Zajęcia 1: Przestrzeń probabilistyczna

2024-10-04

Przykład. Weryfikacja równości dwóch wielomianów. $F(x) = \prod_{i=1}^{d}(x - a_i)$ i $G(x)$ w innej postaci. Wymnożenie i sprawdzenie współczynników działa w $O(d^2)$. Można lepiej: wybieramy losowo ze zbioru $\{1, \ldots, 100 \cdot d\}$, 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})
- $P : \mathcal{F} \to \mathbb{R}$ 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 $P\left(\bigcup_{i \in \mathbb{N}} E_i\right) = \sum_{n=1}^{\infty} P(E_i)$

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 $\Omega = \{1, \ldots, 100d\}$, 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 $\left(\frac{1}{100}\right)^k$.

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

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 $E_1, \ldots, E_n$ 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 $R = (r_1, \ldots, r_n) \in \{0,1\}^n$ 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 $\{r_2, \ldots, r_n\}$, potem dopiero

r_1. Teraz mamy:

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

Zajęcia 2: Zestaw 1

2024-10-08

Twierdzenie 2. Dla przestrzeni probabilistycznej $(\Omega, \mathcal{F}, P)$ 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. $P(\emptyset) = P\left(\bigcup_{n=1}^{\infty} \emptyset\right) = \sum_{n=1}^{\infty} P(\emptyset) \implies P(\emptyset) = 0$
2. $1 = P(\Omega) = P(\Omega \setminus A \cup A) = P(\Omega \setminus A) + P(A)$
3. jak wyżej
4. z powyższego i nieujemności $P(B \setminus A)$
5. z punktu drugiego
6. $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B) = P(A) + P(B) - P(A \cap B)$.

□

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

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

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

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

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

$$P\left(\bigcup_{i=0}^{\infty} A_i\right) = P\left(\bigcup_{i=0}^{\infty} (A_i \setminus A_{i-1})\right) = \sum_{i=0}^{\infty} P(A_i \setminus A_{i-1}) = \lim_{n\to\infty}\left[\sum_{i=0}^{n} P(A_i \setminus A_{i-1})\right] = \lim_{n\to\infty} P(A_n).$$

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

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

□

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

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

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

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

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

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

□

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

Dowód. Mamy

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

□

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

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

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

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

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

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

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

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

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

oraz

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

co daje

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

□

Zajęcia 3: Zmienna losowa

2024-10-04

Definicja 5. X jest zmienną losową na przestrzeni $(\Omega, \mathcal{F}, P)$, jeśli $X : \Omega \to \mathbb{R}$ 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 $X_1, \ldots, X_n$ są niezależne, jeśli

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

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

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

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

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

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

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

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

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

□

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

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

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

□

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

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

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

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

Zajęcia 4: Rozkład zmiennej losowej

2024-10-11

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

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

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

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

Dowód. Niech $Y = (X - E[X])^2$. Mamy

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

□

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

$$\forall_{x_1, x_2, 0 \le \lambda \le 1} \; f\left(\lambda x_1 + (1-\lambda) x_2\right) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

Dla funkcji różniczkowalnej jest to równoważne dodatniości drugiej pochodnej.

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

$$E\left[f(X)\right] \ge f(E\left[X\right]).$$

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

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

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

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

bo $E\left[X - E\left[X\right]\right] = 0$. □

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

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

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 $P\left(X = i\right) = \binom{n}{i} p^i \left(1-p\right)^{n-i}$. Mamy

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

więc jest to rozkład.

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

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

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

Można też wprost:

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

□

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

$$E\left[Y \mid Z = z\right] = \sum_y y P\left(Y = y \mid Z = z\right).$$

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

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

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

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

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

Dowód.

$$\sum_y E[X \mid Y = y] P(Y = y) = \sum_y \sum_x xP(X = x \mid Y = y) P(Y = y)$$
$$= \sum_x x \sum_y P(X = x \mid Y = y) P(Y = y) = \sum_x x \sum_y P(X = x \cap Y = y)$$
$$= \sum_x xP(X = x) = E[X].$$

□

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

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

Dowód. Dowód jak liniowość. □

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

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

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

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

Propozycja 6. $E[Y] = E[E[Y \mid Z]]$.

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

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

□

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.

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

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 $E\left[Y_i \mid Y_{i-1}\right] = Y_{i-1}np$.

