum_{(i_1, \dots, i_n) \in S_n} P(X_{i_1} \leq \dots \leq X_{i_n} \cap X_{i_1} \leq s_1 \cap \dots \cap X_{i_n} \leq s_n) \\ &= n! P(X_1 \leq \dots \leq X_n \cap (X_1, \dots, X_n) \leq (s_1, \dots, s_n)) = n! \int_{u_1=0}^{s_1} \dots \int_{u_n=u_{n-1}}^{s_n} \left(\frac{1}{t}\right)^n du_n \dots du_1 \\ &= \frac{n!}{t^n} \int_{u_1=0}^{s_1} \dots \int_{u_n=u_{n-1}}^{s_n} du_n \dots du_1. \end{aligned}$$

Teraz musimy policzyć odpowiednią wartość dla czasów przyjścia. Niech Z_i oznacza i -ty międzyczas.

Mamy

$$\begin{aligned}
 P((T_1, \dots, T_n) \leq (s_1, \dots, s_n) \cap N(t) = n) \\
 &= P\left(Z_1 \leq s_1 \cap Z_2 \leq s_2 - Z_1 \cap \dots \cap Z_n \leq s_n - \sum_{j=1}^{n-1} Z_j \cap Z_{n+1} > t - \sum_{j=1}^n Z_j\right) \\
 &= \int_{z_1=0}^{s_1} \dots \int_{z_n=0}^{s_n - \sum_{j=1}^{n-1} z_j} \int_{z_{n+1}=t - \sum_{j=1}^n z_j}^{\infty} \lambda^{n+1} e^{-\lambda \sum_{j=1}^{n+1} z_j} dz_{n+1} \dots dz_1 \\
 &= \lambda^n e^{-\lambda t} \int_{z_1=0}^{s_1} \dots \int_{z_n=0}^{s_n - \sum_{j=1}^{n-1} z_j} dz_n \dots dz_1 = \lambda^n e^{-\lambda t} \int_{u_1=0}^{s_1} \dots \int_{u_n=u_{n-1}}^{s_n} du_n \dots du_1,
 \end{aligned}$$

gdzie trzecie przejście jest policzeniem najbardziej wewnętrznej całki, a później podstawiamy $u_i = \sum_{j=1}^i z_j$ (całkujemy funkcję stałą, więc znaczenie ma tak naprawdę tylko długość przedziału).

Wiemy, że $P(N(t) = n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!}$, więc prawdopodobieństwo warunkowe będzie takie, jakie ma być. $\square$

14. Rozkład normalny

2025-01-09

Definicja 68. Mówimy, że ciągła zmienna losowa X ma standardowy rozkład normalny (z parametrami 0 i 1), jeśli jej gęstość to funkcja $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Taki rozkład oznaczamy $N(0, 1)$.

Uwaga. Jeśli $X \sim N(0, 1)$, to $-X \sim N(0, 1)$.

Lemat 20. Zachodzi

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Dowód. Mamy

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \cdot \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{z^2+x^2}{2}} dz dx = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta = \int_0^{2\pi} 1 d\theta = 2\pi,$$

gdzie druga równość to przejście na współrzędne biegunowe a później całkujemy przez podstawienie. To daje nam tezę. $\square$

Lemat 21. Wartość oczekiwana i wariancja rozkładu $N(0, 1)$ to kolejno 0 i 1.

Dowód. Mamy

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} = 0.$$

Drugą całkę liczymy przez części:

$$\begin{aligned}
 \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \left(x \cdot \left(-e^{-\frac{x^2}{2}} \right) \Big|_0^{\infty} + \int_0^{\infty} e^{-\frac{x^2}{2}} dx \right) \\
 &= \frac{2}{\sqrt{2\pi}} \cdot \frac{\sqrt{2\pi}}{2} = 1.
 \end{aligned}$$

$\square$

Definicja 69. Definiujemy funkcję Phi jako

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz,$$

czyli dystrybuantę standardowego rozkładu normalnego.

Uwaga. Zachodzi $\Phi(-a) = 1 - \Phi(a)$.

Definicja 70. Zmienna losowa X ma rozkład normalny $N(\mu, \sigma^2)$ dla $\mu \in \mathbb{R}$ i $\sigma > 0$, jeśli istnieje $Z \sim N(0, 1)$ takie, że $X = \mu + \sigma Z$.

Lemat 22. Jeśli $X \sim N(\mu, \sigma^2)$, to zachodzi $E[X] = \mu$ i $\text{Var}(X) = \sigma^2$.

Dowód.

$$\begin{aligned} E[X] &= E[\mu + \sigma Z] = \mu + \sigma E[Z] = \mu. \\ \text{Var}(X) &= \text{Var}(\mu + \sigma Z) = \sigma^2 \text{Var}(Z) = \sigma^2. \end{aligned}$$

□

Lemat 23. Niech $X \sim N(\mu, \sigma^2)$. Dystrybuanta i gęstość X to

$$\begin{aligned} F_X(x) &= \Phi\left(\frac{x - \mu}{\sigma}\right), \\ f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}. \end{aligned}$$

Dowód.

$$\begin{aligned} F_X(x) &= P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right). \\ f_X(x) &= \Phi'\left(\frac{x - \mu}{\sigma}\right) \frac{1}{\sigma} = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}. \end{aligned}$$

□

Uwaga. Dla $X \sim N(\mu, \sigma^2)$ zachodzi

$$\begin{aligned} P(|X - \mu| \leq \sigma) &\approx 0,68 \\ P(|X - \mu| \leq 2\sigma) &\approx 0,95 \\ P(|X - \mu| \leq 3\sigma) &\approx 0,997 \end{aligned}$$

Lemat 24. Funkcja tworząca momentów rozkładu normalnego $N(\mu, \sigma^2)$ to

$$M_X(t) = e^{\frac{t^2\sigma^2}{2} + t\mu}.$$

Dowód.

$$\begin{aligned} M_X(t) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{tx} \cdot e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\sigma u + \mu)} \cdot e^{-\frac{u^2}{2}} du \\ &= \frac{1}{\sqrt{2\pi}} e^{t\mu} \int_{-\infty}^{\infty} e^{-\frac{(u - t\sigma)^2}{2}} \cdot e^{\frac{t^2\sigma^2}{2}} du = e^{t\mu + \frac{t^2\sigma^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(u - t\sigma)^2}{2}} du = e^{t\mu + \frac{t^2\sigma^2}{2}}, \end{aligned}$$

gdzie drugie przejście to podstawienie $u = \frac{x - \mu}{\sigma}$ a ostatnie to zauważenie, że ta całka to dystrybuanta rozkładu normalnego $N(t\sigma, 1)$. □

Przykład. Mamy

$$\begin{aligned} M'_X(t) &= (\mu + t\sigma^2) e^{\frac{t^2\sigma^2}{2} + t\mu} \implies E[X] = M'_X(0) = \mu \\ M''_X(t) &= (\mu + t\sigma^2) e^{\frac{t^2\sigma^2}{2} + t\mu} + \sigma^2 e^{\frac{t^2\sigma^2}{2} + t\mu} \implies E[X^2] = \mu^2 + \sigma^2 \end{aligned}$$

Twierdzenie 40. Niech $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ będą niezależne. Wówczas

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Dowód.

$$M_{X+Y}(t) = e^{\frac{t^2\sigma_1^2}{2} + t\mu_1 + \frac{t^2\sigma_2^2}{2} + t\mu_2} = e^{\frac{t^2(\sigma_1^2 + \sigma_2^2)}{2} + t(\mu_1 + \mu_2)},$$

a więc $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. $\square$

Twierdzenie 41 (Nierówność Czernowa). Niech $X \sim N(\mu, \sigma^2)$. Wówczas

$$P\left(\left|\frac{X - \mu}{\sigma}\right| \geq a\right) \leq 2e^{-\frac{a^2}{2}}.$$

Dowód. Zmienna $Z = \frac{X - \mu}{\sigma}$ ma standardowy rozkład normalny. Dla $t \geq 0$ mamy

$$P(Z \geq a) = P(e^{tZ} \geq e^{at}) \leq \frac{E[e^{tZ}]}{e^{at}} = e^{\frac{t^2}{2} - at} \leq e^{-\frac{a^2}{2}},$$

gdzie ostatnie przejście to podstawienie $t = a$.

$$P(Z \leq -a) = P(-Z \geq a) \leq e^{-\frac{a^2}{2}},$$

bo $-Z$ ma taki sam rozkład jak Z . $\square$

Definicja 71. Mówimy, że ciąg dystrybuant $F_1, F_2, \dots$ zbiega do dystrybuanty F (względem rozkładu), jeżeli

$$\forall a \in \mathbb{R} \quad F \text{ jest ciągłe w } a \implies \lim_{n \rightarrow \infty} F_n(a) = F(a).$$

Oznaczamy to $F_n \xrightarrow{D} F$ (zbieżność punktowa).

