



METODY PROBABILISTYCZNE INFORMATYKI - OPRACOWANIE PYTAŃ EGZAMINACYJNYCH

„Co dalej? Boże... co dalej? Co my w ogóle robimy tutaj? Nie wiem.”

POPEŁNIONE PRZEZ

ZAŁATANY PONTON

V

NAHTAMATU

Kraków
Anno Domini 2025

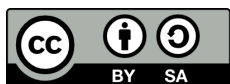
Spis treści

1	Pytanie 1	1
1.1	Rozkład dwumianowy	1
1.2	Rozkład geometryczny	2
2	Pytanie 2	5
2.1	Problem kolekcjonera kuponów	5
2.2	Oczekiwana liczba porównań Quicksorta	5
3	Pytanie 3	8
3.1	Wariancja	8
3.2	Nierówność Markowa	10
3.3	Nierówność Czebyszewa	11
3.3.1	Definicja	11
3.3.2	Kolekcjoner kuponów	11
4	Pytanie 4	12
4.1	Nierówność Czernowa	12
4.1.1	Definicja	12
4.1.2	Próby Poissona	12
4.1.3	Rzuty monetą	14
5	Pytanie 5	15
5.1	Kule i urny - ograniczenie górne	15
6	Pytanie 6	18
6.1	Rozkład Poissona	18
6.1.1	Podstawowe własności	18
6.1.2	Ograniczenia Czernowa	20
7	Pytanie 7	21
7.1	Aproksymacja Poissona	21
7.1.1	Definicja	21
7.1.2	Kule i urny	23

8 Pytanie 8	25
8.1 Granica kolekcjonera kuponów	25
9 Pytanie 9	28
9.1 Definicja łańcucha Markowa	28
9.2 Nieprzywiedlność i okres stanu	29
9.3 Okres nieprzywiedlnego łańcucha Markowa	30
9.4 Prawdopodobieństwa przejść pomiędzy stanami w nieprzywiedlnym i nieokresowym łańcuchu Markowa	31
9.5 Stany powracające i chwilowe	31
9.6 Własność nieprzywiedlnego, skończonego łańcucha Markowa	32
10 Pytanie 10	34
10.1 Rozkład stacjonarny	34
11 Pytanie 11	37
11.1 Norma całkowitego wahanja	37
11.2 Sprzęganie rozkładów prawdopodobieństwa	38
12 Pytanie 12	40
12.1 Sprzęganie łańcuchów Markowa	40
12.2 Lemat o sprzęganiu łańcuchów Markowa	41
12.3 Lemat o monotoniczności	41
13 Pytanie 13	43
13.1 Twierdzenie o geometrycznej zbieżności	43
14 Pytanie 14	45
14.1 Losowe spacery w grafie	45
15 Pytanie 15	47
15.1 Rozkład jednostajny	47
16 Pytanie 16	50
16.1 Rozkład wykładniczy	50
16.2 Funkcja Gamma i rozkład Gamma	54
17 Pytanie 17	57
17.1 Kule i urny z feedbackiem	57
18 Pytanie 18	60
18.1 Proces Poissona	60
18.1.1 Definicja	60
18.1.2 Związek z rozkładem Poissona	61

19 Pytanie 19	65
19.1 Międzyczasy	65
20 Pytanie 20	68
20.1 Scalanie i rozdzielanie procesów Poissona	68
20.1.1 Scalanie	68
20.1.2 Rozdzielanie	69
21 Pytanie 21	71
21.1 Warunkowe czasy pojawiania się zdarzeń w procesie Poissona	71
22 Pytanie 22	73
22.1 Standardowy rozkład normalny	73
22.1.1 Wyprowadzenie	73
22.1.2 Właściwości	75
22.2 Uogólniony rozkład normalny	76
23 Pytanie 23	78
23.1 Centralne Twierdzenie Graniczne	78
23.1.1 Podstawowa wersja	78
23.1.2 Warianty	81
24 Wprowadzenie siły dwóch wyborów	83
24.1 Wprowadzenie modelu eksperymentu	83
24.2 Lematy pomocnicze	83
24.3 Oznaczenia	85
25 Pytanie 24	86
25.1 Ograniczenie górne przez iterację ograniczeń	86
26 Pytanie 25	92
26.1 Ograniczenie górne przez drzewa świadczące	92
27 Pytanie 26	97
27.1 Ograniczenie dolne	97
28 Dodatkowe materiały	101
28.1 Aproksymacja szeregu harmonicznego	101
28.2 Wartość oczekiwana iloczynu	101
28.3 Wartość oczekiwana naturalnej zmiennej losowej	102
28.4 Równość funkcji tworzących	102
28.5 Ograniczenie górne Czernowa prób Poissona	103

Licencja



Ten utwór jest dostępny na licencji Creative Commons Uznanie autorstwa na tych samych warunkach 4.0 Międzynarodowe.

Pytanie 1

Rozkład dwumianowy, rozkład geometryczny i ich własności. Własność bez pamięci, wartość oczekiwana, wariancja, wyższe momenty. Funkcje tworzące momentów.

1.1 Rozkład dwumianowy

Definicja 1.1.1. Mówimy, że zmienna losowa X ma **rozkład dwumianowy** z parametrami n, p (Oznaczana poprzez $B(n, p)$), jeśli dla $j = 0, 1, \dots, n$:

$$P(X = j) = \binom{n}{j} p^j (1 - p)^{(n-j)}$$

Jest tak, gdy powtarzamy jakiś eksperyment wielokrotnie (n razy, gdzie p to szansa powodzenia) i liczymy, ile razy eksperyment się powiódł.

Fakt 1.1.1. Rozkład dwumianowy jest poprawnie zdefiniowany

$$\sum_{j=0}^n P(X = j) = (p + (1 - p))^n = 1$$

Twierdzenie 1.1.1. Niech X ma rozkład dwumianowy z parametrami n, p . Wtedy

$$\mathbb{E}[X] = np$$

Dowód.

$$X = \sum_{i=1}^n X_i, \quad X_i = \begin{cases} 1, & \text{sukces w } i\text{-tej próbie} \\ 0 & \text{wpp} \end{cases}$$

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n P(X_i = 1) = \sum_{i=1}^n p = np$$

□

Twierdzenie 1.1.2. Niech X ma rozkład dwumianowy z parametrami n, p . Wtedy

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

Dowód.

$$\text{Var}[X] = \text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i]$$

Gdzie X_i to indykatory dla kolejnych zdarzeń. Wariancja dla jednego indykatora:

$$\begin{aligned}\text{Var}[X_i] &= \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] \\ &= P(X_i = 0) \cdot (0 - p)^2 + P(X_i = 1) \cdot (1 - p)^2 \\ &= (1-p)p^2 + p(1-p)^2 \\ &= (1-p)p\end{aligned}$$

Sumując po i dostajemy:

$$\sum_{i=1}^n \text{Var}[X_i] = \sum_{i=1}^n (1-p)p = np(1-p)$$

□

Twierdzenie 1.1.3. Funkcja tworząca momenty:

$$M_X(t) = (1-p + pe^t)^n$$

Dowód.

$$\begin{aligned}M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{j=0}^n \binom{n}{j} p^j (1-p)^{n-j} e^{tj} \\ &= \sum_{j=0}^n \binom{n}{j} (pe^t)^j (1-p)^{n-j} \\ &= ((1-p) + pe^t)^n\end{aligned}$$

□

1.2 Rozkład geometryczny

Definicja 1.2.1. Mówimy, że zmienna losowa X ma **rozkład geometryczny** z parametrem $p \in (0, 1)$ jeśli dla $n > 0$

$$P(X = n) = (1-p)^{n-1} \cdot p$$

Twierdzenie 1.2.1 (Lemat 2.8 P&C). Rozkład geometryczny jest **bez pamięci** tzn. jeśli X ma rozkład geometryczny z parametrem p to

$$\forall_{n,k} P(X = n + k \mid X > k) = P(X = n)$$

Dowód.

$$\begin{aligned} P(X = n + k \mid X > k) &= \frac{P(X = n + k \wedge X > k)}{P(X > k)} \\ &= \frac{P(X = n + k)}{P(X > k)} \\ &= \frac{(1 - p)^{n+k-1} \cdot p}{(1 - p)^k} \\ &= (1 - p)^{n-1} \cdot p \\ &= P(X = n) \end{aligned}$$

□

Twierdzenie 1.2.2. Niech X ma rozkład geometryczny z parametrem p . Wtedy tworząca tej zmiennej wynosi

$$M_X(t) = \frac{pe^t}{1 - (1 - p)e^t}$$

dla $t < -\ln(1 - p)$.

Dowód.

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \sum_{i=1}^{\infty} (1 - p)^{i-1} p e^{it} \\ &= \frac{p}{1 - p} \cdot \sum_{i=1}^{\infty} ((1 - p)e^t)^i \\ &= \frac{p}{1 - p} \cdot \left(\left(\sum_{i=0}^{\infty} ((1 - p)e^t)^i \right) - 1 \right) \\ &= \frac{p}{1 - p} \cdot \left(\frac{1}{1 - (1 - p)e^t} - 1 \right) \\ &= \frac{p}{1 - p} \cdot \frac{(1 - p)e^t}{1 - (1 - p)e^t} \\ &= \frac{pe^t}{1 - (1 - p)e^t} \end{aligned}$$

Skorzystaliśmy tutaj z faktu, że szereg

$$\sum_{i=1}^{\infty} ((1-p)e^t)^i$$

jest zbieżny. Dzieje się tak gdy

$$(1-p)e^t < 1$$

$$e^t < \frac{1}{1-p}$$

$$t < -\ln(1-p)$$

□

Wniosek 1.2.1. W oparciu na funkcję tworzącą momenty, możemy obliczyć wartość oczekiwaną, wariancję, oraz dowolne wyższe momenty. Wiemy, że $M_X^{(n)}(0) = \mathbb{E}[X^n]$. Obliczamy więc $M_X^{(1)}(t)$ oraz $M_X^{(2)}(t)$. Otrzymujemy:

$$M_X^{(1)}(t) = p(1 - (1-p)e^t)^{-2}$$

$$M_X^{(2)}(t) = 2p(1-p)(1 - (1-p)e^t)^{-3}e^{2t} + p(1 - (1-p)e^t)^{-2}e^t$$

Po podstawieniu $t = 0$ otrzymujemy znane już wartości:

$$M_X^{(1)}(0) = \mathbb{E}[X] = \frac{1}{p}$$

$$M_X^{(2)}(0) = \mathbb{E}[X^2] = \frac{2-p}{p^2} \implies \text{Var}[X] = \frac{1-p}{p^2}$$

Pytanie 2

Problem kolekcjonera kuponów (wartość oczekiwana). Oczekiwana liczba porównań w algorytmie sortowania Quicksort.

2.1 Problem kolekcjonera kuponów

Wyobraźmy sobie problem, który jest bliski wielu osobom. Próbuje przepchać program na satori ale jak na złość mamy ANS. Sfrustrowani zaczynamy pisać własne testy w nadziei że znajdziemy przypadek brzegowy. I tutaj pojawia się pytanie – jeśli generujemy testy losowo a możliwych przypadków jest n to ile testów potrzebujemy w oczekiwaniu wygenerować aby mieć pewność, że pokryliśmy każdy przypadek?

Problem ten, jak wiele podobnych, możemy modelować za pomocą zbierania kuponów – mamy ich do zebrania n a szansa na uzyskanie i -tego rodzaju jeśli zebraliśmy już $i - 1$ wynosi $p_i = 1 - \frac{i-1}{n}$. Niech X_i oznacza czas czekania na i -ty kupon jeśli mamy już $i - 1$ innych. Wtedy $X = \sum_{i=1}^n X_i$ jest tym czego szukamy – czasem otrzymania każdego kuponu (pokrycia wszystkich przypadków testowych).

Zauważmy jeszcze, że X_i ma rozkład geometryczny z parametrem p_i zatem $\mathbb{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \frac{n}{n-i+1} = n \sum_{i=1}^n \frac{1}{i} = n \cdot H_n = n \ln n + \Theta(n)$$

Ostatnia równość wynika z 28.1.1.

2.2 Oczekiwana liczba porównań Quicksorta

Quicksort jaki jest każdy widzi – pamiętamy z ASD, że jego złożoność to pesymistycznie $\mathcal{O}(n^2)$, ale w losowym przypadku $\Theta(n \lg n)$.

Twierdzenie 2.2.1 (2.11 P&C). Rozważmy standardowy algorytm Quicksort, w którym pivota wybieramy losowo, niezależnie i jednostajnie. Wtedy oczekiwana liczba porównań wynosi $2n \ln n + \mathcal{O}(n)$.

Dowód. Niech $x_1, \dots, x_n$ będzie wejściowym ciągiem n różnych liczb. Niech $y_1, \dots, y_n$ będzie posortowaną permutacją tych wartości.

Definiujemy indykatory dla $i < j$; niech

$$X_{i,j} = \begin{cases} 1 & \text{jeśli } y_i, y_j \text{ zostały porównane chociaż raz} \\ 0 & \text{wpp.} \end{cases}$$

Łączna liczba porównań X wynosi $X = \sum_{i=0}^{n-1} \sum_{j=i+1}^n X_{i,j}$. Oczekiwana liczba porównań wynosi zatem

$$\mathbb{E}[X] = \sum_{i=0}^{n-1} \sum_{j=i+1}^n \mathbb{E}[X_{i,j}]$$

Zastanówmy się kiedy elementy y_i, y_j są porównywane. Na pewno któryś z nich musi zostać wybrany jako pivot. Ale ponadto muszą być w momencie tego wyboru na jednej liście, która jest aktualnie sortowana. Niech $Y^{i,j} = \{y_i, \dots, y_j\}$.

Jeśli wybrany zostanie pivot który leży poza tą listą, to nie dojdzie do „rozspójnienia” tej listy i kiedyś będzie mogło nadal dojść do porównania y_i z y_j .

Jeśli wybrany zostanie pivot z tej listy różny od y_i oraz y_j , to te 2 elementy już nigdy nie zostaną ze sobą porównane, jako że będą znajdować się na 2 oddzielnych listach.

W takim razie $X_{i,j} = 1$ wtedy i tylko wtedy, gdy pierwszym pivotem wybranym ze zbioru $Y^{i,j}$ jest element y_i lub element y_j .

Jako, że losowanie jest jednostajne i w ogóle, to każdy element z listy ma dokładnie takie same szanse na „zostanie pivotem”. Jako, że elementów na liście jest $j - i + 1$, to prawdopodobieństwo, że wybierzemy y_i lub y_j wynosi $\frac{2}{j-i+1}$, czyli $\mathbb{E}[X_{i,j}] = \frac{2}{j-i+1}$.

Aby policzyć ostateczny wynik sumujemy się po wszystkich parach $i < j$:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{2}{j-i+1} \\ &= 2 \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{1}{k} \\ &= 2 \sum_{k=2}^n \sum_{i=1}^{n-k+1} \frac{1}{k} \\ &= 2 \sum_{k=2}^n \frac{n+1-k}{k} \\ &= 2 \left((n+1) \sum_{k=2}^n \frac{1}{k} \right) - 2(n-1) \\ &= 2 \left((n+1) \left(\sum_{k=1}^n \frac{1}{k} \right) - (n+1) \right) - 2(n-1) \end{aligned}$$

Teraz korzystamy z 28.1.1 i dostajemy

$$\begin{aligned}\mathbb{E}[X] &= 2(n+1) \cdot H_n - \Theta(n) \\ &= 2(n+1) \cdot (\ln n + \Theta(1)) - \Theta(n) \\ &= 2n \ln n + \Theta(n)\end{aligned}$$

□

Pytanie 3

Własności wariancji. Nierówność Markowa. Nierówność Czebyszewa i jej zastosowanie w problemie kolekcjonera kuponów.

3.1 Wariancja

Definicja 3.1.1. Wariancję zmiennej losowej X definiujemy jako

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

Czyli jest to drugi moment zmiennej X przesuniętej o swoją wartość oczekiwaną. Intuicyjnie jest to miara tego, jakiego odchylenia od wartości oczekiwanej możemy się spodziewać.

Operator wariancji nie jest liniowy.

Definicja 3.1.2. Odchylenie standardowe zmiennej losowej X definiujemy jako

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

Definicja 3.1.3. Kowariancję zmiennych losowych X oraz Y definiujemy jako

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

Twierdzenie 3.1.1.

$$\forall_{a,b \in \mathbb{R}} \text{Var}[bX + a] = b^2 \text{Var}[X]$$

Dowód.

$$\begin{aligned}
 \text{Var}[bX + a] &= \mathbb{E}[(bX + a)^2] - \mathbb{E}[bX + a]^2 \\
 &= \mathbb{E}[b^2X^2 + 2abX + a^2] - (b\mathbb{E}[X] + a)^2 \\
 &= b^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + a^2 - b^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X] - a^2 \\
 &= b^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) \\
 &= b^2 \text{Var}[X]
 \end{aligned}$$

□

Twierdzenie 3.1.2. Dla dowolnych zmiennych losowych X, Y zachodzi

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

Dowód. Rozpisujemy $\text{Var}[X + Y]$ z definicji.

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

□

Twierdzenie 3.1.3. Dla niezależnych zmiennych losowych X, Y

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

a co za tym idzie

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

Dowód.

$$\begin{aligned}
 \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X]) \cdot (Y - \mathbb{E}[Y])] \\
 &= {}^1\mathbb{E}[X - \mathbb{E}[X]] \cdot \mathbb{E}[Y - \mathbb{E}[Y]] \\
 &= (\mathbb{E}[X] - \mathbb{E}[X]) \cdot (\mathbb{E}[Y] - \mathbb{E}[Y]) \\
 &= 0
 \end{aligned}$$

□

¹Możemy to zrobić przez 28.2.1

Twierdzenie 3.1.4. Niech $X_1, \dots, X_n$ będą parami niezależne. Wtedy

$$\text{Var}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{Var}[X_i]$$

Dowód. Skoro nasze zmienne są parami niezależne, to dla dowolnych $X_i \neq X_j$ mamy $\text{Cov}(X_i, X_j) = 0$. W takim razie

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^n X_i\right] &= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right)^2\right] \\ &= \sum_{i=1}^n \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] + \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[(X_i - \mathbb{E}[X_i]) \cdot (X_j - \mathbb{E}[X_j])] \\ &= \sum_{i=1}^n \text{Var}[X_i] + \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}[X_i] \end{aligned}$$

□

3.2 Nierówność Markowa

Twierdzenie 3.2.1. Jeśli X jest zmienną losową, która przyjmuje nieujemne wartości to

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

Dowód. Niech I będzie indykátorem

$$I = \begin{cases} 1 & \text{gdy } X \geq a \\ 0 & \text{wpp.} \end{cases}$$

Skoro $X \geq 0$ to $I \leq \frac{X}{a}$. Zatem

$$P(X \geq a) = P(I = 1) = \mathbb{E}[I] \leq \frac{\mathbb{E}[X]}{a}$$

□

3.3 Nierówność Czebyszewa

3.3.1 Definicja

Twierdzenie 3.3.1 (Twierdzenie 3.6 P&C). Dla dowolnego $a > 0$

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

Dowód. Korzystamy z nierówności Markowa

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

□

3.3.2 Kolekcjoner kuponów

Niech $X_1, \dots, X_n$ opisują czasy czekania na i -ty kupon oraz $X = \sum X_i$ – łączny czas czekania.