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

Ostatecznie $E\left[\sum Y_i\right] = \sum E\left[Y_i\right] = \sum (np)^i = \frac{1}{1-np}$. 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 $P\left(X = n\right) = (1-p)^{n-1} p$ dla $n \in \mathbb{N}_1$. Mamy $\sum_{n\in\mathbb{N}_1} P\left(X = n\right) = 1$ (szereg geometryczny), więc jest to rozkład.

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

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

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

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

□

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

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

Dowód.

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

□

Zajęcia 5: Zestaw 2

2024-10-15

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

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

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

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

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

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

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

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

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

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

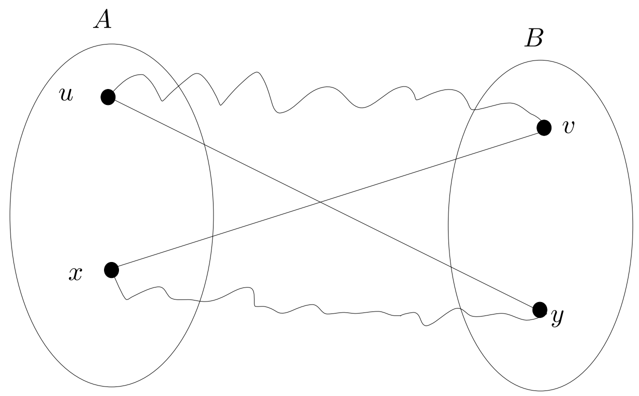
Dowód. Dla dowolnego grafu G losujemy podgraf H – mamy wszystkie wierzchołki i losujemy każdą krawędź niezależnie z prawdopodobieństwem p na należenie do H. W tak otrzymanym grafie najprawdopodobniej będą cykle C_4. Rozważamy każdy z nich i z każdego usuwamy po jednej krawędzi, w ten sposób otrzymując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

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

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

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

□



Rysunek 1: C_4 w grafie H

Zajęcia 6: Miara i całka

2024-10-18

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

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

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

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

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

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

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

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

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

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

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

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 $\sigma\left(\{[a,b] : a < b\}\right) = \mathcal{B}\left(\mathbb{R}\right)$, gdzie $\mathcal{B}\left(\mathbb{R}\right)$ 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 $\mathcal{B}\left(\mathbb{R}\right) = \sigma\left(A_o\right)$.

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

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

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

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

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

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

Definicja 22. Funkcję $s : \Omega \to [0, \infty)$ 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

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

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

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

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

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

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

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

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

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

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

dla każdego $x \in \Omega$.

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

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

jest też

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

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 $X : (\Omega, \Sigma) \to \mathbb{R}$
$E[X]$	$\int_\Omega X\,d\mu$
$\mathrm{Var}(X)$	$\int_\Omega (X - E[X])^2\,d\mu$

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

Przykład. Mamy wagi o sumie 1: $p_1, p_2, \ldots, p_n$. 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 $P(\{0\}) = \frac{1}{3}$, $P\left(\left(0, \frac{1}{3}\right)\right) = 0$, $P(A) = \lambda(A)$ 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 $[a_1a_2\ldots a_k] = \{(x_n)_{n=1}^{\infty} \in \Omega : a_i = x_i\}$ będzie zbiorem wszystkich ciągów zaczynających się w dany sposób. Przez cylinder $C = \{[a_1a_2\ldots a_k] : a_i \in \{0, 1\}, k \in \mathbb{N}\}$ rozumiemy zbiór zawierający wszystkie takie rodziny ciągów o wspólnym początku. Rozważamy σ-algebrę $\Sigma = \sigma(C)$. Mamy $P([a_1 \ldots a_k]) = \frac{1}{2^k}$.

Zajęcia 7: Zestaw 3

2024-10-22

Twierdzenie 21. Złożenie dwóch funkcji mierzalnych jest mierzalne.