Twierdzenie 42 (Lévy'iego o ciągłości). Niech $Y_1, Y_2, \dots$ będzie ciągiem zmiennych losowych gdzie Y_i ma dystrybuantę F_i i funkcję tworzącą momentów M_i . Niech Y będzie zmienną losową o dystrybuanie F i funkcji tworzącej momentów M . Jeżeli

$$\forall t \quad \lim_{n \rightarrow \infty} M_n(t) = M(t),$$

to $F_n \xrightarrow{D} F$.

Twierdzenie 43 (Centralne Twierdzenie Graniczne). Niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi o takich samych rozkładach z wartością oczekiwaną μ i wariancją σ^2 . Dalej połóżmy $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Wówczas dla dowolnych a i b zachodzi

$$P\left(a \leq \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq b\right) \xrightarrow{D} \Phi(b) - \Phi(a).$$

Dowód. Niech $Z_i = \frac{X_i - \mu}{\sigma}$. Wówczas Z_i mają takie same rozkłady i $E[Z_i] = 0$, $\text{Var}(Z_i) = 1$. Rozważmy zmienną losową $Y_n = \sum_{i=1}^n \frac{Z_i}{\sqrt{n}} = \frac{\sqrt{n}}{n} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$. Chcemy pokazać, że

$$\lim_{n \rightarrow \infty} E[e^{tY_n}] = e^{\frac{t^2}{2}},$$

co da nam tezę z twierdzenia Lévy'iego. Oznaczając $M(t) = E[e^{tZ_i}]$ mamy

$$E[e^{tY_n}] = E\left[e^{\frac{t}{\sqrt{n}} \sum_{i=1}^n Z_i}\right] = \left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n.$$

Definiujemy $L(t) = \ln M(t)$. Przeliczamy:

1. $L(0) = \ln M(0) = \ln 1 = 0$, bo $M(0) = E[e^{0 \cdot Z_i}] = 1$
2. $L'(0) = \frac{M'(0)}{M(0)} = E[Z_i] = 0$
3. $L''(0) = \frac{M(0)M''(0) - (M'(0))^2}{(M(0))^2} = E[Z_i^2] = 1$

Z ciągłości logarytmu wystarczy pokazać, że

$$nL\left(\frac{t}{\sqrt{n}}\right) \rightarrow \frac{t^2}{2}.$$

Korzystając dwukrotnie z reguły de l'Hospitala dostajemy

$$\lim_{n \rightarrow \infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} = \lim_{n \rightarrow \infty} \frac{-L'\left(\frac{t}{\sqrt{n}}\right) n^{-\frac{3}{2}}}{-2n^{-2}} = \lim_{n \rightarrow \infty} \frac{-L''\left(\frac{t}{\sqrt{n}}\right) n^{-\frac{3}{2}} t^2}{-2n^{-\frac{3}{2}}} = \lim_{n \rightarrow \infty} L''\left(\frac{t}{\sqrt{n}}\right) \frac{t^2}{2} = \frac{t^2}{2}.$$

□

Twierdzenie 44. Niech $X_1, \dots, X_n$ będzie ciągiem niezależnych zmiennych losowych z $E[X_i] = \mu_i$ i $\text{Var}(X_i) = \sigma_i^2$. Niech zachodzi

1. $\exists_{M>0} \forall_{i \in [n]} P(|X_i| < M) = 1$
2. $\lim_{n \rightarrow \infty} \sum_{i=1}^n \sigma_i^2 = +\infty$.

Wówczas

$$P\left(a \leq \frac{\sum_{i=1}^n (X_i - \mu_i)}{\sqrt{\sum_{i=1}^n \sigma_i^2}} \leq b\right) \xrightarrow{D} \Phi(b) - \Phi(a).$$

Twierdzenie 45 (Berry-Esséen). Istnieje taka stała C , że zachodzi następujące: niech $X_1, \dots, X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie ze skończoną wartością oczekiwaną μ i wariancją σ^2 . Dalej niech $\rho = E[|X_i - \mu|^3] < \infty$ i $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Mamy

$$\left| P\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq a\right) - \Phi(a) \right| \leq C \cdot \frac{\rho}{\sigma^3 \sqrt{n}}.$$

Przykład (Badanie opinii publicznej). Chcemy oszacować poparcie dla jakiejś partii. Zakładamy, że każda osoba niezależnie i z takim samym rozkładem ją popiera lub nie. Zakładamy, że liczność próbki n jest na tyle duża, że możemy stosować CTG. Chcemy znaleźć minimalną wartość n potrzebną do zagwarantowania, że oszacowany przez nas ułamek poparcia $\tilde{p}$ odbiega od prawdziwej wartości p o co najwyżej 0,05 z prawdopodobieństwem 0,95.

Dowód. Niech $X_1, \dots, X_n$ będą indykatorami poparcia partii przez kolejne osoby. Mamy $E[X_i] = p$ oraz $\text{Var}(X_i) = p(1-p)$. Niech $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. To jest właśnie wartość $\tilde{p}$. Definiujemy

$$Z_n = \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}},$$

co z CTG ma rozkład normalny. Teraz wystarczy przeliczyć

$$\begin{aligned} P(|\bar{X}_n - p| \geq \delta) &= P\left(\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \geq \frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right) + P\left(\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \leq -\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right) \\ &= 2\left(1 - \Phi\left(\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right)\right) \leq 0,05. \end{aligned}$$

To daje nam

$$\Phi\left(\frac{\sqrt{n}\delta}{\sqrt{p(1-p)}}\right) \geq 0,975 = \Phi(1,96) \implies n \geq \left(\frac{1,96\sqrt{p(1-p)}}{\delta}\right)^2,$$

bo Φ jest monotoniczne. Teraz wystarczy podstawić wartość δ i skorzystać z $p(1-p) \leq \frac{1}{4}$ żeby dostać $n \geq 385$.

15. Sprzęganie łańcuchów Markowa

2025-01-16

Pomysł. Wiemy, kiedy łańcuch Markowa ma rozkład stacjonarny. Pytamy się, jak szybko łańcuch zbiega do tego rozkładu.

Przykład. Tasujemy n kart – wybieramy kartę jednostajnie, niezależnie i kładziemy ją na wierzchu talii. Stan talii (permutacja kart) tworzy łańcuch Markowa. Taki łańcuch spełnia warunki na posiadanie rozkładu stacjonarnego. Możemy przeliczyć, że rozkład stacjonarny π jest rozkładem jednostajnym. Zatem takie tasowanie ma sens – w granicy każda permutacja jest tak samo prawdopodobna.

Definicja 72. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S . Normą całkowitego wahan (total variation distance) tych rozkładów nazywamy wartość

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq S} |\mu(A) - \nu(A)|.$$

Propozycja 26. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S . Niech $B = \{x \in S : \mu(x) \geq \nu(x)\}$. Zachodzi

$$\|\mu - \nu\|_{TV} = \mu(B) - \nu(B) = \nu(B^c) - \mu(B^c).$$

Dowód. Niech $A \subseteq S$. Zachodzi $\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B)$, gdzie pierwsza nierówność to wzięcie tylko tych elementów A dla których różnica rozkładów jest dodatnia a potem rozszerzamy się na wszystkie takie elementy. Analogicznie mamy $\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c) = 1 - \nu(B) - 1 + \mu(B)$. Zatem te wartości są równe i świadczą o maksimum. $\square$

Propozycja 27. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S . Zachodzi

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|.$$

Dowód. Z poprzedniej propozycji dostajemy, że dla $B = \{x \in S : \mu(x) \geq \nu(x)\}$ jest

$$\|\mu - \nu\|_{TV} = \frac{1}{2} (\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)) = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|.$$

$\square$

Propozycja 28. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S . Zachodzi

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup \left\{ \sum_{x \in S} f(x) \mu(x) - \sum_{x \in S} f(x) \nu(x) : \max_{x \in S} |f(x)| \leq 1 \right\}.$$