Aby w ogóle móc liczyć coś nierównością Czebyszewa potrzebujemy obliczyć $\text{Var}[X]$.

Skorzystamy tutaj z bardzo wygodnego twierdzenia 3.1.4 a następnie z 1.2.1 aby dostać

$$\begin{aligned} \text{Var}[X] &= \sum_{i=1}^n \text{Var}[X_i] \\ &= \sum_{i=1}^n \frac{1 - p_i}{p_i^2} \\ &\leq \sum_{i=1}^n \frac{1}{p_i^2} \\ &= \sum_{i=1}^n \left(\frac{n}{n - i + 1} \right)^2 \\ &= n^2 \cdot \sum_{i=1}^n \frac{1}{i^2} \\ &\leq n^2 \frac{\pi^2}{6} \end{aligned}$$

Teraz wkładamy to do nierówności Czebyszewa:

$$P(|X - nH_n| \geq nH_n) \leq \frac{\text{Var}[X]}{n^2 H_n^2} = \frac{\pi^2}{6 H_n^2} = O\left(\frac{1}{\ln^2 n}\right)$$

Pytanie 4

Ogólny schemat nierówności Czernowa. Nierówność Czernowa dla sum niezależnych prób Poissona. Zastosowanie tej nierówności: niezależne rzuty sprawiedliwą monetą.

4.1 Nierówność Czernowa

4.1.1 Definicja

Łączymy ze sobą dwie rzeczy – funkcje tworzące momenty, oraz nierówność Markowa.

Twierdzenie 4.1.1.

$$\forall_{t>0} P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$$

oraz

$$\forall_{t<0} P(X \leq a) = P(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$$

w szczególności

$$P(X \geq a) \leq \min_{t>0} \left\{ \frac{\mathbb{E}[e^{tx}]}{e^{ta}} \right\}$$

Dowód. Niezależnie od tego jakie wartości przyjmuje X oraz ile wynosi t to e^{tX} oraz e^{ta} zawsze będą dodatnie. Monotoniczność e^{tx} przy ustalonym t zależy jedynie od znaku zatem przejścia między prawdopodobieństwami zachodzą.

Ograniczenie górne uzyskujemy korzystając z nierówności Markowa zastosowanej do (dodatnich) wartości e^{tX} oraz e^{ta} . □

4.1.2 Próby Poissona

Definicja 4.1.1. Próbami Poissona nazywany ciąg zmiennych losowych $X_1, \dots, X_n$, dla których

$$P(X_i = 1) = p_i \wedge P(X_i = 0) = 1 - p_i$$

Ponadto definiujemy

$$\mu = \mathbb{E}\left[\sum X_i\right] = \sum \mathbb{E}[X_i] = \sum p_i$$

Dodatkowo, jeżeli $\forall_{i,j \in [n]} p_i = p_j$, to nazywamy to **próbami Bernoulliego**.

Twierdzenie 4.1.2. Niech $X_1, \dots, X_n$ to niezależne próby Poissona. Dodatkowo oznaczamy $X = \sum_{i=1}^n X_i$ i $\mu = \mathbb{E}[X]$. Wtedy

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

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

$$3. \text{ jeśli } R \geq 6\mu, \text{ to } P(X \geq R) \leq 2^{-R}$$

Dowód. Liczymy funkcję tworzącą

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

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

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

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 $\delta = \frac{R}{\mu} - 1 \geq 5$.

$$P(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu \leq \left(\frac{e}{1 + \delta} \right)^{(1+\delta)\mu} \leq \left(\frac{e}{6} \right)^R \leq \left(\frac{1}{2} \right)^R = 2^{-R}.$$

□

4.1.3 Rzuty monetą

Przykład 4.1.1. Chcemy ograniczyć z góry prawdopodobieństwo, że przy n rzutach monetą wyrzucimy orła więcej niż $\frac{3}{4}n$ razy. Widzimy, że nasze rzuty to niezależne próby Poissona o $p = \frac{1}{2}$, $\mu = \frac{n}{2}$, $\delta = \frac{1}{2}$, a więc możemy użyć wzoru 2. W takim razie mamy

$$P\left(X \geq \frac{3}{4}n\right) \leq e^{\frac{-\mu\delta^2}{3}} = e^{\frac{-\frac{n}{2} \cdot \frac{1}{4}}{3}} = e^{\frac{-n}{24}}$$

Pytanie 5

Kule i urny: obciążenie najcięższej urny prawie zawsze jest co najwyżej $\frac{3 \ln n}{\ln \ln n}$.

5.1 Kule i urny - ograniczenie górne

Zanim zaczniemy, zaprezentujemy dwa proste lematy potrzebne w oszacowaniu

Lemat 5.1.1. Dla dowolnych $n \geq M$

$$\binom{n}{M} \left(\frac{1}{n}\right)^M \leq \frac{1}{M!}$$

Dowód.

$$\binom{n}{M} \left(\frac{1}{n}\right)^M = \frac{n!}{M! \cdot (n-M)! n^M} = \frac{1}{M!} \cdot \frac{(n-M+1) \cdots n}{n^M} \leq \frac{1}{M!}$$

□

Lemat 5.1.2. Dla dowolnego n

$$\frac{1}{n!} \leq \left(\frac{e}{n}\right)^n$$

Dowód. Korzystamy z rozwinięcia e^k w szereg Taylora:

$$e^k = \sum_{i=0}^{\infty} \frac{k^i}{i!} > \frac{k^k}{k!}$$

Przekształcając otrzymujemy

$$\frac{e^k}{k^k} > \frac{1}{k!}$$

co daje nierówność z tezy.

□

Rozważmy bardzo prosty model - wrzucamy sobie n kul do n urn niezależnie i jednostajnie. Oczywiście średnio w jednej urnie spodziewamy się zobaczyć jedną kulę, ale ile spodziewamy

się zobaczyć kul w najbardziej zapełnionej urnie? Na to pytanie odpowiemy twierdzeniem.

Twierdzenie 5.1.1 (Lemat 5.1 P&C). Jeśli wrzucamy n kul do n urn to prawdopodobieństwo, że najcięższa urna zawiera **co najmniej** $M = \frac{3 \ln n}{\ln \ln n}$ kul wynosi co najwyżej $\frac{1}{n}$ dla odpowiednio dużych n .

Dowód. Nie ma co się zrażać mnogością logarytmów; sam w sobie dowód jest względnie prosty – stosujemy dwa razy *union-bound*, a ograniczenie z tezy po prostu pałujemy naszymi lematami, na egzaminie raczej nie będziecie potrzebowali obliczeń.

Prawdopodobieństwo, że ustalony podzbiór M kul wyląduje w ustalonej urnie wynosi $\left(\frac{1}{n}\right)^M$. Różnym podzbiorów jest $\binom{n}{M}$, zatem z union bounda dostajemy ograniczenie na prawdopodobieństwo, że w ustalonej urnie jest co najmniej M kul wynosi

$$\binom{n}{M} \cdot \left(\frac{1}{n}\right)^M$$

Korzystamy teraz z obu lematów i ograniczamy prawdopodobieństwo na to, że istnieje urna w której jest co najmniej M kul przez co najwyżej

$$n \left(\frac{e}{M}\right)^M$$

Teraz wstawiamy magiczne M z tezy i dostajemy:

$$n \left(\frac{e}{M}\right)^M \leq n \left(\frac{e \ln \ln n}{3 \ln n}\right)^{(3 \ln n)/(\ln \ln n)}$$

Zauważamy, że $e \leq 3$

$$\leq n \left(\frac{\ln \ln n}{\ln n}\right)^{(3 \ln n)/(\ln \ln n)}$$

Aby pokazać postulowaną w tezie nierówność bierzemy obustronnie logarytm

$$\begin{aligned} n \left(\frac{\ln \ln n}{\ln n}\right)^{(3 \ln n)/(\ln \ln n)} &\leq \frac{1}{n} \\ \ln n + ((\ln \ln \ln n) - (\ln \ln n)) \left(\frac{3 \ln n}{\ln \ln n}\right) &\leq -\ln n \end{aligned}$$

Wymnażamy i przenosimy na jedną stronę

$$-\ln n + \frac{3(\ln n)(\ln \ln \ln n)}{\ln \ln n} \leq 0$$

Sprowadzamy do wspólnego mianownika

$$\frac{(\ln n) \cdot (3(\ln \ln \ln n) - (\ln \ln n))}{\ln \ln n} \leq 0$$

Ponieważ $\ln n$ i $\ln \ln n$ są od pewnego momentu dodatnie to nierówność sprowadza się do pokazania, że

$$\ln \ln n \geq 3 \ln \ln \ln n$$

co już jest trywialne.

□

Pytanie 6

Rozkład Poissona i jego własności: momenty, suma niezależnych zmiennych, tworząca momentów i ograniczenia Chernowa.

6.1 Rozkład Poissona

6.1.1 Podstawowe własności

Definicja 6.1.1. Mówimy, że zmienna losowa X ma rozkład **Poissona** z parametrem λ jeśli

$$\forall_{n \in \mathbb{N}} P(X = n) = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$$

Aby upewnić się, że jest to poprawny rozkład policzmy $\sum_{n=0}^{\infty} P(X = n)$

$$\begin{aligned} \sum_{n=0}^{\infty} P(X = n) &= \sum_{n=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \cdot e^{\lambda} = 1 \end{aligned}$$

Twierdzenie 6.1.1. Niech X ma rozkład Poissona z parametrem λ . Wtedy

$$\mathbb{E}[X] = \lambda$$

Dowód.

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{n=0}^{\infty} n \cdot P(X = n) \\
 &= \sum_{n=0}^{\infty} n \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} \\
 &= e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\
 &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{n!} \\
 &= \lambda e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \\
 &= \lambda
 \end{aligned}$$

□

Twierdzenie 6.1.2 (Lemat 5.3 P&C). Jeśli zmienna X ma rozkład Poissona z parametrem λ to

$$M_X(t) = \exp(\lambda(e^t - 1))$$

Dowód.

$$\begin{aligned}
 M_X(t) &= \mathbb{E}[e^{tX}] \\
 &= \sum_{n=0}^{\infty} e^{tn} \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!} \\
 &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\
 &= \exp(-\lambda) \cdot \exp(\lambda e^t) \\
 &= \exp(\lambda(e^t - 1))
 \end{aligned}$$

W przedostatnim przejściu korzystamy z faktu, że $\sum_{n=0}^{\infty} \frac{x^n}{n!} = \exp(x)$

□

Twierdzenie 6.1.3 (Lemat 5.2 P&C). Jeśli X ma rozkład Poissona z parametrem λ_X a Y rozkład Poissona z parametrem λ_Y , a ponadto obie zmienne są niezależne to $X + Y$ ma rozkład Poissona z parametrem $\lambda_X + \lambda_Y$

Dowód. Ponieważ X i Y są niezależne to

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \cdot M_Y(t) \\ &= \exp(\lambda_X(e^t - 1)) \cdot \exp(\lambda_Y(e^t - 1)) \\ &= \exp((\lambda_X + \lambda_Y)(e^t - 1)) \end{aligned}$$

Skoro rozkład zmiennej $X+Y$ tworzony jest przez funkcję, która wygląda jak rozkład Poissona, to musi być ona rozkładem Poissona z parametrem $\lambda_X + \lambda_Y$ $\square$

Twierdzenie 6.1.4. Niech X ma rozkład Poissona z parametrem λ . Wtedy

$$\text{Var}[X] = \lambda$$

Dowód. Liczymy drugą pochodną $M_X(t) = \exp(\lambda(e^t - 1))$ i wychodzi. $\square$

6.1.2 Ograniczenia Czernowa

Twierdzenie 6.1.5. Niech X będzie zmienną o rozkładzie Poissona z parametrem μ . Wtedy:

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

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

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. $\square$

Pytanie 7

Aproksymacja Poissona oraz jej zastosowanie do problemu kul i urn: obciążenie najcięższej urny jest prawie zawsze co najmniej $\frac{\ln n}{\ln \ln n}$.

7.1 Aproksymacja Poissona

7.1.1 Definicja

Czasem mamy do czynienia ze zmiennymi, które pojedynczo zachowują się grzecznie, ale jako całość są powiązane w sposób, który istotnie utrudnia ich analizę. Z pomocą przychodzi Aproksymacja Poissona, w której uniezależnimy wszystkie zmienne, a następnie będziemy analizować ich zachowanie pod pewnymi warunkami.

Bardziej formalnie opisuje to poniższe twierdzenie.

Twierdzenie 7.1.1 (Twierdzenie 5.6 P&C). Niech

$$X_1^{(k)}, \dots, X_n^{(k)}$$

opisują (faktyczne) rozmieszczenie k kul w n urnach.

Ponadto, niech

$$Y_1^{(m)}, \dots, Y_n^{(m)}$$

będą niezależnymi zmiennymi z rozkładem Poissona z parametrem $\lambda = \frac{m}{n}$

Wtedy

$$\forall_m : P\left(X_1^{(k)} = k_1, \dots, X_n^{(k)} = k_n\right) = P\left(Y_1^{(m)} = k_1, \dots, Y_n^{(m)} = k_n \mid \sum_{i=1}^n Y_i^{(m)} = k\right)$$

Dowód. Policzmy najpierw lewą stronę równości

$$\begin{aligned} P\left(X_1^{(k)} = k_1, \dots, X_n^{(k)} = k_n\right) &= \frac{1}{n^k} \cdot \binom{k}{k_1} \cdot \binom{k-k_1}{k_2} \cdot \dots \cdot \binom{k_n}{k_n} \\ &= \frac{k!}{k_1! \cdot \dots \cdot k_n! \cdot n^k} \end{aligned}$$

Policzmy teraz prawą stronę

$$P\left(Y_1^{(m)} = k_1, \dots, Y_n^{(m)} = k_n \mid \sum_{i=1}^n Y_i^{(m)} = k\right) = \frac{P\left(Y_1^{(m)} = k_1 \wedge \dots \wedge Y_n^{(m)} = k_n\right)}{P\left(\sum Y_i^{(m)} = k\right)}$$

Korzystamy z faktu, że nasze zmienne są niezależne, oraz suma n Poissonów z parametrem $\lambda = \frac{m}{n}$ ma rozkład Poissona z parametrem m

$$\begin{aligned} &= \left(\prod_{i=1}^n e^{-\lambda} \cdot \frac{\lambda^{k_i}}{k_i!}\right) \cdot \frac{k!}{e^{-m} m^k} = \frac{k!}{k_1! \cdot \dots \cdot k_n!} \cdot \frac{e^{-n\lambda} \cdot \lambda^k}{e^{-m} m^k} \\ &= \frac{k!}{k_1! \cdot \dots \cdot k_n!} \cdot \frac{e^{-m} \cdot \left(\frac{m}{n}\right)^k}{e^{-m} \cdot m^k} \\ &= \frac{k!}{k_1! \cdot \dots \cdot k_n! \cdot n^k} \end{aligned}$$

Po obu stronach wyszło to samo, fajnie. □

Skoro umiemy zamieniać kule i urny na warunkowe Poissony to fajnie byłoby coś umieć o nich powiedzieć.

Twierdzenie 7.1.2. Niech $f(x_1, \dots, x_n)$ będzie funkcją zwracającą nieujemne wartości. Wtedy

$$\mathbb{E}\left[f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right] \leq e\sqrt{m} \cdot \mathbb{E}\left[f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right)\right]$$

Dowód.

$$\begin{aligned} \mathbb{E}\left[f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right)\right] &= \sum_{k=0}^{\infty} \mathbb{E}\left[f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right) \mid \sum Y_i^{(m)} = k\right] \cdot P\left(\sum Y_i^{(m)} = k\right) \\ &\geq \mathbb{E}\left[f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right) \mid \sum Y_i^{(m)} = m\right] \cdot P\left(\sum Y_i^{(m)} = m\right) \\ &= \mathbb{E}\left[f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right] \cdot e^{-m} \cdot \frac{m^m}{m!} \\ &\geq \mathbb{E}\left[f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right] \cdot \frac{1}{e\sqrt{m}} \end{aligned}$$

□

7.1.2 Kule i urny

W twierdzeniu 5.1.1 pokazaliśmy, że górne ograniczenie na liczbę kul w najcięższej urnie to z dużym prawdopodobieństwem $O\left(\frac{\ln n}{\ln \ln n}\right)$.

Teraz pokażemy, że dolne ograniczenie to z dużym prawdopodobieństwem $\Omega\left(\frac{\ln n}{\ln \ln n}\right)$.

Twierdzenie 7.1.3 (Lemat 5.12 P&C). Dla wystarczająco dużego n , jeśli wrzucamy n kul do n urn to prawdopodobieństwo, że najcięższa urna zawiera **co najwyżej** $M = \frac{\ln n}{\ln \ln n}$ kul wynosi co najwyżej $\frac{1}{n}$.

Dowód. Rozważmy tę sytuację w modelu Poissona – liczba kul w ustalonej urnie ma rozkład Poissona z parametrem $\lambda = \frac{n}{n} = 1$.

W takim razie, prawdopodobieństwo, że ustalona urna zawiera co najmniej M kul wynosi

$$\sum_{k=\lceil M \rceil}^{\infty} e^{-1} \cdot \frac{1^k}{k!} \geq \frac{1}{eM!}$$

Prawdopodobieństwo, że każda urna zawiera mniej niż M kul wynosi zatem co najwyżej

$$\left(1 - \frac{1}{eM!}\right)^n \leq \exp\left(-\frac{n}{eM!}\right)$$

Z faktu, że $1 - x \leq e^{-x}$.

Jeśli nasze M jest na tyle fajne, że zachodzi

$$\exp\left(-\frac{n}{eM!}\right) \leq \frac{1}{n^2}$$

to wtedy na mocy twierdzenia 7.1.2 prawdopodobieństwo, że w prawdziwym modelu każda urna ma mniej niż M kul wynosi co najwyżej

$$e\sqrt{n} \cdot \frac{1}{n^2} < \frac{1}{n}$$

Pozostaje pokazać, że $M = \frac{\ln n}{\ln \ln n}$ jest wystarczające dla dużych n .