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

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

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

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

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

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

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

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

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

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

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

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

Przejście na funkcje zespolone analogicznie z definicji. □

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

$$\int_E f + g \, \mathrm{d}\mu = \int_E f \, \mathrm{d}\mu + \int_E g \, \mathrm{d}\mu \, .$$

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

Mamy

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

Teraz

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

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

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

$$\int_E f\,\mathrm{d}\mu+\int_E g\,\mathrm{d}\mu=\lim_{n\to\infty}\int_E s_n\,\mathrm{d}\mu+\lim_{n\to\infty}\int_E t_n\,\mathrm{d}\mu=\lim_{n\to\infty}\left(\int_E s_n\,\mathrm{d}\mu+\int_E t_n\,\mathrm{d}\mu\right)$$
$$=\lim_{n\to\infty}\int_E s_n+t_n\,\mathrm{d}\mu=\int_E f+g\,\mathrm{d}\mu\,.$$

Przejście na funkcje rzeczywiste i zespolone prosto z definicji. □

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

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

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

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

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

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

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

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

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

Zajęcia 8: Wariancja

2024-10-24

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

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

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 $x_1, \ldots, x_n$. 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 $\{x_i, x_{i+1}, \ldots, x_j\}$. Mamy $P(Q \text{ wybrał za pivot } x_i \text{ lub } x_j \mid Q \text{ wybrał pivota spośród } \{x_i, \ldots, x_j\}) = \frac{\frac{2}{\|Z\|}}{\frac{j-i+1}{\|Z\|}} = \frac{2}{j-i+1}$, gdzie Z jest zbiorem wartości, na których odpalony jest algorytm w momencie decyzji, czy elementy zostaną porównane. Jak widać wielkość Z nie ma znaczenia.

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

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

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

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

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

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

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

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

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

Definicja 28. Wariancja zmiennej X to wartość

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

Definicja 29. Odchylenie standardowe zmiennej X to wartość

$$\sigma = \sqrt{\mathrm{Var}\left(X\right)}.$$

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

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

Definicja 30. Kowariancja zmiennych losowych X, Y to wartość

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

Lemat 1. Dla dowolnych zmiennych losowych X, Y zachodzi

$$\mathrm{Var}\left(X + Y\right) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(Y\right) + 2\,\mathrm{Cov}\left(X, Y\right).$$

Dowód.

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

□

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

$$E\left[XY\right] = E\left[X\right]E\left[Y\right].$$

Dowód.

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

□

Wniosek. Dla niezależnych zmiennych losowych X, Y zachodzi

$$\mathrm{Cov}\left(X, Y\right) = 0.$$

Wniosek. Dla niezależnych zmiennych losowych X, Y zachodzi

$$\mathrm{Var}\left(X + Y\right) = \mathrm{Var}\left(X\right) + \mathrm{Var}\left(Y\right).$$

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

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

Dowód. Mamy

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

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

Można też prościej rozbijając na pojedyncze próby:

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

Zajęcia 9: Nierówności koncentrujące

2024-10-25

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

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

Dowód.

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

gdzie nierówność to nierówność Markowa. □

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

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

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

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

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

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

Dowód.

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

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

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

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

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

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

Dowód.

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

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

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

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

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

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 $E\left[Y^2\right] = 2p^{-2}(1-p) + p^{-1}$.

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

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

oraz

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

Dowód. Dla $t > 0$ zachodzi $P(X \geq a) = P\left(e^{tX} \geq e^{ta}\right) \leq \frac{E\left[e^{tX}\right]}{e^{ta}} = \frac{M_X(t)}{e^{ta}}$, 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. □

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

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

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

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

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 $P\left(X \geq \frac{3}{4}n\right) \leq \frac{e^{\frac{1}{2}n(e^t-1)}}{e^{\frac{3}{4}nt}} = e^{\frac{1}{2}ne^t - \frac{1}{2}n - \frac{3}{4}nt} = e^{\frac{1}{2}nf(t)}$, 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

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

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

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

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 $P(Y \geq 2nH_n) = P\left(\bigcup_{i=1}^{n}\{Y_i \geq 2nH_n\}\right) \leq \sum_{i=1}^{n} P(Y_i \geq 2nH_n) \leq n \cdot \frac{1}{n^2} = \frac{1}{n}$, 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	$\mathrm{Uni}(\{0,1\})$	$\mathrm{Geo}\left(\frac{n+i-1}{n}\right)$
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 $p_1, \ldots, p_n \in (0,1)$, zmienne losowe X_i takie, że $P(X_i = 1) = p_i$, $X_i \in \{0,1\}$.