Dowód. ($\geq$) Jeśli $\max_{x \in S} |f(x)| \leq 1$, to mamy

$$\begin{aligned} \frac{1}{2} \left| \sum_{x \in S} f(x) \mu(x) - \sum_{x \in S} f(x) \nu(x) \right| &\leq \frac{1}{2} \sum_{x \in S} |f(x) (\mu(x) - \nu(x))| \\ &\leq \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)| = \|\mu - \nu\|_{TV}. \end{aligned}$$

($\leq$) Bierzemy $B = \{x \in S : \mu(x) \geq \nu(x)\}$ i definiujemy funkcję

$$f^*(x) = \begin{cases} 1, & x \in B \\ -1, & x \in B^c \end{cases},$$

która daje

$$\frac{1}{2} \left(\sum_{x \in S} f^*(x) \mu(x) - \sum_{x \in S} f^*(x) \nu(x) \right) = \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)| = \|\mu - \nu\|_{TV}.$$

□

Przykład. Wracając do przykładu z tasowaniem kart: biorąc obecny rozkład D taki, że $\|D - \pi\|_{TV} < \varepsilon$ dostajemy dla $A = \{x\}$

$$\varepsilon > \max_{A \subseteq S} |D(A) - \pi(A)| \geq \left| D(x) - \frac{1}{n!} \right|.$$

A więc D różni się od jednostajnego o co najwyżej ε .

Notacja. Dalej będziemy rozważać łańcuch Markowa $(X_t)_{t \in \mathbb{N}}$ (skończony, nieprzywiedlny, nieokresowy) o macierzy przejścia P , zbiorze stanów S i rozkładzie stacjonarnym $(\pi_x)_{x \in S}$. Przez $P^t(x, \cdot)$ oznaczamy rozkład X_t przy założeniu $X_0 = x$.

Definicja 73. Definiujemy

$$\begin{aligned} \Delta_x(t) &= \|P^t(x, \cdot) - \pi\|_{TV}, \\ \tau_x(\varepsilon) &= \min \{t : \Delta_x(t) \leq \varepsilon\}. \end{aligned}$$

Mamy też maksima tych wartości:

$$\begin{aligned} \Delta(t) &= \max_{x \in S} \Delta_x(t), \\ \tau_{\text{mix}}(\varepsilon) &= \max_{x \in S} \tau_x(\varepsilon). \end{aligned}$$

Ostatnią z tych wartości nazywamy czasem mieszania łańcucha Markowa. Będziemy też (bez większego powodu) oznaczać $\tau_{\text{mix}} = \tau_{\text{mix}}(\frac{1}{4})$.

Definicja 74. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S . Sprzęganiem μ i ν nazywamy dowolną parę zmiennych losowych (X, Y) taką, że X ma rozkład μ , a Y ma rozkład ν . W szczególności te zmienne nie muszą być niezależne.

Propozycja 29. Niech (X, Y) będzie sprzęganiem μ i ν . Zachodzi

$$\|\mu - \nu\|_{TV} \leq P(X \neq Y).$$

Ponadto istnieje sprzęganie dla którego zachodzi równość.

Dowód. Dla dowolnego $A \subseteq S$ mamy

$$\mu(A) - \nu(A) = P(X \in A) - P(Y \in A) \leq P(X \in A \cap Y \notin A) \leq P(X \neq Y).$$

Analogicznie $\nu(A) - \mu(A) \leq P(X \neq Y)$. To daje żadaną nierówność.

Teraz skonstruujemy sprzęganie spełniające równość. Niech $B = \{x \in S : \mu(x) \geq \nu(x)\}$. Niech $p_1 = \mu(B) - \nu(B)$, $p_2 = \nu(B^c) - \mu(B^c)$. Mamy $p_1 = p_2 = \|\mu - \nu\|_{TV}$. Niech $p_3 = 1 - p_1 = 1 - p_2$.

Rzucamy monetą z prawdopodobieństwem orła p_3 . Jeśli wypadnie orzeł to ustalamy $X = Y = s$, gdzie s wybieramy z S z rozkładem $\left(\frac{1}{p_3} \min(\mu(s), \nu(s)) : s \in S\right)$. Jeśli wypadnie reszka ustalamy $X = x$ i $Y = y$, gdzie x jest wybierany losowo z S z rozkładem $\left(\frac{1}{p_1} \max(\mu(x) - \nu(x), 0) : x \in S\right)$, a y z rozkładem $\left(\frac{1}{p_2} \max(\nu(x) - \mu(x), 0) : x \in S\right)$. W przypadku reszki jedna zmienna przyjmuje tylko te wartości, na których μ jest większe, a druga tylko te, na których ν jest większe. Mamy więc $P(X \neq Y) = 1 - p_3 = \|\mu - \nu\|_{TV}$, a (X, Y) faktycznie jest sprzęganiem μ i ν – zmienne mają odpowiednie rozkłady. $\square$

Definicja 75. Sprzęganiem łańcuchów Markowa X, Y o macierzy przejścia P i zbiorze stanów S jest dowolny łańcuch Markowa $(Z_t = (X_t, Y_t))_{t \in \mathbb{N}}$ na przestrzeni stanów $S \times S$ taki, że

$$P(X_{t+1} = x' \mid Z_t = (x, y)) = P(x, x')$$

$$P(Y_{t+1} = y' \mid Z_t = (x, y)) = P(y, y')$$

dla każdego $t \geq 0, x, y, x', y' \in S$.

Sprzęgane łańcuchy to dwie równoległe kopie jednego procesu. Nie zawsze mają one te same stany, ale też nie zawsze są niezależne. Nie ustalamy nic o stanach początkowych. Będą nas interesować takie sprzęgania, które sprowadzają obie kopie do tego samego stanu i potem je tak utrzymują.

Lemat 25. Niech $((X_t, Y_t))_{t \in \mathbb{N}}$ będzie sprzęganiem łańcuchów (skończonych, nieprzywiedlnych, nieokresowych) z macierzą przejścia P i zbiorem stanów S . Niech $T \in \mathbb{N}$ i $\varepsilon > 0$ będą takie, że dla każdego $x, y \in S$ zachodzi

$$P(X_T \neq Y_T \mid X_0 = x, Y_0 = y) \leq \varepsilon.$$

Wtedy czas mieszania łańcucha z macierzą P jest ograniczony:

$$\forall_{x \in S} \Delta_x(T) \leq \varepsilon$$

$$\tau_{\text{mix}}(\varepsilon) \leq T.$$

Dowód. Zauważmy, że sprzęganie spełnia założenia niezależnie od tego, w jaki sposób ustalimy X_0 i Y_0 . Ustalmy dowolne $x \in S$. Niech $X_0 = x$ i niech Y_0 będzie wybrany losowo z rozkładu stacjonarnego π . Wtedy Y_t ma rozkład π dla każdego t .

Niech $A \subseteq S$. Mamy

$$\begin{aligned} P(X_T \in A) &\geq P(Y_T \in A \cap X_T = Y_T) = 1 - P(Y_T \notin A \cup X_T \neq Y_T) \\ &\geq 1 - P(Y_T \notin A) - P(X_T \neq Y_T) \geq P(Y_T \in A) - \varepsilon = \pi(A) - \varepsilon. \end{aligned}$$

Analogicznie $P(X_T \in A^c) \geq \pi(A^c) - \varepsilon$, czyli $P(X_T \in A) \leq \pi(A) + \varepsilon$.

Mamy zatem

$$\forall_{x \in S} \Delta_x(T) = \max_{A \subseteq S} |P^T(x, A) - \pi(A)| \leq \varepsilon,$$

a z tego wynika

$$\tau_{\text{mix}}(\varepsilon) \leq T.$$

$\square$

Przykład. Wracamy do przykładu z kartami. Niech P będzie macierzą przejścia dla tasowania. X_t to oryginalny łańcuch Markowa, a Y_t definiujemy tak: mamy ustalony stan początkowy y i jeśli X_t powstało z X_{t-1} przez przełożenie karty C na szczyt, to Y_t powstaje z Y_{t-1} przez przełożenie tej samej karty. W takiej sytuacji $((X_t, Y_t))_{t \in \mathbb{N}}$ jest sprzęganiem.