Bierzemy zatem obustronnie logarytm z zadanego warunku

$$\begin{aligned} -\frac{n}{eM!} &\leq -2 \ln n \\ \frac{n}{eM!} &\geq 2 \ln n \\ \frac{n}{2e \ln n} &\geq M! \end{aligned}$$

Znowu bierzemy logarytm obustronnie (bo możemy, lol)

$$\ln n - \ln \ln n - \ln(2e) \geq \ln(M!)$$

Wykorzystamy teraz *magiczne oszacowanie*

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

i dostajemy

$$\begin{aligned} \ln(M!) &\leq \ln M + M \ln M - M \\ &= M \cdot ((\ln \ln n) - (\ln \ln \ln n)) + \ln M - M \\ &= (M \cdot (\ln \ln n) - M) - (M \cdot (\ln \ln \ln n) - \ln M) \\ &= (\ln n - M) - (M \cdot (\ln \ln \ln n) - \ln M) \end{aligned}$$

Teraz korzystamy z faktu, że $\ln M \in o(M \cdot (\ln \ln \ln n))$

$$\leq (\ln n - M) = \ln n - \frac{\ln n}{\ln \ln n}$$

I jeszcze korzystamy z faktu, że $(\ln \ln n)^2 \in o(\ln n)$ a zatem $\ln \ln n \in o\left(\frac{\ln n}{\ln \ln n}\right)$. Możemy więc zamienić $\frac{\ln n}{\ln \ln n}$ na $\ln \ln n + \ln 2e$

$$\leq \ln n - \ln \ln n - \ln(2e)$$

czyli nasze M działa. Uff. □

Pytanie 8

Problem kolekcjonera kuponów: granica prawdopodobieństwa, że nie zbierzemy wszystkich n kuponów po $n \ln n + cn$ krokach.

8.1 Granica kolekcjonera kuponów

Twierdzenie 8.1.1. Niech X będzie liczbą zebranych kuponów aż do zebrania wszystkich n rodzajów. Wtedy dla dowolnej stałej c

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

Dowód. O zbieraniu kuponów możemy myśleć jak o wrzucaniu kul do urn – wrzucenie kuli do odpowiedniej urny odpowiada zebraniu odpowiedniego kuponu.

Będziemy zatem liczyć prawdopodobieństwo, że po wrzuceniu $m = n \ln n + cn$ kul do n urn jakaś urna nadal pozostaje pusta.

Rozważmy ten problem w modelu Poissona, a potem pokażemy jak wyciągnąć z tego wynik dla rzeczywistego modelu. Mamy zatem $\lambda = \frac{m}{n} = \ln n + c$

Prawdopodobieństwo, że ustalona urna jest pusta wynosi

$$e^{-\lambda} \cdot \frac{\lambda^0}{0!} = e^{-(\ln n + c)} = \frac{e^{-c}}{n}$$

Ponieważ w modelu Poissona urny są niezależne to prawdopodobieństwo, że żadna urna nie jest pusta (czyli każda ma co najmniej jedną kulę) wynosi

$$\left(1 - \frac{e^{-c}}{n}\right)^n$$

Nazwijmy to zdarzenie $\mathcal{E}$. Z powyższego faktu mamy

$$\lim_{n \rightarrow \infty} P(\mathcal{E}) = e^{-e^{-c}}$$

Wszystko fajnie i w ogóle, ale my byśmy chcieli dostać rzeczywiste prawdopodobieństwo $\mathcal{E}$, którego nie możemy sobie tak po prostu przenieść z Poissona na rzeczywisty model, bo pamiętamy z twierdzenia 7.1.1, że wolno nam jedynie przejść równością warunkową tj. $P(\mathcal{E} \mid X = m)$, a tego nie znamy.

Aby sobie z tym poradzić rozbijemy nasze zdarzenie $\mathcal{E}$ na dwie części. Ustalamy $\delta = \sqrt{m \ln m}$ i rozbijamy za pomocą prawdopodobieństwa całkowitego:

$$P(\mathcal{E}) = P(\mathcal{E} \mid |X - m| \leq \delta) \cdot P(|X - m| \leq \delta) + P(\mathcal{E} \mid |X - m| > \delta) \cdot P(|X - m| > \delta)$$

Teraz chcemy pokazać dwie rzeczy. Po pierwsze, że drugi składnik jest pomijalnie mały (zbiega do zera). Po drugie, że pierwszy składnik zbiega do $P(\mathcal{E} \mid X = m)$ czyli tego co próbujemy obliczyć.

1. Szacujemy $P(|X - m| > \delta)$ przy pomocy nierówności Czebyszewa.

Ponieważ X ma rozkład Poissona z parametrem $\mu = m$ to

$$\text{Var}[X] = \mu = m$$

W takim razie z nierówności Czebyszewa

$$P(|X - m| > \delta) \leq \frac{\text{Var}[X]}{\delta^2} = \frac{m}{m \ln m} = \frac{1}{\ln m} \in o(1)$$

2. Szacujemy różnicę między tym czego szukamy a tym co mamy:

$$|P(\mathcal{E} \mid |X - m| \leq \delta) - P(\mathcal{E} \mid X = m)|$$

Zauważamy dość naturalny fakt – im więcej kul wrzucamy tym większa szansa na to, że każda ma jakąś kulę. Innymi słowy

$$P(\mathcal{E} \mid X = m) \geq P(\mathcal{E} \mid X = m - \delta)$$

oraz

$$P(\mathcal{E} \mid |X - m| \leq \delta) \leq P(\mathcal{E} \mid X = m + \delta)$$

Możemy zatem zastąpić odpowiednie wyrażenia przez ich oszacowania aby dostać słabsze ograniczenie:

$$|P(\mathcal{E} \mid |X - m| \leq \delta) - P(\mathcal{E} \mid X = m)| \leq P(\mathcal{E} \mid X = m + \delta) - P(\mathcal{E} \mid X = m - \delta)$$

Wyrażenie po prawej stronie oddaje sytuację, kiedy wrzuciliśmy $m - \delta$ kul, ale nadal jakaś urna pozostaje pusta, natomiast po dorzuceniu kolejnych 2δ kul została ona zapełniona.

Prawdopodobieństwo, że konkretna kula trafi do konkretnej pustej urny wynosi $\frac{1}{n}$, zatem prawdopodobieństwo, że jakaś kula trafi do tej urny jest ograniczone przez union bound:

$$P(\mathcal{E} \mid X = m + \delta) - P(\mathcal{E} \mid X = m - \delta) \leq \frac{2\delta}{n} = \frac{2\sqrt{m \ln m}}{n} = 2\sqrt{\frac{m \ln m}{n^2}}$$

Przypominamy sobie, że $m = n \ln n + cn$, zatem $m \ln m \in o(n^2)$. W takim razie nasze oszacowanie zbiega do zera, a co za tym idzie, szacowana różnica też.

Korzystając z powyższych faktów, dochodzimy do wniosku, że

$$\begin{aligned} \lim_{n \rightarrow \infty} P(\mathcal{E}) &= P(\mathcal{E} \mid |X - m| \leq \delta) \cdot P(|X - m| \leq \delta) + P(\mathcal{E} \mid |X - m| > \delta) \cdot P(|X - m| > \delta) \\ &= \lim_{n \rightarrow \infty} P(\mathcal{E} \mid |X - m| \leq \delta) \cdot (1 - o(1)) + P(\mathcal{E} \mid |X - m| > \delta) \cdot o(1) \\ &= \lim_{n \rightarrow \infty} (P(\mathcal{E} \mid X = m) + o(1)) \cdot (1 - o(1)) \\ &= \lim_{n \rightarrow \infty} P(\mathcal{E} \mid X = m) \end{aligned}$$

A to jest dokładnie to co chcieliśmy pokazać. □

Pytanie 9

Łańcuch Markowa. Nieprzywiedlność, okres stanu i okres łańcucha. Prawdopodobieństwa przejść pomiędzy stanami w nieprzywiedlnym i nieokresowym łańcuchu Markowa. Stany powracające (dodatnie i zerowe) i chwilowe. W każdym nieprzywiedlnym, skończonym łańcuchu Markowa oczekiwany czas przejścia pomiędzy dwoma stanami jest skończony.

9.1 Definicja łańcucha Markowa

Definicja 9.1.1. Procesem stochastycznym nazywamy dowolny zbiór zmiennych losowych $\{X_t : t \in T\}$. Zwykle t oznacza moment w czasie, a X_t jest **stanem** tego procesu w czasie t . Zbiór stanów często oznaczany jest jako S .

Definicja 9.1.2. Proces jest **skończony**, jeśli zmienne X_t przyjmują skończenie wiele wartości.

Definicja 9.1.3. Proces jest **dyskretny**, jeśli zmienne X_t przyjmują wartości ze zbioru przeliczalnego.

Definicja 9.1.4. Proces jest **z czasem dyskretnym**, jeśli T jest przeliczalne (najczęściej $T = \mathbb{N}$).

Definicja 9.1.5. Łańcuchem Markowa nazywamy proces stochastyczny z czasem dyskretnym $\{X_t\}_{t \in \mathbb{N}}$ dla którego

1. dla każdego $t \geq 0$ oraz $(a_0, \dots, a_t)$ takiego, że $P(\bigcap_{i=0}^t X_i = a_i) > 0$ zachodzi

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

2. dla każdego $t \geq 0$ i $x, y \in S$ zachodzi

$$P(X_{t+1} = x \mid X_t = y) = P(X_t = x \mid X_{t-1} = y)$$

Właściwość 1. mówi nam, że aby dostać rozkład zmiennej X_t wystarczy, że znamy rozkład

zmiennej X_{t-1} tzn. łańcuch Markowa jest bez pamięci. Warto zauważyć, że **nie oznacza to**, że X_t jest niezależne od $X_{t-2}, X_{t-3}, \dots$ – jest, ale cała ta zależność jest zawarta w zależności od stanu X_{t-1} .

Za właściwość 2. mówi nam, że bez znaczenia na czas, prawdopodobieństwo przejścia z określonego stanu x do stanu y jest zawsze takie samo.

Warto zaznaczyć że niektóre źródła definiują łańcuchy Markowa jako procesy spełniające wyłącznie właściwość 1., a procesy spełniające 1. oraz 2. nazywają łańcuchami Markowa czasu homogenicznego, jednak na probabilu dla uproszczenia terminologii używamy powyższej definicji.

Definicja 9.1.6. Oznaczamy dalej $p_{ij} = P(X_t = j \mid X_{t-1} = i)$.

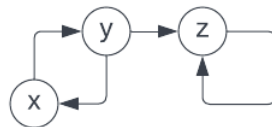
Definicja 9.1.7. Macierzą przejścia łańcucha Markowa nazywamy macierz P zadaną współczynnikami p_{ij} .

Definicja 9.1.8. $p_{ij}(n) = P(X_n = j \mid X_0 = i)$

9.2 Nieprzywiedlność i okres stanu

Definicja 9.2.1. Dla $i \in S$ definiujemy $\mathcal{T}(i) = \{t \geq 1 \mid p_{ii}(t) > 0\}$. Jest to zbiór takich t , że jesteśmy w stanie dojść z i do i w t krokach.

Definicja 9.2.2. Okres stanu dla $i \in S$ definiujemy jako $o(i) = \gcd(\mathcal{T}(i))$



Rysunek 9.1: Stan z ma okres 1

Definicja 9.2.3. Stan j jest **osiągalny** ze stanu i jeśli istnieje $n \geq 0$ takie, że $p_{ij}(n) > 0$. Zapisujemy $i \rightarrow j$.

Definicja 9.2.4. Stany i oraz j są **wzajemnie skomunikowane** jeśli i jest osiągalne z j oraz j jest osiągalne z i . Zapisujemy $i \leftrightarrow j$.

Lemat 9.2.1. Relacja skomunikowania jest relacją równoważności.

Dowód. Rozważamy trzy warunki bycia relacją równoważności

1. $i \leftrightarrow i$

Możemy dojść z i do i w 0 krokach – $p_{ii}(0) = 1$

$$2. i \leftrightarrow j \implies j \leftrightarrow i$$

Koniunkcja jest przemienna, możemy zatem zamienić kolejność warunków w definicji.

$$3. i \leftrightarrow j \wedge j \leftrightarrow k \implies i \leftrightarrow k$$

Skoro $i \leftrightarrow j$ to mamy n dla którego $P_{ij}(n) > 0$.

Podobnie mamy m dla którego $P_{jk}(m) > 0$.

W takim razie $p_{ik}(n+m) \geq p_{ij}(n) \cdot P_{jk}(m) > 0$ zatem k jest osiągalne z i .

Analogicznie pokazujemy, że i jest osiągalne z k , czyli stany te są skomunikowane.

□

Dodatkowo w grafie skierowanym klasy równoważności relacji $\leftrightarrow$ tworzą silnie spójne składowe.

Definicja 9.2.5. Łańcuch jest **nieprzywiedlny (nieredukowalny)** jeśli wszystkie stany są parami skomunikowane. Wtedy jego graf skierowany jest silnie spójny.

Definicja 9.2.6. Stan i jest **okresowy** jeśli $o(i) > 1$, czyli jego okres jest większy od 1. Łańcuch jest okresowy jeśli posiada co najmniej jeden stan okresowy. Stan lub łańcuch, który nie jest okresowy nazywamy **nieokresowym**.

9.3 Okres nieprzywiedlnego łańcucha Markowa

Lemat 9.3.1. W nieprzywiedlnym łańcuchu Markowa wszystkie stany mają ten sam okres.

Dowód. Niech $i, j \in S$ - dwa stany łańcucha.

Z nieprzywiedlności mamy:

$$\exists_m p_{ij}(m) > 0$$

$$\exists_l p_{ji}(l) > 0$$

Niech $n \in \mathcal{T}(j)$

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

$$p_{ii}(m+l) \geq p_{ij}(m) \cdot p_{ji}(l) > 0$$

To znaczy chcemy dojść z i do i w $m+n+l$ krokach, więc możemy iść z i do j , z j do j i z j do i . W drugim przypadku pomijamy n kroków z j do j .

$$m+n+l \in \mathcal{T}(i) \implies o(i) \mid m+n+l$$

$$m+l \in \mathcal{T}(i) \implies o(i) \mid m+l$$

Mamy teraz $o(i) \mid n$, a więc $o(i)$ jest dzielnikiem każdego elementu $\mathcal{T}(j)$, a więc jest $\leq$ od jego nwd, a więc $o(i) \leq o(j)$. Analogicznie dowodzimy w drugą stronę, otrzymując równość. □

9.4 Prawdopodobieństwa przejść pomiędzy stanami w nieprzywiedlnym i nieokresowym łańcuchu Markowa

Lemat 9.4.1. Jeśli $\{X_t \mid t \in \mathbb{N}\}$ jest nieokresowym i nieprzywiedlnym łańcuchem Markowa, to:

$$\forall_{i,j \in S} \exists_{n_0} \forall_{n \geq n_0} p_{ij}(n) > 0$$

Dowód. Lemat Schura:

$$\forall_{x \in \mathbb{N}^+, \gcd(x)=1} \exists_{m_0} \forall_{m \geq m_0} \exists_{r \geq 1, x_1, \dots, x_r \in x, l_1, \dots, l_r \in \mathbb{N}} \sum_{i \in [r]} l_i x_i = m$$

Niech $i, j \in S$. Z lematu Schura mamy:

$$\exists_{m_0} \forall_{m \geq m_0} \exists_{r \geq 1, m_1, \dots, m_r \in \mathcal{T}(j), l_1, \dots, l_r \in \mathbb{N}} \sum_{i \in [r]} l_i m_i = m$$

$$p_{jj}(m) \geq \prod_{i \in [r]} p_{jj}(l_i m_i) \geq \prod_{i \in [r]} p_{jj}(m_i)^{l_i} > 0$$

Z kolei z nieprzywiedlności mamy:

$$i \rightarrow j \implies \exists_{m_1} p_{ij}(m_1) > 0$$

Niech $n_0 = m_0 + m_1$

$$\forall_{n \geq n_0} p_{ij}(n) \geq p_{ij}(m_1) p_{jj}(n - m_1) > 0$$

□

9.5 Stany powracające i chwilowe

Definicja 9.5.1. Definiujemy **piwszy czas pojawienia się w j** jako:

$$T_j = \min(n \in \mathbb{N} \mid X_n = j)$$

$$T_j^+ = \min(n \in \mathbb{N}_1 \mid X_n = j)$$

W drugim przypadku pomijamy stan początkowy $n = 0$.

Definicja 9.5.2. Definiujemy **prawdopodobieństwo pierwszego spotkania w zadanym momencie** $f_{ij}(n)$ jako

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

Definicja 9.5.3. Definiujemy **prawdopodobieństwo pierwszego spotkania** f_{ij} jako

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

Definicja 9.5.4. Stan i jest **powracający (rekurencyjny)** jeśli $f_{i,i} = 1$, a **chwilowy** jeśli $f_{i,i} < 1$.

Mówimy, że łańcuch jest rekurencyjny jeśli każdy jego stan jest rekurencyjny.

Definicja 9.5.5. Definiujemy **czas pierwszego spotkania**

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

dodatkowo

$$\mathbb{E}[T_{i,j}] = \mathbb{E}[T_j^+ \mid X_0 = i] = \sum_{n=1}^{\infty} n P(T_j^+ = n \mid X_0 = i) = \sum_{n=1}^{\infty} n f_{ij}(n)$$

Definicja 9.5.6. Stan powracający i jest **dodatni** jeśli $\mathbb{E}[T_{i,i}] < \infty$, w przeciwnym wypadku jest **zerowy**.

9.6 Własność nieprzywiedlnego, skończonego łańcucha Markowa

Twierdzenie 9.6.1. W skończonym, nieprzywiedlnym łańcuchu Markowa zachodzi

$$\forall_{x,y \in S} \mathbb{E}[T_{x,y}] < \infty$$

Dowód. Nieprzywiedlnosc oraz skonczenosc daja nam

$$\exists_{r>0, \varepsilon>0} \forall_{x,y \in S} \exists_{j \in [r]} p_{x,y}(j) > \varepsilon$$

Mimo tego, że linia ta może początkowo być trudna do przetworzenia, jest całkiem prosta. Nieprzywiedlnosc mówi nam że dla każdego (x, y) , y jest osiągalne z x , a więc $\exists_n p_{x,y}(n) > 0$. Nasze r to po prostu maksimum po tych n dla wszystkich par (x, y) , a ε to minimum z wartości $p_{x,y}(n)$.