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

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

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

Zajęcia 10: Zestaw 4

2024-10-29

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

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

Dowód. Rozbijamy naszą zmienną X na sumę $X = \sum_{\{i,j,k,l\}\in\binom{[n]}{4}} X_{ijkl}$, gdzie X_{ijkl} jest indykatorem mówiącym, czy na wybranych wierzchołkach jest klika. Mamy $E[X] = \sum_{\{i,j,k,l\}\in\binom{[n]}{4}} E[X_i] = \binom{n}{4} \cdot p^6$. Liczymy $E\left[X^2\right]$ rozbijając na indykatory:

$$E\left[X^2\right] = \sum_{\substack{\{ijkl\}\in\binom{[n]}{4}\\ \{efds\}\in\binom{[n]}{4}}} E\left[X_{ijkl}X_{efds}\right] =$$

$$\binom{n}{4}\left[\binom{4}{4}\binom{n-4}{0}p^6 + \binom{4}{3}\binom{n-4}{1}p^9 + \binom{4}{2}\binom{n-4}{2}p^{11} + \binom{4}{1}\binom{n-4}{3}p^{12} + \binom{4}{0}\binom{n-4}{4}p^{12}\right]$$

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

Po przeliczeniu dostajemy

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

□

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

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

a do tego

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

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

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

Niech X_i będzie indykatorem 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

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

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

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

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

Dowód. Zachodzi

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

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

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

□

Zajęcia 11: Nierówności Czernowa

2024-11-05

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

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

Dowód. Liczymy funkcję tworzącą

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

Zatem

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

Ustalmy $t>0$, mamy

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

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

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

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

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

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

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

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

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

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

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

□

Twierdzenie 38. Niech $X_1, \ldots, X_n$ 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 $2 \cdot e^{-\frac{n}{2}\cdot 6\frac{\ln n}{n}\cdot\frac{1}{3}} = \frac{2}{n}$.

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

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

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

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

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

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

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

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

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

Rozwijamy w szereg Taylora:

$$e^t = 1 + t + \frac{t^2}{2} + \ldots + \frac{t^i}{i!} + \ldots$$
$$e^{-t} = 1 - t + \frac{t^2}{2} + \ldots + (-1)^i \frac{t^i}{i!} + \ldots$$

Z tego wynika

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

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

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

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

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 $Y_1, \ldots, Y_n$ 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 $X = \sum_{i=1}^{n} X_i = 2Y - \mu$ mamy

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

oraz

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

□

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

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

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

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

Wektor $\overline{b}$ wyznaczamy, losując.

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

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

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

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

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

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

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

Zajęcia 12: Rozkład Poissona

2024-11-08

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

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

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

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

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

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

Jakie jest maksymalne zapełnienie urny?

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

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

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

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

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

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

Kończymy za pomocą union bounda:

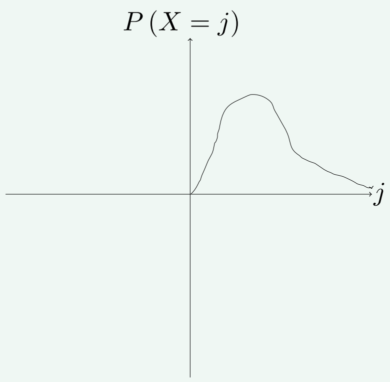
$$P\left(\text{maksymalne zapełnienie} > \frac{4\ln n}{\ln\ln n}\right) = P\left(\bigcup_{i\in[n]}\bigcup_{k>\frac{4\ln n}{\ln\ln n}}\{i\text{-ta urna ma dokładnie } k \text{ kul}\}\right)$$
$$\leq \sum_{\substack{i\in[n]\\ k>\frac{4\ln n}{\ln\ln n}}} P\left(i\text{-ta urna ma dokładnie } k \text{ kul}\right)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

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

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

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

dla każdego $j \in \mathbb{N}$.



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

Mamy

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

czyli faktycznie jest to rozkład.

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

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

Czyli zwiększając liczbę prób w nieskończoność dostajemy rozkład Poissona.

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

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

$$\mathrm{Var}\left(X\right) = \mu.$$

Dowód.

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

Podobnie

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

□

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

Dowód.