Jeśli karta C została położona na szczyt w pewnym kroku, to potem będzie w tym samym miejscu w obu łańcuchach. Zatem liczba kroków T po której łańcuchy się schodzą ma rozkład jak kolekcjoner kuponów. Prawdopodobieństwo, że karta C nie została przełożona po $n \ln n + cn$ krokach to $(1 - \frac{1}{n})^{n \ln n + cn} \leq e^{-(\ln n + c)} = \frac{e^{-c}}{n}$. Zatem z union bounda łańcuchy nie zejść się z prawdopodobieństwem co najwyżej e^{-c} .

Lemat 26 (O monotoniczności). Niech P będzie macierzą przejścia skończonego, nieprzywiedlnego i nieokresowego łańcucha Markowa ze zbiorem stanów S i rozkładem stacjonarnym π . Dla każdego $t \geq 0, x \in S$ zachodzi

$$\Delta_x(t+1) \leq \Delta_x(t).$$

Dowód. Ustalmy $t \geq 0$ i $x \in S$. Niech (X_t, Y_t) będzie sprzężaniem rozkładów $P^t(x, \cdot)$ i π spełniającym $P(X_t \neq Y_t) = \Delta_x(t)$ (przedtem pokazaliśmy, że istnieje sprzężanie, dla którego ta równość zachodzi). Definiujemy (X_{t+1}, Y_{t+1}) w następujący sposób: jeśli $X_t = Y_t$, wykonujemy krok łańcucha zgodnie z macierzą P (na obu współrzędnych taki sam), a w przeciwnym wypadku wykonujemy dwa niezależne kroki. Zauważmy, że Y_{t+1} dalej ma rozkład π . Mamy

$$\Delta_x(t) = P(X_t \neq Y_t) \geq P(X_{t+1} \neq Y_{t+1}) \geq \|P^{t+1}(x, \cdot) - \pi\|_{TV} = \Delta_x(t+1).$$

□

Twierdzenie 46 (O geometrycznej zbieżności). Niech P będzie macierzą przejścia skończonego, nieprzywiedlnego i nieokresowego łańcucha Markowa ze zbiorem stanów S i rozkładem stacjonarnym π . Wtedy istnieją $\alpha \in (0, 1)$ i $C > 0$ takie, że

$$\forall n \in \mathbb{N} \quad \Delta(n) \leq C\alpha^n.$$

Dowód. Ustalmy $r \geq 1$ takie, że dla każdego $x, y \in S$ jest $P^r(x, y) > 0$ (dla konkretnych dwóch istnieje, bo łańcuch jest nieokresowy i nieprzywiedlny, a ze skończoności można wziąć maksimum).

Niech $m_y = \min_{x \in S} P^r(x, y)$ dla $y \in S$. Jest to najmniejsze z prawdopodobieństw, z jakimi da się przejść do y krokiem macierzy P^r . Niech $m = \sum_{y \in S} m_y \leq 1$ (ta suma ogranicza z dołu dowolny wiersz, a wiersz sumuje się do 1).

Niech $((X_t, Y_t))_{t \in \mathbb{N}}$ będzie sprzężaniem łańcuchów o macierzy przejścia P^r zadany w następujący sposób: mając zadane $X_t = x, Y_t = y$ (początkowe wartości wybieramy dowolnie) oznaczamy $\mu = P^r(x, \cdot), \nu = P^r(y, \cdot)$ i $B = \{x \in S : \mu(x) \geq \nu(x)\}$. Niech $p_1 = \mu(B) - \nu(B), p_2 = \nu(B^c) - \mu(B^c)$ i $p_3 = 1 - p_1 = 1 - p_2$.

Rzucamy monetą z prawdopodobieństwem orła p_3 . Jeśli wypadnie orzeł, to ustalamy $X_{t+1} = Y_{t+1} = s$, gdzie s wybieramy z S z rozkładem $(\frac{1}{p_3} \min(\mu(s), \nu(s)) : s \in S)$. Jeśli wypadnie reszka ustalamy $X_{t+1} = x$ i $Y_{t+1} = y$, gdzie x wybieramy losowo z S z rozkładem $(\frac{1}{p_1} \max(\mu(x) - \nu(x), 0) : x \in S)$, a y z rozkładem $(\frac{1}{p_2} \max(\nu(x) - \mu(x), 0) : x \in S)$.

W ten sposób skonstruowaliśmy sprzężanie, dla którego zachodzi

$$\forall t \geq 0, y \in S \quad P(X_{t+1} = Y_{t+1} = y) \geq p_3 \cdot \frac{1}{p_3} \min(\mu(y), \nu(y)) \geq m_y.$$

Mamy więc

$$\forall t \geq 1 \quad P(X_t = Y_t) = \sum_{y \in S} P(X_t = Y_t = y) \geq \sum_{y \in S} m_y = m,$$

a z tego wynika $P(X_t \neq Y_t) \leq (1 - m)^t$. Teraz mając zadane $n = rt + j$ dla $j \in \{0, \dots, r-1\}$ możemy zapisać

$$\Delta_x(n) \leq \Delta_x(rt) = \|P^{rt}(x, \cdot) - \pi\|_{TV} \leq P(X_t \neq Y_t) \leq (1 - m)^t = \alpha^{rt} \leq C\alpha^n,$$

gdzie położyliśmy $\alpha = (1 - m)^{\frac{1}{r}}$ i $C = \alpha^{-r}$.

□

Twierdzenie 47. Rozważmy graf G o n wierzchołkach i stopniu maksymalnym Δ . Definiujemy następujący łańcuch Markowa: niech M_0 będzie ustalonym k -elementowym zbiorem niezależnym w G . W kroku łańcucha losujemy wierzchołek v jednostajnie z M_t i w jednostajnie z $V(G)$. Ustalamy $M_{t+1} = M_t \setminus \{v\} \cup \{w\}$, jeśli taki zbiór jest niezależny i ma k wierzchołków. Inaczej $M_{t+1} = M_t$.

Jeśli $k \leq \frac{n}{3(\Delta+1)}$, to taki łańcuch jest skończony, nieprzywiedlny i nieokresowy a jego rozkład stacjonarny jest jednostajny po wszystkich k -elementowych zbiorach niezależnych. Do tego jego czas mieszania jest ograniczony: $\tau_{\text{mix}} \leq \ln(4k) \cdot \frac{kn}{n-3(k-1)(\Delta+1)}$.

Dowód. Nieokresowość wynika z tego, że łańcuch może stać w miejscu. Niech $A \neq B$ będą k -elementowymi zbiorami niezależnymi i $u \in B \setminus A$. Pokażemy, że z A da się dojść do zbioru zawierającego u bez usuwania żadnych wspólnych elementów A i B , co indukcyjnie pozwoli nam na dojście do B – w ten sposób pokażemy nieprzywiedlność.

Jeśli $|N(u) \cap A| = 0$, to przejście $A \rightarrow A \setminus \{v\} \cup \{u\}$ dla dowolnego $v \in A \setminus B$ jest tym, co chcemy.

Jeśli $N(u) \cap A = \{v\}$, to $v \notin B$ i można przejść $A \rightarrow A \setminus \{v\} \cup \{u\}$.

Jeśli $N(u) \cap A = \{v_1, \dots, v_\ell\}$, to $v_1 \notin B$. Możemy wyrzucić v_1 i zamiast niego wziąć dowolny element $W = V(G) \setminus \left(\bigcup_{w \in A \setminus \{v_1\}} N[w] \cup N[u] \right)$. Powtarzając ten proces aż do v_ℓ znajdziemy się w końcu w poprzednim przypadku i będziemy mogli dodać u do naszego zbioru. Pozostaje wykazać, że $W \neq \emptyset$. Mamy $|W| \geq n - (k-1)(\Delta+1) - (\Delta+1) + 1 \geq \frac{2}{3}n + 1 > 0$, gdzie jedynka wynika z tego, że u zostało policzone w sąsiedztwie swoim oraz v_2 .

Niech P będzie macierzą przejścia tego łańcucha. Mamy $P(x, y) = P(y, x)$, bo albo między stanami da się przejść wybierając odpowiedni wierzchołek ze zbioru i całego grafu (czyli z prawdopodobieństwem $\frac{1}{nk}$), albo się nie da i obie strony to 0. Z tego wynika, że dla rozkładu jednostajnego π jest $\pi(x)P(x, y) = \pi(y)P(y, x)$, a z tego wynika jego stacjonarność.