Następnie chcemy pokazać, że

$$P(T_y^+ > kr \mid X_0 = x) \leq P(T_y^+ > (k-1)r \mid X_0 = x)(1 - \varepsilon)$$

Dlaczego tak jest? Otóż wiemy, że dla każdej możliwej wartości $z = X_{(k-1)r}$ zachodzi

$$\exists_{j \in [r]} p_{z,y}((k-1)r + j) > \varepsilon$$

a więc z prawdopodobieństwem przynajmniej ε odwiedzimy y w następnych r krokach. W takim razie, jeśli $T_y^+ > (k-1)r$, to $P(T_y^+ \leq kr) \geq \varepsilon$, a więc prawdopodobieństwo tego, że nie dojdziemy do y jest ograniczone od góry przez $1 - \varepsilon$, co daje nam naszą nierówność.

Następnie, poprzez prostą indukcję można pokazać, że

$$P(T_y^+ > kr \mid X_0 = x) \leq (1 - \varepsilon)^k$$

Teraz, przechodząc do finalnego dowodu i korzystając z 28.3.1

$$\begin{aligned} \mathbb{E}[T_{x,y}] &= \mathbb{E}[T_y^+ \mid X_0 = x] \\ &= \sum_{t=0}^{\infty} P(T_y^+ > t \mid X_0 = x) \\ &\leq \sum_{k=0}^{\infty} r P(T_y^+ > kr \mid X_0 = x) \\ &\leq r \sum_{k=0}^{\infty} (1 - \varepsilon)^k \\ &< \infty \end{aligned}$$

□

Pytanie 10

Rozkład stacjonarny. Istnienie, unikalność oraz interpretacja.

10.1 Rozkład stacjonarny

Definicja 10.1.1. Rozkładem stacjonarnym nazywamy wektor π taki, że $\pi = \pi P$ oraz $\sum_{i \in S} \pi_i = 1$

Intuicyjnie rozkład stacjonarny opisuje jak często asymptotycznie odwiedzamy każdy ze stanów niezależnie od tego skąd zaczęliśmy. Rozkład stacjonarny nie zawsze istnieje - np. łańcuch na liczbach naturalnych, taki, że $p_{n,n+1} = 1$ w oczywisty sposób nie ma rozkładu stacjonarnego.

Twierdzenie 10.1.1. Istnienie rozkładu stacjonarnego

Niech $z \in S$, $\overline{\pi}_z = (\overline{\pi}_{z,y})_{y \in S}$. Dodatkowo niech

$$\mathbb{E}_z[A] := \mathbb{E}[A \mid X_0 = z]$$

$$P_z(A) := P(A \mid X_0 = z)$$

$$\overline{\pi}_{z,y} := \mathbb{E}_z[\text{liczba wizyt w } y \text{ przed pierwszym powrotem do } z] = \sum_{t=0}^{\infty} P_z(X_t = y \wedge T_z^+ > t)$$

Dla $z \in S$, $\mathbb{E}_z[T_z^+] < \infty$ zachodzi:

- $\overline{\pi}_z = \overline{\pi}_z P$
- $\pi = \frac{\overline{\pi}_z}{\mathbb{E}_z[T_z^+]}$ jest rozkładem stacjonarnym

Dowód. Druga część prosto wynika z pierwszej, ponieważ oczywiście z definicji $\overline{\pi}_{z,y}$ mamy

$$\sum_{y \in S} \overline{\pi}_{z,y} = \mathbb{E}_z[T_z^+]$$

Pozostaje nam więc tylko udowodnić część pierwszą

$$\begin{aligned}
\sum_{x \in S} \overline{\pi_{z,x}} p_{x,y} &= \sum_{x \in S} \sum_{t=0}^{\infty} P_z(X_t = x \wedge T_z^+ > t) p_{x,y} \\
&= \sum_{t=0}^{\infty} P_z(X_{t+1} = y \wedge T_z^+ \geq t+1) \\
&= \sum_{t=1}^{\infty} P_z(X_t = y \wedge T_z^+ \geq t) \\
&= \overline{\pi_{z,y}} - P_z(X_0 = y \wedge T_z^+ > 0) + \sum_{t=1}^{\infty} P_z(X_t = y \wedge T_z^+ = t) \\
&= \overline{\pi_{z,y}} - P_z(X_0 = y) + P_z(X_{T_z^+} = y) \\
&= \overline{\pi_{z,y}}
\end{aligned}$$

$P_z(X_0 = y) = P_z(X_{T_z^+} = y)$ ponieważ oba są indykatorami $y = z$. $\square$

Twierdzenie 10.1.2. Skończony, nieprzywiedlny łańcuch Markowa ma unikalny rozkład stacjonarny

Dowód. Weźmy rozkłady stacjonarne π, ϕ . Ustalmy $x \in S$ taki, że $\frac{\pi_x}{\phi_x}$ jest najmniejsze (możemy to zrobić, ponieważ łańcuch jest skończony). Z definicji rozkładu stacjonarnego

$$\pi_x = \sum_{y \in S} \pi_y p_{y,x} = \sum_{y \in S} \frac{\pi_y}{\phi_y} \phi_y p_{y,x} \geq \sum_{y \in S} \frac{\pi_x}{\phi_x} \phi_y p_{y,x} = \frac{\pi_x}{\phi_x} \sum_{y \in S} \phi_y p_{y,x} = \frac{\pi_x}{\phi_x} \phi_x = \pi_x$$

To, że po obu stronach mamy to samo mówi nam, że powyższa nierówność to tak naprawdę równość. W takim razie wiemy, że

$$\forall_{y, p_{y,x} > 0} \quad \frac{\pi_y}{\phi_y} = \frac{\pi_x}{\phi_x}$$

Widzimy więc, że wszystkie stany z których da się bezpośrednio dojść do x mają taki sam iloraz π do ϕ . Możemy następnie analogicznie pokazać, że wszystkie stany z których da się dojść w 2, 3, ... krokach do x mają taki sam iloraz. Ponieważ łańcuch jest nieprzywiedlny i skończony, z każdego stanu da się dojść w skończonej liczbie kroków do x . W takim razie wszystkie stany mają taki sam iloraz, a więc

$$\pi = \frac{\pi_x}{\phi_x} \phi$$

A więc $\pi = \phi$, bo $\sum_{i \in S} \pi_i = \sum_{i \in S} \phi_i = 1$ $\square$

Twierdzenie 10.1.3. Każdy skończony, nieprzywiedlny łańcuch Markowa

1. Ma unikalny rozkład stacjonarny $\pi = (\pi_i)_{i \in S}$
2. $\forall_{i \in S} \quad \pi_i = \frac{1}{\mathbb{E}[T_i^+ | X_0 = i]}$

Dodatkowo, jeśli łańcuch jest nieokresowy

$$3. \forall_{i,j \in S} \lim_{t \rightarrow \infty} p_{j,i}(t) = \pi_i$$

Dowód. Punkt 1. udowodniliśmy przed chwilą. Po chwili zastanowienia, punkt 2. prosto z niego wynika. Dla danego i , z twierdzenia 10.1.1 wiemy, że istnieje rozkład stacjonarny π w którym

$$\pi_i = \frac{\sum_{t=0}^{\infty} \mathbb{P}(X_t = i \wedge T_i^+ > t \mid X_0 = i)}{\mathbb{E}[T_i^+ \mid X_0 = i]} = \frac{1}{\mathbb{E}[T_i^+ \mid X_0 = i]}$$

Ponieważ rozkład stacjonarny jest unikalny, to jest to prawdziwe dla każdego i .

Za to do dowodu punktu 3. będziemy potrzebowali więcej narzędzi i można go znaleźć w 13.1.1. □

Pytanie 11

Norma całkowitego wahanía rozkładów prawdopodobieństwa, własności normy, sprzężanie rozkładów prawdopodobieństwa. Związek między normą a sprzężaniem

11.1 Norma całkowitego wahanía

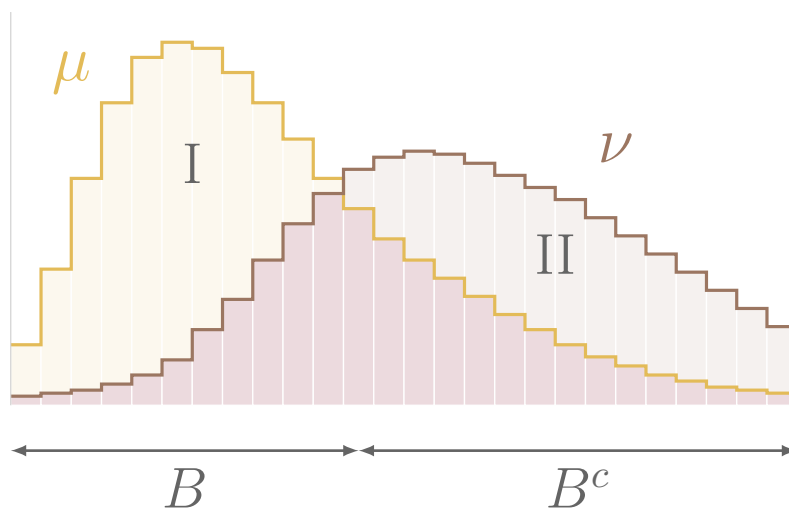
Definicja 11.1.1. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S . Normą całkowitego wahanía (total variation distance) tych rozkładów nazywamy wartość

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

Lemat 11.1.1. 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. Spróbujmy przekazać intuicję tego, czym jest norma całkowitego wahanía. Poniżej mamy wykres, na którym zaznaczone są rozkłady μ oraz ν oraz zbiory B i B^c



Wiemy, że pole pod $\mu = 1$ oraz pole pod $\nu = 1$. W takim razie, możemy zauważyć że pola I i II są sobie wzajemnie równe. Dodatkowo, patrząc na definicję normy całkowitego wahanía, można prosto zauważyć, że jest ona równa polu I (bo B to zbiór w którym μ najbardziej dominuje nad ν) oraz polu II (analogicznie). Prosto widzimy, że

$$\text{Pole I} = \mu(B) - \nu(B)$$

$$\text{Pole II} = \nu(B^c) - \mu(B^c)$$

Co kończy dowód □

Lemat 11.1.2. 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 poprzedniego lematu dostajemy, że dla $B = \{x \in S : \mu(x) \geq \nu(x)\}$ jest

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

□

11.2 Sprzęganie rozkładów prawdopodobieństwa

Definicja 11.2.1. 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.

Lemat 11.2.1. 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 usta-

lamy $X = x$ 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. $\square$

Pytanie 12

Lemat o sprzęganiu łańcuchów Markowa, lemat o monotoniczności.

12.1 Sprzęganie łańcuchów Markowa

Będziemy rozważać łańcuch Markowa $(X_t)_{t \in \mathbf{N}}$ (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 12.1.1. Definiujemy

$$\Delta_x(t) = \|P^t(x, \cdot) - \pi\|_{TV}$$

$$\tau_x(\varepsilon) = \min \{t : \Delta_x(t) \leq \varepsilon\}$$

Mamy też maksima tych wartości:

$$\Delta(t) = \max_{x \in S} \Delta_x(t)$$

$$\tau_{\text{mix}}(\varepsilon) = \max_{x \in S} \tau_x(\varepsilon)$$

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

Definicja 12.1.2. Sprzęganiem łańcuchów Markowa X, Y o macierzy przejścia P i zbiorze stanów S jest dowolny łańcuch Markowa $(Z_t = (X_t, Y_t))_{t \in \mathbf{N}}$ 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.

12.2 Lemat o sprzęganiu łańcuchów Markowa

Lemat 12.2.1. 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

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

□

12.3 Lemat o monotoniczności

Lemat 12.3.1 (O monotoniczności). Niech P będzie macierzą przejścia skończonego, nieprzywiedlnego i nieokresowego łańcucha Markowa ze zbiorem stanów S i rozkładem stacjonarnym π . Dla każdych $t \geq 0, x \in S$ zachodzi

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

Dowód. Ustalmy $t \geq 0$ i $x \in S$. Niech (X_t, Y_t) będzie sprzęganiem rozkładów $P^t(x, \cdot)$ i π spełniającym $P(X_t \neq Y_t) = \Delta_x(t)$ (przedtem pokazaliśmy, że istnieje sprzęganie, 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

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

□

Pytanie 13

Twierdzenie o geometrycznej zbieżności rozkładu stanu łańcucha do rozkładu stacjonarnego.

13.1 Twierdzenie o geometrycznej zbieżności

Twierdzenie 13.1.1 (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

$$\forall_{n \in \mathbf{N}} \Delta(n) \leq C\alpha^n.$$

Dowód. Ustalmy $r \geq 1$ takie, że dla każdych $x, y \in S$ 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 $((X_t, Y_t))_{t \in \mathbf{N}}$ będzie sprzężaniem łańcuchów o macierzy przejścia P^r zadany w następujący sposób: Niech $(X_t, Y_t) = (x_1, x_2)$. Zakładamy, że jeżeli łańcuchy się zeszły, to już się nie rozchodzą (czyli jeżeli X_t oraz Y_t są równe, to X_{t+1} i Y_{t+1} również będą równe). W przeciwnym przypadku rzucamy monetą, na której orzeł wypada z prawdopodobieństwem m :

- Jeżeli wypadnie orzeł, to $X_{t+1} = Y_{t+1} = x$, gdzie x wybieramy losowo z rozkładem:

$$\left(\frac{1}{m} m_y \mid y \in S \right)$$

- Jeżeli wypadnie reszka, to $X_{t+1} = x'_1$, $Y_{t+1} = x'_2$, gdzie x'_1 i x'_2 wybieramy kolejno z rozkładami:

$$\left(\frac{1}{1-m} (P^r(x_1, y) - m_y) \mid y \in S \right)$$

$$\left(\frac{1}{1-m} (P^r(x_2, y) - m_y) \mid y \in S \right)$$

W ten sposób skonstruowaliśmy sprzężanie, dla którego zachodzi:

$$P(X_{t+1} = Y_{t+1}) = \sum_{y \in S} P(X_{t+1} = Y_{t+1} = y) \geq \sum_{y \in S} m \cdot \frac{m_y}{m} = m$$

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 $C = \alpha^{-r}$. □

Pytanie 14

Losowe spacery w grafie jako zastosowanie łańcuchów Markowa.

14.1 Losowe spacery w grafie

Definicja 14.1.1. Spacerem losowym na nieskierowanym grafie G nazywamy łańcuch Markowa, którego stany odpowiadają wierzchołkom grafu. Prawdopodobieństwo przejścia ze stanu v do stanu u wynosi $\frac{1}{\deg(v)}$ gdy $(v, u) \in E$ i 0 w przeciwnym przypadku.

Twierdzenie 14.1.1 (Lemat 7.12 P&C). Spacer losowy na grafie G jest nieokresowy wtedy i tylko wtedy gdy G nie jest dwudzielny

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

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

Twierdzenie 14.1.2 (Lemat 7.13 P&C). Spacer losowy na spójnym, niedwudzielnym grafie G posiada rozkład stacjonarny $\bar{\pi}$ w którym $\pi_v = \frac{\deg(v)}{2|E|}$

Dowód. Pokażemy, że tak zadane $\bar{\pi}$ faktycznie jest rozkładem stacjonarnym. Mamy

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

a więc faktycznie jest to rozkład. Mamy też

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

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

Definicja 14.1.2. Czas pokrycia grafu G to chwila (indeks łańcucha Markowa), w której spacer odwiedził już każdy wierzchołek. Taką zmienną losową oznaczamy C_G .

Definicja 14.1.3. Zmienna losowa r_{uv} określa liczbę kroków spaceru losowego przebiegającego od wierzchołka u do wierzchołka v .

Lemat 14.1.1. Dla każdej krawędzi uv w grafie G zachodzi $\mathbb{E}[r_{uv}] + \mathbb{E}[r_{vu}] \leq 2|E|$.

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 $\pi_{uv} = \frac{1}{2|E_G|}$ jest stacjonarny. Po pierwsze $\sum_{uv \in E_D} \pi_{uv} = \sum_{uv \in E_D} \frac{1}{2|E_G|} = 1$, więc jest to rozkład. Mamy też

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

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

Ograniczana wartość $\mathbb{E}[r_{uv}] + \mathbb{E}[r_{vu}]$ 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

$$\mathbb{E}[r_{uv}] + \mathbb{E}[r_{vu}] \leq \mathbb{E}[r_{(vu)(vu)}] = \frac{1}{\pi_{vu}} = 2|E_G|$$

□

Twierdzenie 14.1.3 (Twierdzenie 7.15 P&C). Wartość oczekiwana czasu pokrycia grafu $G = (V, E)$ jest ograniczona od góry przez $2|E|(|V| - 1)$.

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

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

□

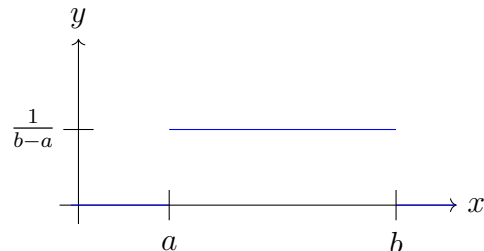
Pytanie 15

Rozkład jednostajny: gęstość, dystrybuanta, momenty, funkcja tworząca momentów, rozkład pod warunkiem, że wylosowano wartość poniżej ustalonego progu, wartość oczekiwana k -tej statystyki n niezależnych prób zmiennych o rozkładzie jednostajnym.

15.1 Rozkład jednostajny

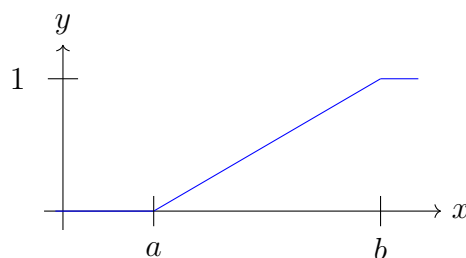
Definicja 15.1.1. Mówimy, że zmienna losowa X ma **rozkład jednostajny** na przedziale $[a, b]$ jeśli gęstość tej zmiennej zadana jest przez funkcję

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



Łatwo można zauważyć, że dystrybuanta takiej zmiennej wynosi

$$F(x) = \begin{cases} 0 & \text{gdy } x < a \\ \frac{x-a}{b-a} & \text{gdy } a \leq x \leq b \\ 1 & \text{gdy } x > b \end{cases}$$



Twierdzenie 15.1.1. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy

$$\mathbb{E}[X] = \frac{a+b}{2}$$

Dowód.

$$\begin{aligned}
 \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx \\
 &= \int_a^b x \frac{1}{b-a} dx \\
 &= \frac{1}{b-a} \int_a^b x dx \\
 &= \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b \\
 &= \frac{1}{b-a} \frac{b^2 - a^2}{2} \\
 &= \frac{1}{b-a} \frac{(b-a)(b+a)}{2} \\
 &= \frac{a+b}{2}
 \end{aligned}$$

□

Twierdzenie 15.1.2. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy

$$\text{Var}[X] = \frac{(b-a)^2}{12}$$

Dowód.

$$\begin{aligned}
 \mathbb{E}[X^2] &= \int_a^b x^2 \frac{1}{b-a} dx \\
 &= \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b \\
 &= \frac{1}{b-a} \frac{b^3 - a^3}{3} \\
 &= \frac{1}{b-a} \frac{(b-a)(a^2 + ab + b^2)}{3} \\
 &= \frac{a^2 + ab + b^2}{3}
 \end{aligned}$$

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2} \right)^2 = \frac{(b-a)^2}{12}$$

□

Twierdzenie 15.1.3. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy

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

Twierdzenie 15.1.4. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy dla dowolnych $a \leq c \leq d \leq b$

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

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 15.1.5. Niech $X_1, \dots, X_n$ będą niezależne i wszystkie mają rozkład jednostajny na $[0, 1]$. Ponadto, niech $Y_1, \dots, Y_n$ będą tymi samymi wartościami, posortowanymi rosnąco. Wtedy

$$\mathbb{E}[Y_k] = \frac{k}{n+1}$$

Dowód. Modyfikujemy lekko problem i zamiast wybierać n punktów z odcinka będziemy wybierać $n+1$ punktów z okręgu o obwodzie 1. Nazywamy je $P_0, \dots, P_n$. W ten sposób X_i jest odległością zgodnie ze wskazówkami zegara między punktami P_0, P_i , natomiast Y_k jest odległością od P_0 do k -tego punktu zgodnie ze wskazówkami zegara.