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

□

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

Dowód.

$$P\left(X+Y=j\right) = \sum_{k=0}^{j} P\left(X=k\cap Y=j-k\right) = \sum_{k=0}^{j} P\left(X=k\right)P\left(Y=j-k\right)$$
$$= \sum_{k=0}^{j}\frac{e^{-\mu_1}\mu_1^k}{k!}\cdot\frac{e^{\mu_2}\mu_2^{j-k}}{(j-k)!} = e^{-(\mu_1+\mu_2)}\cdot\frac{1}{j!}\sum_{k=0}^{j}\frac{j!}{k!\,(j-k)!}\mu_1^k\mu_2^{j-k} = e^{-(\mu_1+\mu_2)}\cdot\frac{1}{j!}\left(\mu_1+\mu_2\right)^j.$$

Można też rozważyć funkcję tworzącą sumy:

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

która jest funkcją tworzącą rozkładu Poissona z odpowiednim parametrem. □

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

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

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

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

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. □

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

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

Lemat 5. Mamy ciąg niezależnych zmiennych o rozkładzie Poissona $\left(Y_1^{(k)}, \ldots, Y_u^{(k)}\right)$ 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 $\left(X_1^{(m)}, \ldots, X_u^{(m)}\right)$ modelujących wrzucanie m kul do u urn (niezależnie od wartości k).

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

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

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:

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

co równa się odpowiedniej wartości.

W drugim modelu jest

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

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. □

Lemat 6. Zachodzi

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

Dowód. Korzystamy z

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

Mamy
$$\sum_{i=1}^{n} \ln i - \frac{1}{2}(\ln 1 + \ln n) = \sum_{i=1}^{n-1} \frac{1}{2}(\ln i + \ln(i+1))(i+1-i) \leq \int_1^n \ln x \, dx,$$
gdzie nierówność wynika z tego, że druga suma jest przybliżeniem pola pod $\ln x$ za pomocą trapezów, a logarytm jest funkcją wklęsłą (bo pochodna jest ujemna).

Dla pełności policzymy tę całkę:
$$\int_1^n \ln x \, dx = x \ln x - \int_1^n x \cdot \frac{1}{x} \, dx = n \ln n - n + 1.$$
Przypominając, że $\ln 1 = 0$, 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 44. Niech $f(x_1, \ldots, x_u)$ będzie dowolną nieujemną funkcją. Mamy ciąg niezależnych zmiennych o rozkładzie Poissona $\left(Y_1^{(k)}, \ldots, Y_u^{(k)}\right)$ z parametrem $\frac{k}{u}$ oraz ciąg zmiennych modelujących wrzucanie k kul do u urn $\left(X_1^{(k)}, \ldots, X_u^{(k)}\right)$. Zachodzi dla nich nierówność
$$E\left[f\left(X_1^{(k)}, \ldots, X_u^{(k)}\right)\right] \leq e\sqrt{k} E\left[f\left(Y_1^{(k)}, \ldots, Y_u^{(k)}\right)\right].$$
Funkcja f może być na przykład indykatorem zdarzenia które nas interesuje. Model jest skomplikowany, ale ograniczenie wynikające z rozkładu Poissona też może być ciekawe.

Dowód.
$$\begin{aligned}
E\left[f\left(Y_1^{(k)}, \ldots, Y_u^{(k)}\right)\right] &= \sum_{j=0}^{\infty} E\left[f\left(Y_1^{(k)}, \ldots, Y_u^{(k)}\right) \mid \sum_{i=1}^{u} Y_i^{(k)} = j\right] \cdot P\left(\sum_{i=1}^{u} Y_i^{(k)} = j\right) \\
&\geq E\left[f\left(Y_1^{(k)}, \ldots, Y_u^{(k)}\right) \mid \sum_{i=1}^{u} Y_i^{(k)} = k\right] \cdot P\left(\sum_{i=1}^{u} Y_i^{(k)} = k\right) \\
&= E\left[f\left(X_1^{(k)}, \ldots, X_u^{(k)}\right)\right] \cdot P\left(\sum_{i=1}^{u} Y_i^{(k)} = k\right) = E\left[f\left(X_1^{(k)}, \ldots, X_u^{(k)}\right)\right] \cdot e^{-k} k^k \cdot \frac{1}{k!} \\
&\geq E\left[f\left(X_1^{(k)}, \ldots, X_u^{(k)}\right)\right] \cdot \frac{1}{e\sqrt{k}}.
\end{aligned}$$
Pierwsza nierówność to ograniczenie się do $j = k$, a druga to ograniczenie górne na $k!$. □