Zdefiniujemy sprzęganie $((X_t, Y_t))_{t \in \mathbb{N}}$ w następujący sposób: X_t jest już opisanym łańcuchem, czyli w kroku losujemy wierzchołki v, w . Jeśli $v \in Y_t$, to Y_t wykonuje krok dla wierzchołków v, w . W przeciwnym wypadku losujemy v' jednostajnie z $Y_t \setminus X_t$ i wykonujemy krok dla v', w . Wierzchołki należące do $Y_t \cap X_t$ mają szansę $\frac{1}{k}$ na bycie wybranym, a pozostałe $\frac{k - |Y_t \cap X_t|}{k} \cdot \frac{1}{k - |Y_t \cap X_t|} = \frac{1}{k}$, a więc drugi łańcuch też ma macierz przejścia P .

Zdefiniujmy $d_t = |X_t \setminus Y_t|$. Oczywiście $d_{t+1} \in \{d_t - 1, d_t, d_t + 1\}$. Pokażemy, że jeśli $d_t > 0$, to przejście -1 jest bardziej prawdopodobne od $+1$. Mamy

$$P(d_{t+1} = d_t + 1 \mid d_t > 0) \leq \frac{k - d_t}{k} \cdot \frac{2d_t(\Delta + 1)}{n}$$

$$P(d_{t+1} = d_t - 1 \mid d_t > 0) \geq \frac{d_t}{k} \cdot \frac{n - (k + d_t - 2)(\Delta + 1)}{n},$$

bo w pierwszym przypadku musimy wybrać wspólne v i takie w , że dokładnie jeden z łańcuchów zmieni stan – szacujemy z góry przez sumę sąsiedztw niewspólnych wierzchołków z obu łańcuchów (może być tak, że te sąsiedztwa się pokrywają i oba nie zmieniają stanu, ale to szacowanie z góry więc jest dobrze). W drugim przypadku musimy wybrać różne v, v' i takie w , że oba łańcuchy zmieniają stan – odejmujemy największe możliwe sąsiedztwa wszystkich wierzchołków w obu łańcuchach poza v i v' . Teraz zakładając $d_t > 0$ możemy przeliczyć

$$\begin{aligned} E[d_{t+1} \mid d_t] &= d_t - P(d_{t+1} = d_t - 1) + P(d_{t+1} = d_t + 1) \\ &\leq d_t - \frac{d_t}{k} \cdot \frac{n - (k + d_t - 2)(\Delta + 1)}{n} + \frac{k - d_t}{k} \cdot \frac{2d_t(\Delta + 1)}{n} = d_t \left(1 - \frac{n - (3k - d_t - 2)(\Delta + 1)}{kn} \right) \\ &\leq d_t \left(1 - \frac{n - 3(k-1)(\Delta + 1)}{kn} \right), \end{aligned}$$

gdzie druga nierówność to zastosowanie $d_t \geq 1$. Oznaczmy $C = \frac{n - 3(k-1)(\Delta + 1)}{kn}$.

Mamy $E[d_{t+1} \mid d_t = 0] = 0$, zatem

$$E[d_{t+1}] = E[E[d_{t+1} \mid d_t]] \leq E[d_t](1 - C) \leq d_0(1 - C)^{t+1} \leq d_0 e^{-(t+1)C} \leq k e^{-(t+1)C}.$$

Z nierówności Markowa

$$P(d_t \geq 1) \leq E[d_t] \leq ke^{-tC} \implies P(X_t \neq Y_t) \leq ke^{-tC}.$$

Dla $t \geq \ln(4k) \cdot \frac{1}{C}$ mamy $\tau_{\text{mix}}(\frac{1}{4}) \leq t$ (w tym momencie ważne jest $C > 0$, które wynika z założenia o wartości k). $\square$

Twierdzenie 48. Rozważmy graf G o n wierzchołkach i stopniu maksymalnym Δ . Definiujemy następujący łańcuch Markowa: niech M_0 będzie ustalonym poprawnym c -kolorowaniem wierzchołkowym G . W kroku łańcucha losujemy wierzchołek v jednostajnie z $V(G)$ i kolor ℓ jednostajnie z $[c]$. Jeśli $N(v)$ nie zawiera wierzchołków koloru ℓ , to zmieniamy kolor v na ℓ . Inaczej łańcuch stoi w miejscu.

Jeśli $c \geq 2\Delta + 1$, to taki łańcuch jest skończony, nieprzywiedlny i nieokresowy a jego rozkład stacjonarny jest jednostajny po wszystkich poprawnych c -kolorowaniach wierzchołkowych. Do tego jego czas mieszania jest ograniczony: $\tau_{\text{mix}} \leq \ln(4n) \cdot \frac{cn}{c-2\Delta}$.

Dowód. Nieokresowość wynika z tego, że łańcuch może stać w miejscu. Rozważmy kolorowania X, Y i pewien dowolny porządek na wierzchołkach. Możemy po kolei poprawiać kolory wierzchołków z X tak, żeby uzyskać Y . Jeśli pewien wierzchołek v nie może zmienić koloru, to z powodu jakiegoś v' , który jest dalej w tym porządku. Możemy zmienić kolor v' na jakiś niekonfliktujący, bo $c \geq \Delta + 2$. Robimy tak ze wszystkimi konfliktami i w końcu poprawiamy samo v . Przedstawiony proces dowodzi, że łańcuch jest nieprzywiedlny.

Niech P będzie macierzą przejścia tego łańcucha. Mamy $P(x, y) = P(y, x)$, bo albo między stanami da się przejść wybierając odpowiedni wierzchołek i kolor (czyli z prawdopodobieństwem $\frac{1}{nc}$), albo się nie da i obie strony to 0. Z tego wynika, że dla rozkładu jednostajnego π jest $\pi(x)P(x, y) = \pi(y)P(y, x)$, a z tego wynika jego stacjonarność.

Zdefiniujemy sprzęganie $((X_t, Y_t))_{t \in \mathbb{N}}$ w następujący sposób: wybierzmy jednostajnie wierzchołek v i kolor ℓ . Niech $D_t = \{v \in V(G) : X_t(v) \neq Y_t(v)\}$, $A_t = \{v \in V(G) : X_t(v) = Y_t(v)\}$. Jeśli $v \in D_t$, to oba X_t i Y_t wykonują krok z v, ℓ . W przeciwnym wypadku X_t dalej wykonuje taki sam krok, natomiast krok Y_t jest bardziej przemyślany. Definiujemy $S_X(v) = X_t(N(v)) \setminus Y_t(N(v))$ oraz $S_Y(v) = Y_t(N(v)) \setminus X_t(N(v))$. Są to zbiory kolorów, które występują na sąsiedztwie v tylko w jednym z kolorowań. Do tego definiujemy $B_X(v) = [c] \setminus X(N(v))$, $B_Y(v) = [c] \setminus Y(N(v))$.

Ustalmy porządki $S_X(v) = \{x_1, \dots, x_\alpha\}$, $S_Y(v) = \{y_1, \dots, y_\beta\}$, $B_X(v) = \{z_1, \dots, z_\gamma\}$, $B_Y(v) = \{w_1, \dots, w_\delta\}$. Zauważmy, że $X(N(v)) \setminus S_X(v) = Y(N(v)) \setminus S_Y(v)$.

Jeśli $\alpha \leq \beta$, to definiujemy bijekcję $f : [c] \mapsto [c]$ jako

$$f(t) = \begin{cases} t, & t \in X(N(v)) \setminus S_X(v) \\ y_i, & t = x_i \\ y_{i+\alpha}, & t = z_i, i + \alpha \leq \beta \\ w_{i-\beta+\alpha}, & t = z_i, i + \alpha > \beta \end{cases}.$$

Jeśli $\alpha > \beta$, to definiujemy ją jako

$$f(t) = \begin{cases} t, & t \in X(N(v)) \setminus S_X(v) \\ y_i, & t = x_i, i \leq \beta \\ w_{i-\beta}, & t = x_i, i > \beta \\ w_{i+\alpha-\beta}, & t = z_i \end{cases}.$$

Intuicyjnie znaczy to, że przypisujemy kolorom z $S_X(v)$ kolory z $S_Y(v)$ aż któryś zbiór się skończy. Y_t wykona krok z $v, f(\ell)$.