Mamy $n+1$ łuków między punktami i, ze względu na symetrię, oczekiwana długość łuku między dwoma sąsiednimi punktami wynosi $\frac{1}{n+1}$.

W takim razie oczekiwana wartość Y_k to oczekiwana łączna długość k sąsiednich łuków, która wynosi $\frac{k}{n+1}$. □

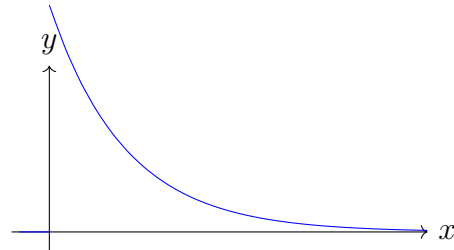
Pytanie 16

Rozkład wykładniczy. Gęstość, dystrybuanta, momenty, funkcja tworząca momentów, własność bez pamięci, rozkład minimum n niezależnych prób. Funkcja Gamma i rozkład Gamma. Związek z rozkładem wykładniczym.

16.1 Rozkład wykładniczy

Definicja 16.1.1. Rozkładem wykładniczym z parametrem λ nazywamy rozkład zadany gęstością

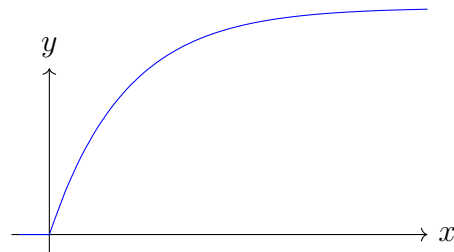
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{wpp.} \end{cases}$$



Intuicyjnie widzimy, że im λ jest mniejsze, tym bardziej ten wykres się "wypłaszcza".

Dystrybuanta takiej zmiennej wynosi

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{wpp.} \end{cases}$$



dodatkowo definiujemy

$$G(x) = P(X > x) = 1 - F(x) = \begin{cases} e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{wpp.} \end{cases}$$

Lemat 16.1.1. Dla $X \sim \text{Exp}(a)$ oraz $Y = \frac{X}{b}$ zachodzi

$$Y \sim \text{Exp}(ab)$$

Dowód.

$$P(Y \leq y) = P(X \leq by) = 1 - e^{-aby}$$

□

Twierdzenie 16.1.1. Dla $X \sim \text{Exp}(\lambda)$ zachodzi

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

$$\mathbb{E}[X^2] = \frac{2}{\lambda^2}$$

$$\text{Var}[X] = \frac{1}{\lambda^2}$$

Dowód.

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} t f(t) dt \\ &= \int_0^{\infty} t \lambda e^{-\lambda t} dt \\ &= - \int_0^{\infty} t (-\lambda e^{-\lambda t}) dt \\ &= - [t e^{-\lambda t}]_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt \\ &= 0 + \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^{\infty} \\ &= \frac{1}{\lambda} \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X^2] &= \int_0^{\infty} t^2 \lambda e^{-\lambda t} dt \\ &= - \int_0^{\infty} t^2 (-\lambda e^{-\lambda t}) dt \\ &= - [t^2 e^{-\lambda t}]_0^{\infty} + \int_0^{\infty} 2t e^{-\lambda t} dt \\ &= 0 + \frac{2}{\lambda} \int_0^{\infty} t \lambda e^{-\lambda t} dt \\ &= 0 + \frac{2}{\lambda} \cdot \frac{1}{\lambda} \\ &= \frac{2}{\lambda^2} \end{aligned}$$

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}$$

□

Alternatywnie, dla uproszczenia można to udowodnić wyłącznie dla $\lambda = 1$ a potem z 16.1.1 rozszerzyć na dowolne λ .

Twierdzenie 16.1.2 (Lemat 8.4 P&C). Rozkład wykładniczy jest **bez pamięci**, tzn. dla $X \sim \text{Exp}(\lambda)$, $s, t \in \mathbb{R}^+$ zachodzi

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

Dowód.

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

□

Jest to bardzo przydatna własność, bowiem sprawia, że możemy „resetować” zmienną o której wiemy, że ma większą wartość niż ustalona.

Twierdzenie 16.1.3 (MGF). Niech X ma rozkład wykładniczy z parametrem λ . Wtedy dla $t < \lambda$

$$M_X(t) = \frac{\lambda}{\lambda - t}$$

Dowód.

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \int_0^\infty e^{tx} \cdot f(x) \, dx \\ &= \int_0^\infty e^{tx} \cdot \lambda e^{-\lambda x} \, dx \\ &= \lambda \int_0^\infty e^{-x(\lambda - t)} \, dx \\ &= \frac{\lambda}{\lambda - t} \end{aligned}$$

□

Twierdzenie 16.1.4 (Lemat 8.5 P&C). Jeśli $X_1, \dots, X_n$ są **niezależnymi** zmiennymi losowymi spełniającymi $X_i \sim \text{Exp}(\lambda_i)$, to

$$\min(X_1, \dots, X_n) \sim \text{Exp}\left(\sum_{i=1}^n \lambda_i\right)$$

oraz

$$P(X_i = \min(X_1, \dots, X_n)) = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}$$

Dowód. Przeprowadzimy dowód dla $n = 2$, który później prostą indukcją można rozszerzyć na $n > 2$.

$$\begin{aligned} P(\min(X_1, X_2) > x) &= P(X_1 > x \wedge X_2 > x) \\ &= P(X_1 > x) \cdot P(X_2 > x) \\ &= e^{-\lambda_1 x} \cdot e^{-\lambda_2 x} \\ &= e^{-(\lambda_1 + \lambda_2)x} \end{aligned}$$

a więc $\min(X_1, X_2) \sim \text{Exp}(\lambda_1 + \lambda_2)$. Teraz pozostaje pokazać, że:

$$P(X_1 \leq X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Zatem liczymy:

$$\begin{aligned} P(X_1 \leq X_2) &= \int_{x_2=-\infty}^{\infty} \int_{x_1=-\infty}^{x_2} f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 = \\ &= \int_{x_2=-\infty}^{\infty} f_{X_2}(x_2) \int_{x_1=-\infty}^{x_2} f_{X_1}(x_1) dx_1 dx_2 = \\ &= \int_{x_2=0}^{\infty} \lambda_2 e^{-\lambda_2 x_2} \int_{x_1=0}^{x_2} \lambda_1 e^{-\lambda_1 x_1} dx_1 dx_2 = \\ &= \lambda_1 \lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \int_{x_1=0}^{x_2} e^{-\lambda_1 x_1} dx_1 dx_2 = \\ &= \lambda_1 \lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \left(\left[\frac{x_2 - 1}{\lambda_1} e^{-\lambda_1 x_1} \right]_0^{x_2} \right) dx_2 = \\ &= \lambda_1 \lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \left(\frac{-1}{\lambda_1} e^{-\lambda_1 x_2} - \frac{-1}{\lambda_1} e^0 \right) dx_2 = \\ &= \lambda_1 \lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \frac{-1}{\lambda_1} (e^{-\lambda_1 x_2} - 1) dx_2 = \\ &= -\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} (e^{-\lambda_1 x_2} - 1) dx_2 = \\ &= -\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2 - \lambda_1 x_2} - e^{-\lambda_2 x_2} dx_2 = \\ &= -\lambda_2 \int_{x_2=0}^{\infty} e^{-x_2(\lambda_2 + \lambda_1)} - e^{-\lambda_2 x_2} dx_2 = \\ &= -\lambda_2 \left(\int_{x_2=0}^{\infty} e^{-x_2(\lambda_2 + \lambda_1)} dx_2 - \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} dx_2 \right) = \end{aligned}$$

$$\begin{aligned}
& -\lambda_2 \left(\left(\int_{x_2=0}^{\infty} \frac{-1}{\lambda_1 + \lambda_2} e^{-x_2(\lambda_1 + \lambda_2)} \right) - \left(\int_{x_2=0}^{\infty} \frac{-1}{\lambda_2} e^{-\lambda_2 x_2} \right) \right) = \\
& \quad -\lambda_2 \left(\frac{1}{\lambda_1 + \lambda_2} - \frac{1}{\lambda_2} \right) = \\
& \quad -\lambda_2 \left(\frac{\lambda_2}{\lambda_2 \cdot (\lambda_1 + \lambda_2)} - \frac{\lambda_1 + \lambda_2}{\lambda_2 \cdot (\lambda_1 + \lambda_2)} \right) = \\
& \quad (-\lambda_2) \cdot \frac{-\lambda_1}{\lambda_2 \cdot (\lambda_1 + \lambda_2)} = \\
& \quad \frac{\lambda_1}{\lambda_1 + \lambda_2}
\end{aligned}$$

□

16.2 Funkcja Gamma i rozkład Gamma

Definicja 16.2.1. Funkcją gamma nazywamy funkcję:

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

dla $a > 0$. Powyższe dwie notacje są równoważne, my będziemy korzystać z tej pierwszej.

Parę faktów o funkcji gamma:

Fakt 16.2.1. $\Gamma(1) = 1$

Dowód.

$$\Gamma(1) = \int_0^{\infty} x e^{-x} \frac{dx}{x} = \int_0^{\infty} e^{-x} dx = [-e^{-x}]_0^{\infty} = 1$$

□

Fakt 16.2.2. $\forall_{a>0} \Gamma(a+1) = a\Gamma(a)$

Dowód.

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

□

Fakt 16.2.3. $\forall_{n \geq 1} \Gamma(n) = (n-1)!$

Wynika to bezpośrednio z poprzedniego faktu.

Definicja 16.2.2. Mówimy, że ciągła zmienna losowa ma **Rozkład Gamma** z parametrem $(a, 1)$, jeżeli jej funkcja gęstości jest równa:

$$f(x) = \frac{1}{\Gamma(a)} x^a e^{-x} \frac{1}{x}$$

Wtedy:

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

Zatem jest to poprawny rozkład.

Definicja 16.2.3. Dla $X \sim \text{Gamma}(a, 1)$ oraz $\lambda > 0$ definiujemy $Y \sim \text{Gamma}(a, \lambda)$ jako

$$Y = \frac{X}{\lambda}$$

Twierdzenie 16.2.1. Funkcja gęstości $Y \sim \text{Gamma}(a, \lambda)$ jest równa

$$f(x) = \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \frac{1}{x}$$

Dowód. Niech $X \sim \text{Gamma}(a, 1)$, $Y = \frac{X}{\lambda}$. Liczymy gęstość

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

□

Fakt 16.2.4.

$$\text{Gamma}(1, 1) \equiv \text{Exp}(1)$$

$$\text{Gamma}(1, \lambda) \equiv \text{Exp}(\lambda)$$

Twierdzenie 16.2.2. Niech $X_1, \dots, X_n$ - niezależne zmienne losowe o rozkładzie wykładniczym z parametrem λ . Wtedy

$$X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda)$$

Dowód. Skorzystamy z twierdzenia 28.4.1. Policzmy funkcję tworzącą dla sumy X_i oraz dla $Y \sim \text{Gamma}(n, \lambda)$.

$$M_{X_i}(t) = \mathbb{E}[e^{tX_i}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - t}$$

$$X = \sum_{i \in [n]} X_i \implies M_X(t) = \prod_{i \in [n]} M_{X_i}(t) = \left(\frac{\lambda}{\lambda - t} \right)^n$$

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{tY}] = \int_0^\infty e^{ty} \frac{1}{\Gamma(n)} (\lambda y)^n \frac{dy}{y} \\ &= \frac{\lambda^n}{(\lambda - t)^n} \int_0^\infty \frac{1}{\Gamma(n)} e^{-(\lambda - t)y} ((\lambda - t)y)^n \frac{dy}{y} \\ &= \left(\frac{\lambda}{\lambda - t} \right)^n \frac{\Gamma(n)}{\Gamma(n)} = \left(\frac{\lambda}{\lambda - t} \right)^n \end{aligned}$$

Obie funkcje tworzące są równe, więc zmienne X i Y mają ten sam rozkład.

□

Pytanie 17

(8.3.2). Problem kul i urn ze wzmocnionym feedbackiem.

17.1 Kule i urny z feedbackiem

Jak zwykle, zanim zaczniemy to pokażemy pomocniczy lemat:

Lemat 17.1.1. Niech X będzie dowolną zmienną losową ze skończoną wartością oczekiwaną, tj. $\mathbb{E}[X] \in \mathbb{R}$. Wtedy

$$P(X < \infty) = 1$$

Dowód. Korzystamy z nierówności Markowa

$$P(X \geq n) \leq \frac{\mathbb{E}[X]}{n}$$

Zatem

$$\lim_{n \rightarrow \infty} P(X \geq n) \leq \lim_{n \rightarrow \infty} \frac{\mathbb{E}[X]}{n} = 0$$

□

Kule i urny jakie są każdy widzi. Rozważmy sobie jednak zabawny model, w którym mamy tylko dwie urny ale z takim twistem, że im więcej kul jest w urnie, tym większa szansa na to, że wrzucimy tam kolejną kulę.

Konkretniej - jeśli w pierwszej urnie jest x kul a w drugiej y to prawdopodobieństwo, że kolejna kula trafi do pierwszej urny wynosi

$$\frac{x^p}{x^p + y^p}$$

a do drugiej

$$\frac{y^p}{x^p + y^p}$$

dla ustalonego p .

Będziemy się zajmować $p > 1$ tzn. więcej kul dostaje cięższa urna.

Twierdzenie 17.1.1. Dla dowolnego $p > 1$ oraz dowolnych warunków początkowych, z prawdopodobieństwem 1 od pewnego momentu kule wpadają tylko do jednej urny.

Dowód. Przyjmijmy, że w obu urnach na początku jest po jednej kuli, uprości to dowód, a rozumowanie pozostaje takie same.

Rozważmy inny, choć podobny, proces. Każda urna dostaje własny, niezależny licznik, który odlicza czas do przyjścia kolejnej kuli do tej konkretnej urny.

Jeśli w pierwszej urnie jest x kul to czas oczekiwania na kolejną wynosi T_x , które ma rozkład wykładniczy z parametrem x^p .

Podobnie dla drugiej urny – jeśli jest w niej y kul to mamy zmienną U_y z parametrem y^p .

Zauważamy teraz fajną rzecz – prawdopodobieństwo, że kolejna kula ląduje w pierwszej urnie wynosi dokładnie

$$\frac{x^p}{x^p + y^p}$$

a w drugiej

$$\frac{y^p}{x^p + y^p}$$

Czyli nasz nowy proces jest taki sam jak oryginalny, cóż za zbieg okoliczności.

Definiujemy czasy nasycenia – opisują one po jakim czasie liczba kul w urnach jest dowolnie duża.

$$F_1 = \sum_{i=1}^{\infty} T_i$$

$$F_2 = \sum_{i=1}^{\infty} U_i$$

Możemy tak zrobić, bo $\mathbb{E}[T_i] = \mathbb{E}[U_i] = \frac{1}{i^p}$, a ponieważ $p > 1$ to $\mathbb{E}[F_1]$ oraz $\mathbb{E}[F_2]$ są skończone.

Tutaj należy uważać ale książka Wam tego nie powie. Otóż a priori nie wiemy, że jeśli zmienna ma skończoną oczekiwaną to z prawdopodobieństwem 1 *zmienna* przyjmuje skończoną wartość. My się powołujemy na lemat 17.1.1 dzięki czemu wiemy, że wartości F_1, F_2 są skończone.

Co więcej, z prawdopodobieństwem 1 są różne.

Bez straty ogólności, przyjmijmy, że $F_2 > F_1$. Oznacza to, że dla pewnego n

$$\sum_{i=1}^n U_i < F_1 < \sum_{i=1}^{n+1} U_i$$

a to z kolei oznacza, że dla wystarczająco dużych m

$$\sum_{i=1}^n U_i < \sum_{i=1}^m T_i < \sum_{i=1}^{n+1} U_i$$

W takim razie, dla odpowiednio dużych m pierwsza urna zawiera m kul a druga urna zawiera jedynie n kul, czyli z prawdopodobieństwem 1 druga urna utknęła na posiadaniu n kul, a to jest to co chcieliśmy pokazać. $\square$

Pytanie 18

(8.4). Proces Poissona. Definicja. Prawdopodobieństwo pojawienia się n zdarzeń w ustalonym odcinku czasowym długości t (Twierdzenia 8.7 i 8.8).

18.1 Proces Poissona

18.1.1 Definicja

Definicja 18.1.1. Stochastycznym procesem liczącym nazywamy proces stochastyczny

$$\{N(t), | t \geq 0\}$$

spełniający

1. $N(t) \in \mathbb{N}_0$
2. $\forall_{s < t} N(s) \leq N(t)$

Intuicyjnie: $N(t)$ mówi ile *jakichś* zdarzeń zaszło od momentu rozpoczęcia procesu do chwili t , a dla $s \leq t$ liczba zdarzeń które zaszły w przedziale czasu $(s, t]$ to $N(t) - N(s)$

Definicja 18.1.2. Procesem Poissona z parametrem λ nazywamy stochastyczny proces liczący $\{N(t) | t \in \mathbb{R}, t \geq 0\}$ taki, że:

1. $N(0) = 0$
2. Proces ma stacjonarne i niezależne przyrosty, tzn.
 - 2a. Stacjonarność: $\forall_{s, t \geq 0}$ zmienne $N(s)$ oraz $N(s + t) - N(t)$ mają taki sam rozkład
 - 2b. Niezależność: $\forall_{t_1 < t_2 \leq t_3 < t_4}$ zmienne $N(t_2) - N(t_1)$ oraz $N(t_4) - N(t_3)$ są niezależne
3. Prawdopodobieństwo jednego zdarzenia w małym przedziale długości t zbiega do λ

$$\lim_{t \rightarrow 0} \frac{P(N(t) = 1)}{t} = \lambda$$

4. Prawdopodobieństwo więcej niż jednego zdarzenia w małym przedziale zbiega do zera

$$\lim_{t \rightarrow 0} \frac{P(N(t) > 1)}{t} = 0$$

18.1.2 Związek z rozkładem Poissona

Powyższa definicja nie jest jedyną możliwą definicją procesu Poissona. Okazuje się, że możemy skorzystać też z nieco wygodniejszej definicji bez warunków 3. i 4., ale za to korzystającej z rozkładu Poissona.

Pokażemy teraz dwa lematy, które dadzą nam równoważność między dwoma definicjami.

Twierdzenie 18.1.1 (Twierdzenie 8.7 P&C). Niech $\{N(t) \mid t \geq 0\}$ będzie procesem Poissona z parametrem λ . Wtedy dla dowolnego $t \geq 0$ oraz $n \in \mathbb{N}$

$$P_n(t) := P(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

Dowód. Zaczynamy od policzenia $P_0(t)$; dowód będzie indukcyjny.

Zauważmy, że z niezależności przyrostów mamy

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

Robimy więc pierwszą rzecz, która nam przychodzi do głowy tj. liczymy pochodną $P_0(t)$, a co.

$$\begin{aligned} P_0'(t) &= \lim_{h \rightarrow 0} \frac{P_0(t+h) - P_0(t)}{h} \\ &= \lim_{h \rightarrow 0} P_0(t) \cdot \frac{P_0(h) - 1}{h} \\ &= \lim_{h \rightarrow 0} P_0(t) \cdot \frac{(1 - P(N(h) = 1) - P(N(h) > 1)) - 1}{h} \\ &= \lim_{h \rightarrow 0} \left(P_0(t) \cdot \left(\frac{-P(N(h) = 1)}{h} - \frac{P(N(h) > 1)}{h} \right) \right) \\ &= P_0(t) \cdot \left(-\lim_{h \rightarrow 0} \frac{P(N(h) = 1)}{h} - \lim_{h \rightarrow 0} \frac{P(N(h) > 1)}{h} \right) \\ &= P_0(t) \cdot (-\lambda - 0) \\ &= -\lambda P_0(t) \end{aligned}$$

Wyniki poszczególnych limesów biorą się z własności 4 i 5 procesu Poissona.

Mamy zatem równanie różniczkowe

$$P'_0(t) = -\lambda P_0(t)$$

$$\frac{P'_0(t)}{P_0(t)} = -\lambda$$

Całkujemy po t i dostajemy

$$\ln P_0(t) = -\lambda t + C$$

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

Ponieważ $P_0(0) = 1$ to $C = 0$, czyli $P_0(t) = e^{-\lambda t}$. Tym samym bazę indukcji mamy udowodnioną.

Podobnie zabawny motyw dzieje się gdy obliczamy kolejne $P_n(t)$. Na początek zaobserwujmy jednak jedną rzecz.

Fakt 18.1.1.

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

Dowód. Jeśli wiemy, że w czasie $t+h$ zaistniało n zdarzeń, to wiemy, że jakieś k (być może 0) zdarzeń musiało zaistnieć w czasie h , a więc $n-k$ zdarzeń zaistniało w czasie t . Aby policzyć prawdopodobieństwo takiej sytuacji wystarczy wymnożyć 2 takie prawdopodobieństwa (bo niezależność) a z racji tego że kolejne składniki sumy opisują zdarzenia które są rozłączne to zsumowanie jest legalne. $\square$

Korzystając z wyżej wymienionego faktu, mamy:

$$\begin{aligned} P_n(t+h) &= \sum_{k=0}^n P_{n-k}(t) \cdot P_k(h) \\ &= P_n(t) \cdot P_0(h) + P_{n-1}(t) \cdot P_1(h) + \sum_{k=2}^n P_{n-k}(t) \cdot P(N(h) = k) \end{aligned}$$

Zrobiliśmy tu bardzo sprytną rzecz – mianowicie rozbiliśmy sumę na trzy części tak, aby przy

liczeniu pochodnych wszystko nam się ładnie zwinęło.

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

Znowu dostajemy równanie różniczkowe

$$\begin{aligned}
P'_n(t) &= -\lambda P_n(t) + \lambda P_{n-1}(t) \\
P'_n(t) + \lambda P_n(t) &= \lambda P_{n-1}(t) \\
e^{\lambda t} (P'_n(t) + \lambda P_n(t)) &= \lambda e^{\lambda t} P_{n-1}(t) \\
e^{\lambda t} P'_n(t) + e^{\lambda t} \lambda P_n(t) &= \lambda e^{\lambda t} P_{n-1}(t) \\
\frac{d}{dt} (e^{\lambda t} \cdot P_n(t)) &= \lambda e^{\lambda t} P_{n-1}(t)
\end{aligned}$$

I z założenia indukcyjnego:

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

Całkujemy obustronnie:

$$\int \frac{d}{dt} (e^{\lambda t} P_n(t)) dt = e^{\lambda t} P_n(t) + C_1$$

$$\int \frac{\lambda^n \cdot t^{n-1}}{(n-1)!} dt = \frac{\lambda^n}{(n-1)!} \cdot \int t^{n-1} dt = \frac{\lambda^n}{(n-1)!} \cdot \frac{t^n}{n} + C_2 = \frac{\lambda^n t^n}{n!} + C_2$$

Definiujemy $C = C_2 - C_1$ by mieć mniej myśleć o stałych:

$$\begin{aligned}
e^{\lambda t} P_n(t) + C_1 &= \frac{\lambda^n t^n}{n!} + C_2 \\
e^{\lambda t} P_n(t) &= \frac{\lambda^n t^n}{n!} + C_2 - C_1 \\
e^{\lambda t} P_n(t) &= \frac{\lambda^n t^n}{n!} + C \\
P_n(t) &= e^{-\lambda t} \frac{\lambda^n t^n}{n!} + C e^{-\lambda t}
\end{aligned}$$

Wiemy, że $P_n(0) = 0$, zatem $C = 0$. W takim razie:

$$P_n(t) = e^{-\lambda t} \frac{\lambda^n t^n}{n!} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

□

Twierdzenie 18.1.2 (Twierdzenie 8.8 P&C). Niech $\{N(t) \mid t \geq 0\}$ będzie procesem stochastycznym liczącym takim, że

1. $N(0) = 0$
2. Proces ma niezależne przyrosty
3. $\forall_{t,s} P(N(s+t) - N(t) = n) = e^{-\lambda s} \frac{(\lambda s)^n}{n!}$

Wtedy jest to proces Poissona z parametrem λ .

Dowód. Pokazujemy co następuje z definicji 18.1.2

1. $N(0) = 0$ z założeń
2. Niezależność i stacjonarność przyrostów również bezpośrednio z założeń
- 3.

$$\lim_{t \rightarrow 0} \frac{P(N(t) = 1)}{t} = \lim_{t \rightarrow 0} \frac{e^{-\lambda t} \frac{\lambda t}{1!}}{t} = \lim_{t \rightarrow 0} \lambda e^{-\lambda t} = \lambda$$

- 4.

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

□

Pytanie 19

(8.4.1). Proces Poissona. Rozkład czasów pomiędzy zdarzeniami (Twierdzenia 8.9, 8.10 i 8.11).

19.1 Międzyczasy

Niech X_1 to czas pierwszego zdarzenia, a X_i dla $i \geq 2$ to czas pomiędzy $(i-1)$ -szym a i -tym zdarzeniem - dalej nazywane międzyczasami.

Twierdzenie 19.1.1 (Twierdzenia 8.9 i 8.10 P&C). Niech $\{N(t) \mid t \geq 0\}$ będzie procesem Poissona z parametrem λ . Wtedy międzyczasy są niezależne i mają rozkład wykładniczy z parametrem λ .

Dowód. Pierwsze pokażmy, że jest to prawda dla X_1

$$P(X_1 > t) = P(N(t) = 0) = e^{-\lambda t}$$

Jest to dopełnienie dystrybuanty zmiennej o rozkładzie wykładniczym z parametrem λ .

A teraz pokażmy, że jest to prawda dla $X_i, i > 1$

$$\begin{aligned} P(X_i > t_i \mid (X_0, X_1, \dots, X_{i-1}) = (t_0, t_1, \dots, t_{i-1})) &= P\left(N\left(\sum_1^i t_j\right) - N\left(\sum_1^{i-1} t_j\right) = 0\right) \\ &= P(N(t_i) = 0) = e^{-\lambda t_i} \end{aligned}$$

A więc X_i też ma rozkład wykładniczy z parametrem λ , i jest niezależne od wartości poprzednich międzyczasów. $\square$

Twierdzenie 19.1.2 (Twierdzenie 8.11 P&C). Niech $\{N(t) \mid t \geq 0\}$ będzie stochastycznym procesem zliczającym takim, że

1. $N(0) = 0$
2. Międzyczasy są niezależne i wszystkie mają rozkład $\text{Exp}(\lambda)$

Wtedy proces ten jest procesem Poissona z parametrem λ

Dowód. Pokażemy kolejne własności z definicji 18.1.2

1. $N(0) = 0$ z definicji
- 2a. Stacjonarność: aby pokazać, że rozkład $N(s+t) - N(s)$ jest taki sam jak rozkład $N(t)$ zrobimy tę samą sztuczkę co przed chwilą - w chwili s resetujemy ostatnią zmienną. Wszystko teraz dzieje się na przedziale długości t bez żadnych zależności od tego co było wcześniej, zatem rozkład liczby zdarzeń musi być taki sam jak rozkład $N(t)$
- 2b. Niezależność: weźmy dowolne dwa przedziały $[b, a] \cap [d, c] = \emptyset$, przy czym $d > a$

W chwili d , „toczy się” pewna zmienna X licząca czas między dwoma zdarzeniami. Możemy ją „zresetować” albo bardziej formalnie warunkować się po tym, że $X > t$ gdzie t jest czasem od poprzedniego zdarzenia do chwili d .

Rozkład X pod warunkiem, że $X > t$ jest wykładniczy z parametrem λ i jest niezależny od tego co się działo przed d , zatem wszystko co zarejestrujemy na przedziale $[d, c]$ jest niezależne od zdarzeń na przedziale $[b, a]$.

3. Niech X_1 będzie zmienną opisującą czas do pierwszego zdarzenia, a X_2 od pierwszego zdarzenia do drugiego. Widzimy, że $P(N(t) = 1) = P(X_1 < t \wedge X_1 + X_2 > t)$

Ponieważ X_1, X_2 mają rozkład wykładniczy a $X_1 + X_2$ nie jest specjalnie ładnym tworem, to będziemy chcieli poradzić sobie nieco inaczej.

Skorzystamy zatem z twierdzenia o trzech funkcjach aby udowodnić zadaną granicę.

Nasze oszacowania będą wyglądały następująco:

$$P(X_1 < t \wedge X_2 > t) \leq P(N(t) = 1) \leq P(X_1 < t)$$

Ograniczenie od dołu jest na pewno mniej prawdopodobnym zdarzeniem – jeśli $X_1 < t \wedge X_2 > t$ to na pewno $N(t) = 1$, ale nie uwzględnia ono sytuacji kiedy $X_1, X_2 < t \wedge X_1 + X_2 > t$.

Podobnie oszacowanie górne – warunek jest konieczny, ale nie wystarczający, zatem zajdzie z większym prawdopodobieństwem.

Możemy zatem policzyć granice ograniczeń.

Ograniczenie dolne:

$$\begin{aligned}
 \lim_{t \rightarrow 0} \frac{P(X_1 < t \wedge X_2 > t)}{t} &= \lim_{t \rightarrow 0} \frac{P(X_1 < t) \cdot P(X_2 > t)}{t} \\
 &= \lim_{t \rightarrow 0} \frac{(1 - \exp(-\lambda t)) \exp(-\lambda t)}{t} \\
 &= \lim_{t \rightarrow 0} \frac{\exp(-\lambda t) - \exp(-2\lambda t)}{t} = \left[\begin{array}{c} 0 \\ 0 \end{array} \right] \\
 &= \lim_{t \rightarrow 0} -\lambda \exp(-\lambda x) + 2\lambda \exp(-2\lambda x) \\
 &= -\lambda + 2\lambda = \lambda
 \end{aligned}$$

Ograniczenie górne:

$$\begin{aligned}
 \lim_{t \rightarrow 0} \frac{P(X_1 < t)}{t} &= \lim_{t \rightarrow 0} \frac{1 - \exp(-\lambda t)}{t} = \left[\begin{array}{c} 0 \\ 0 \end{array} \right] \\
 &= \lim_{t \rightarrow 0} \lambda \exp(-\lambda t) = \lambda
 \end{aligned}$$

Obie granice wyszły nam λ , zatem $\lim_{t \rightarrow 0} \frac{P(N(t)=1)}{t} = \lambda$. Fajnie.

4. Ostatni warunek szacujemy niemal identycznie jak poprzedni.

Podobnie zauważamy, że $0 \leq P(N(t) > 1) \leq P(X_1 < t \wedge X_2 < t)$ – to, że oba czasy są mniejsze niż t nie oznacza jeszcze, że ich suma również taka jest, zatem jest to warunek konieczny, ale nie wystarczający.

$$\begin{aligned}
 \lim_{t \rightarrow 0} \frac{P(X_1 < t \wedge X_2 < t)}{t} &= \lim_{t \rightarrow 0} \frac{P(X_1 < t) P(X_2 < t)}{t} \\
 &= \lim_{t \rightarrow 0} \frac{(1 - \exp(-\lambda t))(1 - \exp(-\lambda t))}{t} \\
 &= \lim_{t \rightarrow 0} \frac{1 + \exp(-2\lambda t) - 2\exp(-\lambda t)}{t} = \left[\begin{array}{c} 0 \\ 0 \end{array} \right] \\
 &= \lim_{t \rightarrow 0} -2\lambda \exp(-2\lambda t) + 2\lambda \exp(-\lambda t) \\
 &= -2\lambda + 2\lambda = 0
 \end{aligned}$$

Wyszła nam granica jaką chcieliśmy, a więc z twierdzenia o trzech funkcjach

$$\lim_{t \rightarrow 0} \frac{P(N(t) > 1)}{t} = 0$$

co kończy dowód.

□

Pytanie 20

(8.4.2). Scalanie i rozdzielanie procesów Poissona (Twierdzenia 8.12 i 8.13).

20.1 Scalanie i rozdzielanie procesów Poissona

Definicja 20.1.1. Mówimy, że procesy Poissona $\{N_1(t), | t \geq 0\}, \{N_2(t), | t \geq 0\}$ są **niezależne** jeśli $\forall_{x,y} N_1(x)$ i $N_2(y)$ są niezależne.

20.1.1 Scalanie

Ta prostsza część.

Twierdzenie 20.1.1 (Twierdzenie 8.12 P&C). Niech N_1, N_2 będą niezależnymi procesami Poissona z parametrami λ_1, λ_2 . Wtedy $N(t) = N_1(t) + N_2(t)$ jest procesem Poissona z parametrem $\lambda_1 + \lambda_2$ a ponadto każde zdarzenie procesu N przyszło z procesu N_1 z prawdopodobieństwem $\frac{\lambda_1}{\lambda_1 + \lambda_2}$

Dowód. Pierwszy warunek mamy za darmo. Żeby pokazać niezależność przyrostów, musimy zauważyć tylko że dla $t_1 < t_2 < t_3 < t_4$ mamy:

$$N(t_2) - N(t_1) = N_1(t_2) - N_1(t_1) + N_2(t_2) - N_2(t_1)$$

$$N(t_4) - N(t_3) = N_1(t_4) - N_1(t_3) + N_2(t_4) - N_2(t_3)$$

a N_1 i N_2 są niezależne z założenia.

Zauważamy, że skoro $N_1(t), N_2(t)$ miały rozkład Poissona, to $N_1(t) + N_2(t)$ również ma rozkład Poissona, tyle, że z parametrem $\lambda_1 + \lambda_2$, zatem otrzymujemy proces Poissona z parametrem $\lambda_1 + \lambda_2$.

Druga część tezy wynika wprost z tego, że czasy między zdarzeniami mają rozkłady wykładnicze. □

20.1.2 Rozdzielanie

Ta smutniejsza część.

Twierdzenie 20.1.2 (Twierdzenie 8.13 P&C). Niech N będzie procesem Poissona z parametrem λ . Każde zdarzenie jest niezależnie typu 1 z prawdopodobieństwem p oraz typu 2 z prawdopodobieństwem $1 - p$.

Wtedy zdarzenia typu 1 tworzą proces Poissona N_1 z parametrem λp a typu 2 proces Poissona N_2 z parametrem $\lambda(1 - p)$. Ponadto, te dwa procesy są niezależne.

Dowód. Niezależność i stacjonarność dziedziczymy z N , tak samo $N_1(0) = 0$. Policzmy zatem

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

Dostaliśmy rozkład Poissona z parametrem $\lambda p t$, czyli N_1 jest procesem Poissona z parametrem $\lambda p t$. Tak samo pokazujemy N_2 .

Pozostaje pokazać niezależność tych procesów. Najpierw pokazujemy, że $N_1(t)$ oraz $N_2(t)$ są niezależne.

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

Wypadałoby jeszcze pokazać, że dla dowolnych t, u $N_1(t)$ oraz $N_2(u)$ są niezależne. Ponieważ rozumowanie jest analogiczne, to założmy, że $t < u$.

Zauważamy bardzo odkrywczą rzecz, mianowicie $N_2(u) = N_2(t) + (N_2(u) - N_2(t))$. Pokazaliśmy już, że $N_1(t)$ oraz $N_2(t)$ są niezależne, więc wystarczy pokazać, że $N_1(t)$ oraz $N_2(u) - N_2(t)$ też są niezależne. A tak jest, dlatego, że oryginalny N był procesem Poissona i rozdzielanie

robiliśmy niezależnie, więc to ile zdarzeń z przedziału (t, u) wpadło do N_2 jest niezależne od $N_1(t)$. $\square$

Pytanie 21

(8.4.3). Warunkowe czasy pojawiania się zdarzeń w procesie Poissona (Twierdzenie 8.14).

21.1 Warunkowe czasy pojawiania się zdarzeń w procesie Poissona

Lemat 21.1.1. 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.