Oznaczmy $d_t = |D_t|$ i $d'(v) = \begin{cases} |N(v) \cap A_t|, & v \in D_t \\ |N(v) \cap D_t|, & v \in A_t \end{cases}$. Zauważmy, że mamy $\sum_{v \in A_t} d'(v) =$

$m' = \sum_{v \in D_t} d'(v)$, gdzie m' jest liczbą krawędzi między wierzchołkami D_t a A_t . Zachodzi

$$P(d_{t+1} = d_t - 1 \mid d_t > 0) \geq \frac{1}{n} \sum_{v \in D_t} \frac{c - 2\Delta + d'(v)}{c} = \frac{1}{cn} ((c - 2\Delta) d_t + m')$$

$$P(d_{t+1} = d_t + 1 \mid d_t > 0) \leq \frac{1}{n} \sum_{v \in A_t} \frac{\max(|S_X(v)|, |S_Y(v)|)}{c} \leq \frac{1}{cn} \sum_{v \in A_t} d'(v) = \frac{m'}{cn},$$

bo w pierwszym przypadku musimy wybrać wierzchołek, na którym kolorowania się nie zgadzają a potem wybrać kolor, który pasuje do obu kolorowań – od wszystkich kolorów odejmujemy maksymalną liczbę sąsiadów v (razy dwa, bo mogą mieć różne kolory w różnych kolorowaniach) i bierzemy poprawkę na kolory, które występują w obu kolorowaniach (zliczyliśmy je podwójnie). W drugim przypadku bierzemy wierzchołek, na którym kolorowania się zgadzają i kolor, który nie pasuje do dokładnie jednego kolorowania. W naszym sprzęganiu powiązaliśmy ze sobą te kolory, więc można je ograniczyć przez większą z wartości $|S_X(v)|, |S_Y(v)|$. Oczywiście jest $E[d_{t+1} \mid d_t = 0] = 0$, a więc pozostaje nam przeliczyć

$$E[d_{t+1} \mid d_t] \leq d_t - \frac{1}{cn} ((c - 2\Delta) d_t + m') + \frac{m'}{cn} = d_t \left(1 - \frac{c - 2\Delta}{cn}\right).$$

Zatem mamy

$$E[d_{t+1}] = E[E[d_{t+1} \mid d_t]] \leq E[d_t] \left(1 - \frac{c - 2\Delta}{cn}\right) \leq d_0 \left(1 - \frac{c - 2\Delta}{cn}\right)^{t+1} \leq n \left(1 - \frac{c - 2\Delta}{cn}\right)^{t+1},$$

czyli

$$P(d_t \geq 1) \leq E[d_t] \leq n \left(1 - \frac{c - 2\Delta}{cn}\right)^t \leq ne^{-t \frac{c - 2\Delta}{cn}},$$

co dla $t \geq \ln(4n) \cdot \frac{cn}{c - 2\Delta}$ daje $\tau_{\text{mix}} \leq t$ (korzystamy w tym momencie z $c > 2\Delta$). $\square$

16. Metoda Monte Carlo

2025-02-07

Definicja 76. Zmienna losowa X (randomizowany algorytm) daje (ε, δ) -aproksymację wartości V , jeśli

$$P(|X - V| \leq \varepsilon V) \geq 1 - \delta.$$

Przykład. Liczba π jest niewymierna, ale możemy ją przybliżać. Losujemy punkt (X, Y) w kwadracie o boku 2. Niech Z będzie indykátorem zdarzenia, że wylosowany punkt leży w kole wpisanym w ten kwadrat. Mamy $P(Z = 1) = \frac{\pi}{4}$. Powtarzamy taki eksperyment m razy i definiujemy $W = \sum_{i=1}^m Z_i$. Teraz $E[W] = \frac{m\pi}{4}$, a więc $W' = \frac{4}{m}W$ jest naturalnym oszacowaniem π . Z nierówności Czernowa dostajemy

$$P(|W' - \pi| \geq \varepsilon \pi) = P\left(\left|W - \frac{m}{4}\pi\right| \geq \varepsilon \frac{m}{4}\pi\right) \leq 2e^{-m\pi \frac{\varepsilon^2}{12}},$$

a więc dla $m \geq \frac{12 \ln(\frac{2}{\delta})}{\pi \varepsilon^2}$ W' jest (ε, δ) -aproksymacją π .

Definicja 77. Mając dany problem obliczeniowy $x \rightarrow R(x)$ mówimy, że mamy w pełni wielomianowy randomizowany schemat aproksymacji (FPRAS – fully polynomial randomized approximation scheme), jeśli dla wejścia x i parametrów ε, δ umiemy generować (ε, δ) -aproksymację $R(x)$ w czasie wielomianowym od $x, \frac{1}{\varepsilon}, \ln(\frac{1}{\delta})$.

Algorytm 3 (Metoda Monte Carlo). Mamy problem $x \rightarrow R(x)$ i chcemy znaleźć jego FPRAS. W tym celu projektujemy zmienną losową X o $E[X] = R(x)$, powtarzamy ją tak dużo razy, aż ograniczenie Czernowa da nam odpowiednią aproksymację.

Zakładając, że X jest indykátorem, chcemy iterować zmienną $m \geq \frac{3 \ln(\frac{2}{\delta})}{R(x)\varepsilon^2}$ razy. Ta liczba jest

wielomianowa od $\frac{1}{\varepsilon}$ i $\ln\left(\frac{1}{\delta}\right)$, ale $\frac{1}{R(x)}$ niekoniecznie jest wielomianowe od x . Jeśli jednak jest, to znaleźliśmy dobrą aproksymację.

Definicja 78. Dysjunktywna postać normalna (DNF – disjunctive normal form) formuły boolowskiej to postać, która jest alternatywą klauzul składających się z koniunkcji literalów, np.

$$(x_1 \wedge \overline{x_2} \wedge x_2) \vee (\overline{x_1} \wedge x_3) \vee (\overline{x_1} \wedge \overline{x_2} \wedge x_4 \wedge x_5).$$

Łatwo jest sprawdzić, czy taka formuła jest spełnialna – wystarczy, że jakaś formuła nie zawiera wyrażenia postaci $x \wedge \overline{x}$.

Przykład. Chcemy znaleźć FPRAS dla problemu zliczenia wartościowań spełniających formułę logiczną F o n zmiennych. Oznaczmy tę wartość przez $C(F)$ i zdefiniujmy $R(F) = \frac{C(F)}{2^n}$. Możemy podejść do sprawy naiwnie: losujemy jednostajnie wartościowanie i oznaczamy przez X indykator tego, czy spełnia ono F . Mamy $E[X] = R(F)$ i po powtórzeniu eksperymentu $m \geq \frac{3 \ln(\frac{2}{\delta})}{R(F)\varepsilon^2}$ razy dostaniemy (ε, δ) -aproksymację. Wartość $\frac{1}{R(F)} = \frac{2^n}{C(F)}$ jest wielomianowa od n dla $C(F) = \frac{2^n}{\text{poly}(n)}$, ale dla mniejszych już niekoniecznie. Zatem przedstawione podejście nie zawsze daje nam FPRAS.

Algorytm 4. Mając zadaną formułę w postaci DNF $F = C_1 \vee C_2 \vee \dots \vee C_t$ o n zmiennych możemy wyznaczyć liczbę wartościowań ją spełniających $C(F)$ za pomocą następującego algorytmu: pozbywamy się klauzul zawierających wyrażenia postaci $x \wedge \overline{x}$ i definiujemy SC_i jako zbiór wartościowań spełniających C_i . Do tego $S = \{(i, a) : i \in [t], a \in SC_i\}$, $RC_i = SC_i \setminus \bigcup_{j=1}^{i-1} SC_j$, $R = \{(i, a) : i \in [t], a \in RC_i\}$.

Wybieramy $i \in [t]$ z prawdopodobieństwem $\frac{|SC_i|}{|S|}$ a następnie wybieramy jednostajnie wartościowanie $a \in SC_i$. Niech X będzie indykatorem tego, czy $a \in RC_i$. Mamy $E[X] = \frac{C(F)}{|S|}$ i taki algorytm daje nam FPRAS.