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

□

Twierdzenie 21.1.1 (Twierdzenie 8.14 P&C). 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 przyścia. 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 (wychodzi $\lambda^n e^{-\lambda t}$ co jest stałą względem pozostałych całek, więc wyciągamy to na przód), 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 wynosiło:

$$\begin{aligned}
 P((T_1, \dots, T_n) \leq (s_1, \dots, s_n) \mid N(t) = n) &= \frac{\lambda^n e^{-\lambda t} \int_{u_1=0}^{s_1} \dots \int_{u_n=u_{n-1}}^{s_n} du_n \dots du_1}{\frac{e^{-\lambda t} (\lambda t)^n}{n!}} \\
 &= \frac{n!}{t^n} \int_{u_1=0}^{s_1} \dots \int_{u_n=u_{n-1}}^{s_n} du_n \dots du_1
 \end{aligned}$$

□

Pytanie 22

Naturalny proces losowy motywujący gęstość rozkładu normalnego. Rozkład normalny. Własności. Funkcja tworząca.

22.1 Standardowy rozkład normalny

22.1.1 Wyprowadzenie

Rozważmy sobie rozkład na $\mathbb{R}^2$ o następujących własnościach:

- gęstość wokół każdego punktu zależy jedynie od odległości od środka układu (w szczególności rotacja nie zmienia rozkładu).
- wartości współrzędnych x i y są od siebie niezależne.
- ten rozkład jest ciągły.

Okazuje się, że istnieje tylko jeden taki rozkład (z dokładnością do stałej), nazywamy go standardowym rozkładem normalnym. Oznaczamy go przez $Z \sim N(0, 1)$.

Zgodnie z definicją, gęstość zależy jedynie od odległości punktu od środka układu współrzędnych. Możemy więc to zapisać jako:

$$f((x, y)) = f(r) = f\left(\sqrt{x^2 + y^2}\right) = g(x)h(y) = g(x)g(y)$$

Ostatnie przekształcenie wynika z tego, że nasza gęstość nie zależy od rotacji, więc możemy obrócić wszystko o 90 stopni. Dodatkowo, dla punktu $(r, 0)$ równanie przyjmie postać:

$$f((r, 0)) = g(r)g(0)$$

gdzie $g(0)$ jest stałą. Z tego powodu możemy wstępnie założyć, że $f = g$ a potem całość odpowiednio przeskalować. Tak więc teraz mamy:

$$f\left(\sqrt{x^2 + y^2}\right) = f(x)f(y)$$

Niech:

$$h(x) = f(\sqrt{x})$$

Wtedy:

$$h(x^2) = f(x)$$

$$h(x^2 + y^2) = h(x^2)h(y^2)$$

$$h(x + y) = h(x)h(y)$$

Z ostatniego punktu można przez indukcję pokazać, że $\forall_{n \in \mathbb{N}} \forall_{x_1, \dots, x_n \in \mathbb{R}} h(x_1 + \dots + x_n) = h(x_1) \dots h(x_n)$. Niech $h(1) = b$. Korzystając z poprzedniego faktu mamy, że $\forall_{n \in \mathbb{N}} h(n) = b^n$.

Teraz chcemy udowodnić to samo dla liczb wymiernych:

$$h\left(\frac{p}{q} + \dots + \frac{p}{q}\right) = h(p) = b^p = h\left(\frac{p}{q}\right)^q$$

Gdzie na początku mamy dokładnie q ułamków $\frac{p}{q}$ w funkcji h . Przekształcając ostatnią równość otrzymujemy:

$$h\left(\frac{p}{q}\right) = b^{\frac{p}{q}}$$

Na koniec chcemy udowodnić to samo dla liczb rzeczywistych (co na wykładzie chyba pominęliśmy). Z MFI pamiętamy, że każdą liczbę rzeczywistą możemy przybliżyć jakimś ciągiem liczb wymiernych, a bardziej formalnie:

$$\forall_{x \in \mathbb{R}} \exists_{q_n} x = \lim_{n \rightarrow \infty} q_n$$

Gdzie q_n jest jakimś ciągiem liczb wymiernych. Z połączenia tego faktu i założenia o ciągłości funkcji h otrzymamy:

$$h(x) = h\left(\lim_{n \rightarrow \infty} q_n\right) = \lim_{n \rightarrow \infty} h(q_n) = \lim_{n \rightarrow \infty} b^{q_n} = b^{\lim_{n \rightarrow \infty} q_n} = b^x$$

Ustalmy:

$$b = e^c \implies h(x) = e^{cx}$$

Teraz podstawiamy to do naszej funkcji gęstości, uwzględniamy skalowanie i otrzymujemy:

$$f(z) = a \cdot e^{cx^2}$$

Gdzie $c < 0$.

Przyjmijmy $c = -\frac{1}{2}$. Teraz chcemy znaleźć stałą a . Oczywiście chcemy, żeby pole pod naszą

funkcją wynosiło 1, więc wystarczy obliczyć odpowiednią całkę. Całki:

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz$$

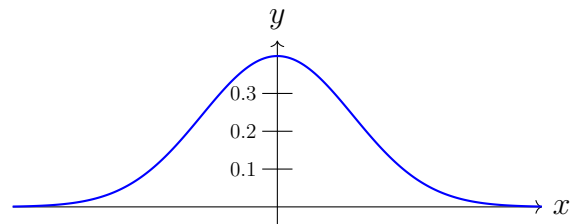
nie jesteśmy w stanie ładnie rozwiązać, więc posłużymy się takim trikiem:

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \cdot \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} \cdot e dr d\theta \\ &= \int_0^{2\pi} \int_0^{\infty} e^{-u} du d\theta \\ &= \int_0^{2\pi} 1 d\theta = 2\pi \end{aligned}$$

Gdzie kolejno druga i trzecia równość to przejście na współrzędne biegunowe, oraz podstawienie $u = \frac{r^2}{2}$ (dlatego przyjęliśmy akurat $c = -\frac{1}{2}$). Dalej mamy:

$$\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{2\pi} \implies a = \frac{1}{\sqrt{2\pi}}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$



Funkcja gęstości prawdopodobieństwa standardowego rozkładu normalnego wygląda jak dzwon.

22.1.2 Właściwości

Twierdzenie 22.1.1. Wartość oczekiwana standardowego rozkładu normalnego wynosi 0, wariancja wynosi 1.

Dowód. Wartość oczekiwana wynosi 0, ponieważ standardowy rozkład normalny jest symetryczny wobec prostej OY Wariancja:

$$\text{Var}[Z] = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = \mathbb{E}[Z^2] =$$

ponieważ $\mathbb{E}[Z] = 0$

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z t^2 e^{-t^2/2} dt = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z (t)(te^{-t^2/2}) dt = \end{aligned}$$

całkowanie przez części

$$-\frac{1}{\sqrt{2\pi}}te^{-t^2/2}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-t^2/2}dt = 1$$

Ponieważ pierwszy wyraz jest równy 0 a drugi jest to dystrybuanta na od $-\infty$ do ∞ więc wynosi ona 1. $\square$

Definicja 22.1.1. Dystrybuantę standardowego rozkładu normalnego oznaczamy jako Φ , gdzie:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$$

Oraz:

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

Ta całka generalnie nie jest do policzenia, jeżeli trzeba skorzystać z dystrybuanty to są do tego specjalne tabele wartości

22.2 Uogólniony rozkład normalny

Definicja 22.2.1. Dla $Z \sim N(0, 1)$ definiujemy (uogólniony) rozkład normalny $X \sim N(\mu, \sigma^2)$ jako

$$X = \mu + \sigma Z$$

Twierdzenie 22.2.1. Dla $X \sim N(\mu, \sigma^2)$ zachodzi

- $\mathbb{E}[X] = \mu$
- $\text{Var}[X] = \sigma^2$
- $F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$
- $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Dowód. Ponieważ zmienna losowa X z $N(\mu, \sigma^2)$ ma ten sam rozkład co $\mu + \sigma Z$ mamy że

$$\mathbb{E}[X] = \mathbb{E}[\mu + \sigma Z] = \mu + \sigma\mathbb{E}[Z] = \mu$$

$$\text{Var}[X] = \text{Var}[\sigma Z + \mu] = \sigma^2 \text{Var}[Z] = \sigma^2$$

$$F_X(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

$$f_X(x) = (F_X(x))' = \left(\Phi\left(\frac{x - \mu}{\sigma}\right)\right)' = \frac{1}{\sigma} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt\right)' = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$\square$

Twierdzenie 22.2.2. Funkcja tworząca momenty rozkładu normalnego $N(\mu, \sigma^2)$ wynosi

$$\mathcal{M}_X(t) = e^{\frac{t^2\sigma^2}{2} + \mu t}$$

Dowód.

$$\begin{aligned} \mathcal{M}_X(t) &= \mathbb{E}[e^{tX}] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2 - 2(\mu + \sigma^2 t)x + (\mu + \sigma^2 t)^2 - (\mu + \sigma^2 t)^2 + \mu^2}{2\sigma^2}\right) dx \\ &= \exp\left(\frac{(\mu + \sigma^2 t)^2 - \mu^2}{2\sigma^2}\right) \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}} dx}_1 \\ &= \exp\left(\frac{\mu^2 + 2\mu\sigma^2 t + \sigma^4 t^2 - \mu^2}{2\sigma^2}\right) \\ &= e^{\mu t + \frac{t^2\sigma^2}{2}} \end{aligned}$$

Całka w 3 linii od dołu jest równa 1, bo jest to całka po gęstości rozkładu $N(\mu + \sigma^2 t, \sigma^2)$. $\square$

Twierdzenie 22.2.3. Niech $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ to niezależne zmienne losowe. Wtedy $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Dowód.

$$\begin{aligned} \mathcal{M}_{X+Y}(t) &= (\mathcal{M}_X(t))(\mathcal{M}_Y(t)) = \left(e^{\frac{t^2\sigma_1^2}{2} + \mu_1 t}\right) \left(e^{\frac{t^2\sigma_2^2}{2} + \mu_2 t}\right) = \\ &= e^{\frac{t^2(\sigma_1^2 + \sigma_2^2)}{2} + t(\mu_1 + \mu_2)} \end{aligned}$$

$\square$

Podobnie możemy pokazać, że dla niezależnych $X_1 \sim N(0, \sigma_1^2), X_2 \sim N(0, \sigma_2^2)$ dostajemy $X_1 + X_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$ oraz $X_1 - X_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$.

Pytanie 23

(9.3). Centralne Twierdzenie Graniczne. Dowód. Warianty mocniejszych wypowiedzi.

23.1 Centralne Twierdzenie Graniczne

23.1.1 Podstawowa wersja

Intuicyjnie: Centralne Twierdzenie Graniczne mówi, że jak mamy niezależne zmienne losowe o takim samym rozkładzie, to dla liczby prób zbiegającej do nieskończoności rozkład średniej arytmetycznej tych wylosowanych wartości będzie zbiegać do rozkładu normalnego. Twierdzenie to uzasadnia występowanie w naturze rozkładu normalnego.

Definicja 23.1.1. Ciąg dystrybuant $F_1, F_2, \dots$ zbiega w dystrybuancie do dystrybuanty F , co oznaczamy jako $F_n \rightarrow F$, jeśli dla każdego $a \in \mathbb{R}$ w którym F jest ciągła zachodzi:

$$\lim_{n \rightarrow \infty} F_n(a) = F(a)$$

Twierdzenie 23.1.1 (Centralne Twierdzenie Graniczne). Niech $\{X_i\}_{i \in \mathbb{N}}$ będą niezależnymi zmiennymi losowymi o takim samym rozkładzie, wartości oczekiwanej μ i wariancji σ^2 . Niech $\widetilde{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Wówczas dla dowolnych a, b

$$\lim_{n \rightarrow \infty} P \left(a \leq \frac{\widetilde{X}_n - \mathbb{E}[\widetilde{X}_n]}{\sqrt{\text{Var}[\widetilde{X}_n]}} \leq b \right) = \Phi(b) - \Phi(a)$$

Dowód. Pierwsze, przekształćmy sobie trochę nasz cel

$$\mathbb{E}[\widetilde{X}_n] = \mu$$

$$\text{Var}[\widetilde{X}_n] = \text{Var} \left[\frac{1}{n} \sum_{i=1}^n X_i \right] = \frac{1}{n^2} \text{Var} \left[\sum_{i=1}^n X_i \right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}$$

A więc

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a \leq \frac{\widetilde{X}_n - \mathbb{E}[\widetilde{X}_n]}{\sqrt{\text{Var}[\widetilde{X}_n]}} \leq b \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left(a \leq \frac{\widetilde{X}_n - \mu}{\sigma} \cdot \sqrt{n} \leq b \right)$$

Dalej, aby dowieść CTG, będziemy musieli przytoczyć *pomocne twierdzonek*, którego (mamy nadzieję) nikt nie będzie musiał dowodzić:

Twierdzenie 23.1.2 (Lévy-Cramér). Niech $\{Y_i\}_{i \in \mathbb{N}}$ będzie sekwencją zmiennych losowych, gdzie Y_i ma dystrybuantę F_i i funkcję tworzącą momenty M_i . Niech Y będzie zmienną losową o dystrybuancie F i funkcji tworzącej momenty M . Jeżeli dla każdego t zachodzi:

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

to dla każdego t takiego, że F jest ciągła w t zachodzi

$$\lim_{n \rightarrow \infty} F_n(t) = F(t)$$

Dowód. Mitzenmacher przytacza to twierdzenie bez dowodu; na wykładzie go również nie było, a więc i my udowodnimy je poprzez założenie go jako aksjomat (haha). $\square$

Przystępujemy teraz do dowodzenia CTG.

Definiujemy $Z_i = \frac{X_i - \mu}{\sigma}$. Wówczas Z_i są niezależnymi zmiennymi losowymi oraz

$$\mathbb{E}[Z_i] = \mathbb{E} \left[\frac{X_i - \mu}{\sigma} \right] = \frac{1}{\sigma} \cdot (\mathbb{E}[X_i] - \mathbb{E}[\mu]) = \frac{1}{\sigma} \cdot (\mu - \mu) = 0$$

$$\text{Var}[Z_i] = \text{Var} \left[\frac{X_i - \mu}{\sigma} \right] = \frac{1}{\sigma^2} \cdot (\text{Var}[X_i - \mu]) = \frac{1}{\sigma^2} \cdot (\text{Var}[X_i] - \text{Var}[\mu]) = \frac{1}{\sigma^2} \cdot (\sigma^2 - 0) = 1$$

$$\text{Var}[Z_i] = \mathbb{E}[Z_i^2] - \mathbb{E}[Z_i]^2 \implies \mathbb{E}[Z_i^2] = \text{Var}[Z_i] + \mathbb{E}[Z_i]^2 = 1 + 0^2 = 1$$

Ponadto mamy, że:

$$\frac{\widetilde{X}_n - \mu}{\sigma} \cdot \sqrt{n} = \frac{\sum_{i=1}^n \frac{X_i}{n} - \mu}{\sigma} \cdot \sqrt{n} = \frac{\sum_{i=1}^n \frac{X_i - \mu}{n}}{\sigma} \cdot \sqrt{n} = \frac{\sqrt{n}}{n} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} = \frac{\sum_{i=1}^n Z_i}{\sqrt{n}}$$

Żeby zastosować teraz przywołane przez nas twierdzenie Levy'ego i tego drugiego musimy

pokazać, że funkcja tworząca momenty zmiennych losowych postaci

$$Y_n = \frac{\sum_{i=1}^n Z_i}{\sqrt{n}}$$

zbiega do funkcji tworzącej momenty zmiennej losowej o standardowym rozkładzie normalnym. Po zastosowaniu tego twierdzenia dostalibyśmy już tezę Centralnego Twierdzenia Granicznego.

W takim razie, chcemy pokazać

$$\lim_{n \rightarrow \infty} M_{Y_n}(t) = \lim_{n \rightarrow \infty} \mathbb{E} \left[e^{t \frac{\sum_{i=1}^n Z_i}{\sqrt{n}}} \right] = e^{\frac{t^2}{2}}$$

Niech $M_{Z_i}(t) = \mathbb{E}[e^{tZ_i}]$ będzie funkcją tworzącą momenty zmiennej Z_i . Zauważamy, że wówczas funkcja tworząca momenty zmiennej losowej $\frac{Z_i}{\sqrt{n}}$ wynosi

$$M_{\frac{Z_i}{\sqrt{n}}}(t) = \mathbb{E} \left[e^{t \cdot \frac{Z_i}{\sqrt{n}}} \right] = M_{Z_i} \left(\frac{t}{\sqrt{n}} \right)$$

Ponieważ Z_i są niezależne i mają ten sam rozkład mamy

$$M_{Y_n}(t) = M_{\sum_{i=1}^n \frac{Z_i}{\sqrt{n}}}(t) = \left(M_{\frac{Z_i}{\sqrt{n}}}(t) \right)^n = \left(M_{Z_i} \left(\frac{t}{\sqrt{n}} \right) \right)^n$$

Teraz wykonujemy *magiczne założenie*. Zdefiniujmy sobie, **for no reason at all**, funkcję L , taką że

$$L(t) = \ln M_{Z_i}(t)$$

Dodatkowo, *również bez jakiegokolwiek przyczyny*, policzmy sobie pierwszą i drugą pochodną $L(0)$.

Zacznijmy od trywialnych obserwacji:

$$M_{Z_i}(0) = 1 \implies L(0) = 0$$

$$L'(0) = (\ln M_{Z_i}(0))' = \frac{1}{M_{Z_i}(0)} \cdot M'_{Z_i}(0) = \frac{M'_{Z_i}(0)}{M_{Z_i}(0)} = \frac{\mathbb{E}[Z_i]}{1} = \mathbb{E}[Z_i] = 0$$

$$L''(0) = \frac{M_{Z_i}(0)M''_{Z_i}(0) - (M'_{Z_i}(0))^2}{(M_{Z_i}(0))^2} = \frac{M''_{Z_i}(0) - 0}{1} = \mathbb{E}[Z_i^2] = 1$$

Przypomnijmy, że chcieliśmy pokazać, że

$$\lim_{n \rightarrow \infty} M_{Y_n}(t) = \lim_{n \rightarrow \infty} \left(M_{Z_i} \left(\frac{t}{\sqrt{n}} \right) \right)^n = e^{\frac{t^2}{2}}$$

po zlogarytmowaniu stronami

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

Pytanie teraz co musimy zrobić by wykazać, że ta granica tyle wynosi.

Jak wszyscy wiemy, kiedy nie wiadomo jak policzyć granicę, to liczymy ją L'Hôpitalem. Zapiszmy więc sobie ten limit tak, byśmy mogli użyć tego twierdzenia (czyli żeby pojawił się symbol nieoznaczony $\frac{0}{0}$).

$$\lim_{n \rightarrow \infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}}$$

No i lecimy z pochodnymi!