Dowód. Zauważmy, że pozbycie się klauzul zawierających $x \wedge \overline{x}$ można wykonać liniowym przejściem po klauzulach. Jeśli C_i ma ℓ_i literalów (czyli ℓ_i zmiennych), to spełnia ją dokładnie $2^{n-\ell_i}$ wartościowań (ustalamy zmienne w klauzuli, reszta dowolnie). Mamy $|S| = \sum_{i=1}^t |SC_i| = \sum_{i=1}^t 2^{n-\ell_i}$, a więc tę liczbę da się łatwo wyznaczyć. RC_i jest zbiorem wartościowań, które spełniają C_i , ale nie spełniają żadnej poprzedniej klauzuli. Jasne, że $C(F) = |R|$ (każde wartościowanie liczymy raz). Nasz algorytm właściwie losuje parę $(i, a) \in S$ i sprawdza, czy $(i, a) \in R$. Co więcej, robi to jednostajnie: najpierw wybieramy i z prawdopodobieństwem $\frac{|SC_i|}{|S|}$, a następnie a z prawdopodobieństwem $\frac{1}{|SC_i|}$ – razem $\frac{1}{|S|}$. Zauważmy, że wylosowanie a da się zrobić łatwo – losujemy jednostajnie wartości literalów nie ustalonych przez C_i . Dla $R(F) = \frac{C(F)}{|S|} = \frac{|R|}{|S|}$ mamy $E[X] = R(F)$. Aby dostać (ε, δ) -aproksymację musimy powtórzyć losowanie $\frac{3 \ln(\frac{2}{\delta})}{R(F)\varepsilon^2}$ razy. Mamy $\frac{1}{R(F)} = \frac{|S|}{|R|} \leq t$, bo każde wartościowanie z R występuje w S co najwyżej t razy. Zatem $\frac{1}{R(F)}$ jest wielomianowe od F i faktycznie mamy FPRAS. $\square$

Definicja 79. Zmienna losowa X (randomizowany algorytm) na skończonej przestrzeni Ω daje ε -jednostajną próbkę Ω , jeśli dla każdego $S \subseteq \Omega$ jest

$$\left| P(X \in S) - \frac{|S|}{|\Omega|} \right| \leq \varepsilon.$$

Innymi słowy: dla rozkładu jednostajnego U i P_X będącego rozkładem X jest $\|P_X - U\|_{TV} \leq \varepsilon$.

Definicja 80. Mając dany problem obliczeniowy $x \rightarrow \Omega(x)$ (gdzie $\Omega(x)$ to zbiór rozwiązań) mówimy, że mamy w pełni wielomianowy prawie jednostajny schemat próbkowania (FPAUS – fully polynomial almost uniform sampler), jeśli dla wejścia x i parametru ε umiemy generować ε -jednostajną próbkę $\Omega(x)$ w czasie wielomianowym od $x, \ln\left(\frac{1}{\varepsilon}\right)$.

Pomysł. Mając zadany FPAUS dla jakiejś przestrzeni będziemy chcieli znajdować FPRAS obliczając wielkość tej przestrzeni.

Twierdzenie 49. Niech $\Omega(G)$ oznacza rodzinę zbiorów niezależnych w grafie G . Zakładając, że mamy dla niej FPAUS, istnieje FPRAS jej rozmiaru.

Dowód. Będziemy chcieli znaleźć (ε, δ) -aprosymację $|\Omega(G)|$. Ustalamy porządek $e_1, \dots, e_k$ na krawędziach G i oznaczamy $G_i = (V(G), \{e_1, \dots, e_i\})$. G_0 nie ma krawędzi, a więc $|\Omega(G_0)| = 2^n$, gdzie $n = |V(G)|$. Do tego $G_k = G$. Niech $r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, $\varepsilon' = \frac{\varepsilon}{2k}$, $\delta' = \frac{\delta}{k}$.

Zauważmy, że mamy $|\Omega(G)| = \frac{|\Omega(G_k)|}{|\Omega(G_{k-1})|} \cdot \dots \cdot \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \cdot |\Omega(G_0)| = r_k \cdot \dots \cdot r_1 \cdot 2^n$. Będziemy chcieli znaleźć s_i będące (ε', δ') -aprosymacją r_i . Niech $W = s_k \cdot \dots \cdot s_1 \cdot 2^n$. Mając $P(|s_i - r_i| > \varepsilon' r_i) < \delta'$ dostajemy

$$P\left(\bigcup_{i=1}^k \{|s_i - r_i| > \varepsilon' r_i\}\right) \leq \sum_{i=1}^k P(|s_i - r_i| > \varepsilon' r_i) < k \cdot \delta' = \delta.$$

Teraz możemy przeliczyć

$$\begin{aligned} P(|W - |\Omega(G)|| \leq |\Omega(G)| \varepsilon) &= P\left(1 - \varepsilon \leq \frac{W}{|\Omega(G)|} \leq 1 + \varepsilon\right) = P\left(1 - \varepsilon \leq \prod_{i=1}^k \frac{s_i}{r_i} \leq 1 + \varepsilon\right) \geq \\ &= P\left(\bigcap_{i=1}^k \left\{1 - \varepsilon' \leq \frac{s_i}{r_i} \leq 1 + \varepsilon'\right\}\right) = P\left(\bigcap_{i=1}^k \{|s_i - r_i| \leq \varepsilon' r_i\}\right) \geq 1 - \delta, \end{aligned}$$

gdzie pierwsza nierówność jest konsekwencją ciągu nierówności

$$1 - \varepsilon \leq \left(1 - \frac{\varepsilon}{2k}\right)^k \leq \prod_{i=1}^k \frac{s_i}{r_i} \leq \left(1 + \frac{\varepsilon}{2k}\right)^k \leq 1 + \varepsilon.$$

Pierwsza z tych nierówności wynika z nierówności Bernoulliego $\left(1 - \frac{\varepsilon}{2k}\right)^k \geq 1 - \frac{\varepsilon}{2}$, a ostatnia to szacowanie $\left(1 + \frac{\varepsilon}{2k}\right)^k = \sum_{i=0}^k \binom{k}{i} \left(\frac{\varepsilon}{2k}\right)^i \leq \sum_{i=0}^k k^i \frac{\varepsilon^i}{2^i k^i} \leq 1 + \sum_{i=1}^k \frac{\varepsilon^i}{2^i} \leq 1 + \varepsilon$, gdzie w przedostatnim kroku założyliśmy $\varepsilon \leq 1$.

Z tego wynika, że jeśli znajdziemy FPRAS dla wartości $R = \frac{|\Omega(H)|}{|\Omega(H')|}$, gdzie $H = H' + e$ dla pewnej krawędzi e , to uzyskamy naszą tezę. Wykorzystamy do tego FPAUS. Bierzemy $\frac{\varepsilon'}{3}$ -jednostajnie zbiór $I \in \Omega(H')$ i ustalamy indykator $X = [I \in \Omega(H)]$. Powtarzamy taki eksperyment m razy: $Y = \frac{1}{m} \sum_{i=1}^m X_i$. Niech $\mu = E[Y]$.

Zauważmy, że mamy $R \geq \frac{1}{2}$ – każdy zbiór niezależny w $\Omega(H') \setminus \Omega(H)$ można jednoznacznie przypisać do zbioru w $\Omega(H)$ odejmując od niego v będące ustalonym końcem e . Zatem w $\Omega(H')$ jest co najwyżej dwa razy więcej zbiorów. Z określenia naszego próbkowania mamy $|\mu - R| = |P(I \in \Omega(H)) - R| \leq \frac{\varepsilon'}{3}$. Z tego mamy $\mu \geq \frac{1}{3}$, bo można założyć $\varepsilon' \leq \frac{1}{2}$. Do tego

$$1 - \frac{2\varepsilon'}{3} \leq 1 - \frac{\varepsilon'}{3R} \leq \frac{\mu}{R} \leq 1 + \frac{\varepsilon'}{3R} \leq 1 + \frac{2\varepsilon'}{3}.$$

Biorąc $m \geq \frac{9 \ln(\frac{2}{\delta'})}{\left(\frac{\varepsilon'}{6}\right)^2} \geq \frac{3 \ln(\frac{2}{\delta'})}{\mu \left(\frac{\varepsilon'}{6}\right)^2}$ dostajemy z nierówności Czernowa $P\left(1 - \frac{\varepsilon'}{6} \leq \frac{Y}{\mu} \leq 1 + \frac{\varepsilon'}{6}\right) \geq 1 - \delta'$. W połączeniu z poprzednim daje nam to

$$P\left(\left(1 - \frac{\varepsilon'}{6}\right) \left(1 - \frac{2\varepsilon'}{3}\right) \leq \frac{Y}{\mu} \cdot \frac{\mu}{R} \leq \left(1 + \frac{\varepsilon'}{6}\right) \left(1 + \frac{2\varepsilon'}{3}\right)\right) \geq 1 - \delta'.$$

Rozpisując lewą stronę nierówności dostajemy

$$\left(1 - \frac{\varepsilon'}{6}\right) \left(1 - \frac{2\varepsilon'}{3}\right) = 1 - \frac{5\varepsilon'}{6} + \frac{\varepsilon'^2}{9} \geq 1 - \varepsilon'$$

i podobnie dla prawej strony. Mamy zatem

$$P\left(1 - \varepsilon' \leq \frac{Y}{R} \leq 1 + \varepsilon'\right) \geq 1 - \delta',$$

czyli przedstawiony proces daje dobrą aproksymację R . Zauważmy jeszcze, że m jest wielomianowe od wejścia, a więc mamy FPRAS. $\square$

Pomysł. Wiemy, że łańcuch Markowa zbiega do rozkładu stacjonarnego. Możemy to wykorzystać do znajdowania FPAUS różnych problemów.

Twierdzenie 50. Rozważmy graf G z $\Delta(G) \leq 4$, w którym do tego nie ma trójkątów¹. Niech $\Omega(G)$ będzie rodziną zbiorów niezależnych w G . Istnieje FPAUS dla $\Omega(G)$.

¹W książce *Probability and Computing* nie ma tego założenia. Jest ono jednak potrzebne w części dowodu, w której rozważamy przypadki.

Dowód. Możemy założyć bez straty ogólności, że G jest spójny – inaczej można próbować z każdej składowej osobno. Zdefiniujemy łańcuch Markowa o stanach z $\Omega(G)$ w następujący sposób: jego krok $K_e^p(I)$ polega na wylosowaniu jednostajnie $e = uv$ z $E(G)$ (ustalamy jakiś porządek na wierzchołkach, żeby móc mówić o pierwszym i drugim wierzchołku krawędzi) i $p \in \{1, 2, 3\}$. Teraz mamy przypadki

$$\begin{aligned} p = 1 &\rightarrow I' = I \setminus \{u, v\} \\ p = 2 &\rightarrow I' = I \setminus \{u\} \cup \{v\} \\ p = 3 &\rightarrow I' = I \setminus \{v\} \cup \{u\} \end{aligned}$$

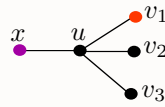
Jeśli $I' \in \Omega(G)$, to $K_e^p(I) = I'$, inaczej $K_e^p(I) = I$. Tak zadany łańcuch jest nieprzywiedlny i nieokresowy – można usuwać odpowiednie wierzchołki aż osiągnie się zbiór pusty, a potem dodawać odpowiednie wierzchołki i otrzymać dowolny zbiór. Nieokresowość wynika z tego, że można stać w miejscu. Rozkład stacjonarny tego łańcucha jest jednostajny, bo prawdopodobieństwo przejścia $I \rightarrow J$ jest takie samo jak $J \rightarrow I$.

Rozważmy sprzęganie $((X_t, Y_t))_{t \in \mathbb{N}}$ łańcuchów takie, że w kroku losujemy jednostajnie p_t, e_t i ustalamy $(X_{t+1}, Y_{t+1}) = (K_{e_t}^{p_t}(X_t), K_{e_t}^{p_t}(Y_t))$. Skracamy zapis $K = K_{e_t}^{p_t}$. Oznaczmy $d_t = |X_t \Delta Y_t|$, gdzie Δ oznacza różnicę symetryczną. Pokażemy, że $E[d_{t+1} \mid d_t] \leq d_t$. Zwykle wykazywaliśmy ostrą nierówność i z niej wynikało ograniczenie na czas mieszania. Z nieostrej też wynika szybkie mieszanie się łańcucha, ale nie będziemy tego pokazywać.

Jeśli $d_t = 0$, to $d_{t+1} = 0$. Rozważmy teraz przypadek $d_t = 1$ – mamy $I, J \in \Omega(G)$ takie, że $J = I \cup \{x\}$ i chcemy pokazać, że dla $D = |K(I) \Delta K(J)|$ zachodzi $E[D] \leq 1$. Niech $e = uv$. Jeśli $u, v \notin N(x)$, to $D = 1$, bo oba zbiory zachowują się tak samo podczas przejścia. Załóżmy, że $u \in N(x)$. Skorzystamy teraz z założeń i rozpatrzmy przypadki.

1. $|N(u) \cap I| \geq 2$. Tutaj zawsze $D \leq 1$, co sprawdzamy rozpatrując wszystkie przypadki.
2. $|N(u) \cap I| = 1$. Patrząc na rysunek: jeśli $v \in \{v_2, v_3\}$, to $D = 1$, bo w obu zbiorach zrobimy to samo (m.in. tu ważny jest brak trójkątów – v nie jest połączone z x). Jeśli $v = x$, to $p \in \{1, 2\} \rightarrow D = 0$ (usuwanie albo dodajemy x do obu) a $p = 3 \rightarrow D = 1$, bo ruch się nie uda. Jeśli $v = v_1$, to $p \in \{1, 2\} \rightarrow D = 1$, ale $p = 3 \rightarrow D = 3$, bo ruch nie uda się dla J , ale uda się dla I . Widzimy, że wartość oczekiwana D w tej sytuacji to 1.
3. $|N(u) \cap I| = 0$. Jeśli $v = x$, to $D = 0$. Jeśli $v = v_i$ dla $v_i \in N(u) \setminus \{x\}$, to $p \in \{1, 2\} \rightarrow D = 1$, ale $p = 3 \rightarrow D = 2$, bo u zostanie dodane do I . Mamy

$$E[D \mid e \in \{ux, uv_i\}] \leq \frac{1}{\Delta(G)} \cdot 0 + \frac{\Delta(G) - 1}{\Delta(G)} \left(\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 2 \right) = \frac{\Delta(G) - 1}{\Delta(G)} \cdot \frac{4}{3} \leq 1.$$



Rysunek 2: Przypadek, gdy I ma w sobie jednego sąsiada u .

Pozostało nam rozważyć przypadek $d_t > 1$. Zastosujemy metodę path coupling – połączymy X_t z Y_t „ścieżką”. Niech $X_t \setminus Y_t = \{x_1, \dots, x_r\}$, $Y_t \setminus X_t = \{y_1, \dots, y_s\}$. Oznaczmy

$$Z_0 = X_t, Z_1 = X_t \setminus \{x_1\}, \dots, Z_r = X_t \cap Y_t, Z_{r+1} = (X_t \cap Y_t) \cup \{y_1\}, \dots, Z_{r+s} = Z_{d_t} = Y_t.$$

Mamy $|Z_{i-1} \Delta Z_i| = 1$, a więc $E[|K(Z_{i-1}) \Delta K(Z_i)|] \leq 1$. Korzystając z nierówności $|A \setminus B| \leq |A \setminus C| + |C \setminus B|$ dostajemy

$$|K(Z_0) \setminus K(Z_{d_t})| \leq |K(Z_0) \setminus K(Z_1)| + |K(Z_1) \setminus K(Z_{d_t})| \leq \dots \leq \sum_{i=1}^{d_t} |K(Z_{i-1}) \setminus Z_i|,$$

a więc

$$\begin{aligned} |K(Z_0) \Delta K(Z_{d_t})| &= |K(Z_0) \setminus K(Z_{d_t})| + |K(Z_{d_t}) \setminus K(Z_0)| \\ &\leq \sum_{i=1}^{d_t} |K(Z_{i-1}) \setminus K(Z_i)| + |K(Z_i) \setminus K(Z_{i-1})| = \sum_{i=1}^{d_t} |K(Z_{i-1}) \Delta K(Z_i)|. \end{aligned}$$

To daje nam

$$d_{t+1} = |X_{t+1} \Delta Y_{t+1}| = |K(Z_0) \Delta K(Z_{d_t})| \leq \sum_{i=1}^{d_t} |K(Z_{i-1}) \Delta K(Z_i)|.$$

Ostatecznie

$$E[d_{t+1} \mid d_t] \leq \sum_{i=1}^{d_t} E[|K(Z_{i-1}) \Delta K(Z_i)|] \leq d_t.$$

□