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{L\left(\frac{t}{\sqrt{n}}\right)}{n^{-1}} &= \left[\frac{0}{0} \right] \\ &= \lim_{n \rightarrow \infty} \frac{-\frac{1}{2}L'\left(\frac{t}{\sqrt{n}}\right)tn^{-\frac{3}{2}}}{-n^{-2}} \\ &= \lim_{n \rightarrow \infty} \frac{L'\left(\frac{t}{\sqrt{n}}\right)t}{2n^{-\frac{1}{2}}} \\ &= \left[\frac{0}{0} \right] \\ &= \lim_{n \rightarrow \infty} \frac{-\frac{1}{2}L''\left(\frac{t}{\sqrt{n}}\right)t^2n^{-\frac{3}{2}}}{-\frac{1}{2}2n^{-\frac{3}{2}}} \\ &= \lim_{n \rightarrow \infty} \frac{t^2 \cdot L''\left(\frac{t}{\sqrt{n}}\right)}{2} \\ &= \lim_{n \rightarrow \infty} \frac{t^2 \cdot 1}{2} \\ &= \frac{t^2}{2} \end{aligned}$$

I w sumie to mieliśmy dowieść. Ale fajnie. □

23.1.2 Warianty

Istnieją różne warianty CTG, które mają swoje zastosowania w różnych sytuacjach. Poniżej podajemy wypowiedzi dwóch takich wariantów.

W pierwszym wariancie usuwamy warunek na to, że wszystkie zmienne X_i muszą mieć taki sam rozkład, ale musimy za to dodać dwa dodatkowe warunki.

Twierdzenie 23.1.3. Niech $\{X_i\}_{i \in \mathbb{N}}$ będzie ciągiem niezależnych zmiennych losowych spełniających $\mathbb{E}[X_i] = \mu_i$ i $\text{Var}[X_i] = \sigma_i^2$. Niech zachodzi

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

Wówczas dla dowolnych a, b zachodzi

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

Za to w drugim wariancie mając dodatkową informację o trzecim momencie, możemy wyznaczyć prędkość zbiegania do rozkładu normalnego

Twierdzenie 23.1.4 (Berry-Esséen). Istnieje taka stała C , że dla każdego ciągu niezależnych zmiennych losowych $\{X_i\}_{i \in \mathbb{N}}$ o tym samym rozkładzie ze skończoną wartością oczekiwaną μ i wariancją σ^2 oraz dla $\rho = \mathbb{E}[|X_i - \mu|^3] < \infty$ i $\widetilde{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ zachodzi

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

Wprowadzenie siły dwóch wyborów

Pytania 24-26 wszystkie są związane z siłą dwóch wyborów i wymagają wypowiedzi tych samych twierdzeń oraz modelu eksperymentu. Dodatkowo, częściowo używają tych samych oznaczeń oraz lematów. Aby uniknąć powtarzania tych segmentów, są one wszystkie obecne tutaj.

24.1 Wprowadzenie modelu eksperymentu

Rozważmy wariant standardowego eksperymentu z kulami i urnami. Rzucamy n kul sekwencyjnie do n urn i dla każdej kuli symulujemy dwa rzuty. Kula trafia do tej urny z dwóch wylosowanych która jest mniej wypełniona, a remisy rozstrzygamy dowolnie. Okazuje się, że znacząco zmienia to rozkład $\max(X_1, \dots, X_n)$, co pokazuje następujące twierdzenie

Twierdzenie 24.1.1. W opisanym powyżej modelu $\forall_{\alpha \geq 1} \exists_{n_0} \forall_{n \geq n_0}$ zachodzi

$$P\left(\frac{\ln \ln n}{\ln 2} - \mathcal{O}_\alpha(1) \leq \max(X_1, \dots, X_n) \leq \frac{\ln \ln n}{\ln 2} + \mathcal{O}_\alpha(1)\right) \geq 1 - \frac{1}{n^\alpha}$$

lub równoważnie

$$P(\log_2 \ln n - \mathcal{O}_\alpha(1) \leq \max(X_1, \dots, X_n) \leq \log_2 \ln n + \mathcal{O}_\alpha(1)) \geq 1 - \frac{1}{n^\alpha}$$

Dodatkowo, jeśli symulujemy d rzutów zamiast dwóch, w powyższym wzorze zamiast $\ln 2$ jest obecne $\ln d$ i musimy zamienić $\mathcal{O}_\alpha(1)$ na jakieś $\mathcal{O}_{d,\alpha}(1)$. Widzimy, że nie zmienia to bardzo naszych ograniczeń, a dowody tych wariantów są bardzo podobne, a więc dla uproszczenia nasze rozumowanie będziemy przeprowadzać dla $d = 2$.

24.2 Lematy pomocnicze

Lemat 24.2.1. Dla $Z \sim \text{Bin}(n, p)$ zachodzi

$$P(Z \geq 2np) \leq e^{-\frac{np}{3}}$$

$$P\left(Z \leq \frac{1}{2}np\right) \leq e^{-\frac{np}{8}}$$

Dowód. Pierwsza nierówność wynika wprost z punktu 2. 4.1.2 z $\delta = 1$, a druga z punktu 2.

28.5.1 z $\delta = \frac{1}{2}$.

$$\mathbb{E}[Z] = np$$

$$P(Z \geq 2np) = P(Z \geq (1 + \delta)\mathbb{E}[Z]) \leq e^{-\frac{1}{3}\mathbb{E}[Z]\delta^2} = e^{-\frac{np}{3}}$$

$$P\left(Z \leq \frac{1}{2}np\right) = P(Z \leq (1 - \delta)\mathbb{E}[Z]) \leq e^{-\frac{1}{2}\mathbb{E}[Z]\delta^2} = e^{-\frac{np}{8}}$$

□

Lemat 24.2.2. Niech $X_1, \dots, X_n$ to zmienne losowe, $Y_1, \dots, Y_n$ to binarne zmienne losowe, $Y_i = f_i(X_1, \dots, X_i)$ (jest wyznaczona przez $X_1, \dots, X_i$) oraz $Z \sim \text{Bin}(n, p)$ jest niezależna od poprzednich zmiennych. Jeśli dla każdego $i \in [n]$ oraz $(x_1, \dots, x_{i-1})$ takiego, że $P((X_1, \dots, X_{i-1}) = (x_1, \dots, x_{i-1})) > 0$ zachodzi

$$P(Y_i = 1 \mid (X_1, \dots, X_{i-1}) = (x_1, \dots, x_{i-1})) \leq p$$

to

$$P\left(\sum_{i=1}^n Y_i > k\right) \leq P(Z > k)$$

Dowód. (dla $n = 3$, dla wyższych n analogicznie)

$$\begin{aligned} P(Y_1 + Y_2 + Y_3 > k) &\leq P(Z_1 + Y_2 + Y_3 > k) \\ &= \sum_{x_1} P(Z_1 + Y_2 + Y_3 > k \mid X_1 = x_1) \cdot P(X_1 = x_1) \\ &\leq \sum_{x_1} P(Z_1 + Z_2 + Y_3 > k \mid X_1 = x_1) \cdot P(X_1 = x_1) \\ &= \sum_{(x_1, x_2)} P(Z_1 + Z_2 + Y_3 > k \mid (X_1, X_2) = (x_1, x_2)) \\ &\quad \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_1 = x_1) \\ &\leq \sum_{(x_1, x_2)} P(Z_1 + Z_2 + Z_3 > k \mid (X_1, X_2) = (x_1, x_2)) \\ &\quad \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_1 = x_1) \\ &= \sum_{(x_1, x_2)} P(Z_1 + Z_2 + Z_3 > k) \\ &\quad \cdot P(X_2 = x_2 \mid X_1 = x_1) \cdot P(X_1 = x_1) \\ &= P(Z_1 + Z_2 + Z_3 > k) \cdot \sum_{(x_1, x_2)} P(X_2 = x_2 \mid X_1 = x_1) P(X_1 = x_1) \\ &= P(Z_1 + Z_2 + Z_3 > k) \end{aligned}$$

□

Lemat 24.2.3. (Lemat dualny dla 24.2.2) Niech $X_1, \dots, X_n$ to zmienne losowe, $Y_1, \dots, Y_n$ to binarne zmienne losowe, $Y_i = f_i(X_1, \dots, X_i)$ (jest wyznaczona przez $X_1, \dots, X_i$) oraz $Z \sim \text{Bin}(n, p)$ jest niezależna od poprzednich zmiennych. Jeśli dla każdego $i \in [n]$ oraz $(x_1, \dots, x_{i-1})$

takiego, że $P((X_1, \dots, X_{i-1}) = (x_1, \dots, x_{i-1})) > 0$ zachodzi

$$P(Y_i = 1 \mid (X_1, \dots, X_{i-1}) = (x_1, \dots, x_{i-1})) \geq p$$

to

$$P\left(\sum_{i=1}^n Y_i > k\right) \geq P(Z > k)$$

a więc też

$$P\left(\sum_{i=1}^n Y_i < k\right) \leq P(Z < k)$$

Dowód. Tak samo jak w poprzednim lemacie. □

24.3 Oznaczenia

Do dowodów ograniczenia górnego przez iterację ograniczeń oraz ograniczenia dolnego przydadzą nam się funkcje pomocnicze. Dla $t \in [n]$ niech

- $h(t)$ to wysokość t -tej kuli, czyli liczba kul w urnie, w której wylądowała t -ta kula zaraz po jej wrzuceniu
- $\nu_i(t)$ to liczba urn zawierających $\geq i$ kul zaraz po wrzuceniu t -tej kuli
- $\mu_i(t)$ to liczba kul o wysokości $\geq i$ zaraz po wrzuceniu t -tej kuli

Prosto widzimy, że

$$\forall_{i \in [n], t \in [n]} \nu_i(t) \leq \mu_i(t)$$

bo w każdej urnie zawierającej $\geq i$ kul jest przynajmniej jedna kula wysokości $\geq i$.

Pytanie 24

Siła dwu wyborów. Model eksperymentu, wypowiedzi twierdzeń. Dowód ograniczenia górnego (ten w książce) przez iterację ograniczeń na prawdopodobieństwa warunkowe pewnych zdarzeń.

W tej sekcji znajduje się wyłącznie dowód ograniczenia górnego przez iterację ograniczeń na prawdopodobieństwa warunkowe. Model eksperymentu oraz wypowiedzi twierdzeń znajdują się w rozdziale 24.

25.1 Ograniczenie górne przez iterację ograniczeń

Przejdźmy wreszcie do pierwszego dowodu ograniczenia górnego twierdzenia 24.1.1. Tak naprawdę będziemy dowodzić że

$$P(\max(X_1, \dots, X_n) \geq \log_2 \ln n + \mathcal{O}(1)) < \frac{1}{n^{\alpha-1}}$$

ale jak się temu przyjrzymy to możemy zauważyć, że z odpowiednim α da nam to co chcemy.

Niech

$$\beta_4 = \frac{n}{4}$$
$$\beta_{i+1} = 2 \frac{\beta_i^2}{n} \text{ dla } 4 \leq i \leq i^*$$

i^* zostanie zdefiniowane później.

Lemat 25.1.1.

$$\beta_{i+4} = \frac{n}{2^{2^i+1}}$$

a więc

$$\beta_{i+4} \leq \frac{n}{2^{2^i}}$$

Dowód. Przeprowadźmy dowód indukcyjny. Dla $i = 0$

$$\beta_4 = \frac{n}{4} = \frac{n}{2^{1+1}} = \frac{n}{2^{2^0+1}} = \beta_{0+4}$$

Jeśli lemat zachodzi dla i to

$$\beta_{i+5} = 2 \frac{\beta_{i+4}^2}{n} = 2 \frac{\left(\frac{n}{2^{2^i+1}}\right)^2}{n} = 2 \frac{n}{2^{2(2^i+1)}} = 2 \frac{n}{2^{2^{i+1}+2}} = \frac{n}{2^{2^{i+1}+1}} = \beta_{(i+1)+4}$$

a więc zachodzi też dla $i+1$ □

Zdefiniujmy teraz $\mathcal{E}_i$ jako zdarzenie zachodzące jeśli $\nu_i(n) \leq \beta_i$. Widzimy, że

$$P(\mathcal{E}_4) = P\left(\nu_4(n) \leq \frac{n}{4}\right) = 1$$

bo oczywiście $\geq \frac{n}{4}$ kul może mieć co najwyżej $\frac{n}{4}$ urn.

Chcemy teraz wykazać, że jeśli $\mathcal{E}_i$ zaszło, to prawie na pewno $\mathcal{E}_{i+1}$ też zaszło. Dla $t \in [n]$ definiujemy binarną zmienną losową Y_t

$$Y_t = \begin{cases} 1 & \text{dla } h(t) \geq i+1 \wedge \nu_i(t-1) \leq \beta_i \\ 0 & \text{wpp.} \end{cases}$$

Zauważmy, że dla $(K_1, \dots, K_{t-1})$ będącego zmiennymi losowymi reprezentującymi urny do których trafiły kolejne kule oraz $(\omega_1, \dots, \omega_{t-1})$ takiego, że $\forall_{i \in [t-1]} \omega_i \in [n]$ zachodzi

$$P(Y_t = 1 \mid (K_1, \dots, K_{t-1}) = (\omega_1, \dots, \omega_{t-1})) \leq \left(\frac{\beta_i}{n}\right)^2$$

Jest tak, ponieważ aby zaszło $h(t) \geq i+1$ t -ta kula musi obiema symulacjami trafić w urnę o przynajmniej i kulach, ale wiemy z $\nu_i(t-1) \leq \beta_i$ że takich urn jest co najwyżej β_i z n możliwych.

Jeśli $\mathcal{E}_i$ zaszło, to $\mu_{i+1}(n) = \sum_{t \in [n]} Y_t$, ponieważ z $\mathcal{E}_i$ wiemy, że $\nu_i(t-1) \leq \beta_i$ jest prawdziwe dla każdego t , a więc $Y_t = 1 \iff h(t) \geq i+1$, a więc suma Y_t zlicza liczbę kul o wysokości $\geq i+1$, co jest dokładnie definicją $\mu_{i+1}(n)$.

Niech $p_i := \left(\frac{\beta_i}{n}\right)^2$. Z lematu 24.2.2 dostajemy

$$P\left(\sum_{t \in [n]} Y_t > k\right) \leq P(\text{Bin}(n, p_i) > k)$$

Teraz przy użyciu tego pokażemy, że $P(\neg \mathcal{E}_{i+1})$ jest małe, jeżeli $\mathcal{E}_i$ zaszło. Najpierw policzymy

prawdopodobieństwo warunkowe:

$$\begin{aligned} P(\neg \mathcal{E}_{i+1} \mid \mathcal{E}_i) &= P(\nu_{i+1}(n) > \beta_{i+1} \mid \mathcal{E}_i) \\ &\leq P(\mu_{i+1}(n) > \beta_{i+1} \mid \mathcal{E}_i) \\ &= P\left(\sum_{t \in [n]} Y_t > 2np_i \mid \mathcal{E}_i\right) \end{aligned}$$

Tutaj w ostatniej równości robimy dwie rzeczy: korzystamy z faktu, że jeżeli $\mathcal{E}_i$ zaszło, to $\mu_{i+1} = \sum_{t \in [n]} Y_t$, oraz z definicji β_{i+1} i p_i podstawiamy $\beta_{i+1} = 2np_i$

$$\begin{aligned} P\left(\sum_{t \in [n]} Y_t > 2np_i \mid \mathcal{E}_i\right) &\leq \frac{P\left(\sum_{t \in [n]} Y_t > 2np_i\right)}{P(\mathcal{E}_i)} \\ &\leq \frac{P(\text{Bin}(n, p_i) > 2np_i)}{P(\mathcal{E}_i)} \end{aligned}$$

W drugiej nierówności korzystamy z faktu z lematu 24.2.2. Przekształcając dalej:

$$\begin{aligned} \frac{P(\text{Bin}(n, p_i) > 2np_i)}{P(\mathcal{E}_i)} &\leq \frac{1}{e^{\frac{np_i}{3}} P(\mathcal{E}_i)} \\ &\leq \frac{1}{n^\alpha P(\mathcal{E}_i)} \end{aligned}$$

Na koniec korzystamy z nierówności Czernowa (lemat 24.2.1), oraz zakładamy, że $np_i \geq 3\alpha \ln(n)$. Ostatecznie otrzymujemy:

$$P(\neg \mathcal{E}_{i+1} \mid \mathcal{E}_i) \leq \frac{1}{n^\alpha P(\mathcal{E}_i)}$$

Teraz możemy ograniczyć $P(\neg \mathcal{E}_{i+1})$:

$$\begin{aligned} P(\neg \mathcal{E}_{i+1}) &= P(\neg \mathcal{E}_{i+1} \mid \mathcal{E}_i) P(\mathcal{E}_i) + P(\neg \mathcal{E}_{i+1} \mid \neg \mathcal{E}_i) P(\neg \mathcal{E}_i) \\ &\leq P(\neg \mathcal{E}_{i+1} \mid \mathcal{E}_i) P(\mathcal{E}_i) + P(\neg \mathcal{E}_i) \\ &\leq \frac{1}{n^\alpha} + P(\neg \mathcal{E}_i) \end{aligned}$$

Gdzie na końcu ponownie korzystamy z założenia, że $np_i \geq 3\alpha \ln(n)$. Pokazujemy przez indukcję:

$$P(\neg \mathcal{E}_{i+1}) \leq \frac{1}{n^\alpha} + P(\neg \mathcal{E}_i) \leq \frac{1}{n^\alpha} + \frac{i}{n^\alpha} = \frac{i+1}{n^\alpha}$$

Ponieważ wiemy, że $P(\mathcal{E}_4) = 1$, zatem $P(\neg \mathcal{E}_4) = 0 \leq \frac{4}{n^\alpha}$

Teraz wracamy do i^* i definiujemy je jako minimalne i takie, że $np_i < 3\alpha \ln n$. Możemy zauważyć, że to co pokazaliśmy z indukcji jest prawdziwe jedynie dla $i < i^*$, lecz działa jeszcze dla $P(\neg \mathcal{E}_{i^*})$ bo pracowaliśmy na $\mathcal{E}_{i+1}$, a więc $P(\neg \mathcal{E}_{i^*}) \leq \frac{i^*}{n^\alpha}$

Chcemy jeszcze znaleźć i^* więc rozważamy następujący ciąg nierówności:

$$np_i < 3\alpha \ln n$$

Podstawiamy $p_i = \left(\frac{\beta_i}{n}\right)^2$

$$n \left(\frac{\beta_i}{n}\right)^2 < 3\alpha \ln n$$

Tworzymy mocniejszą nierówność, korzystając z faktu, że $\beta_i \leq \frac{n}{2^{2^{i-4}}}$

$$n \left(\frac{\frac{n}{2^{2^{i-4}}}}{n}\right)^2 < 3\alpha \ln n$$

$$\frac{n}{3\alpha \ln n} < 2^{2^{i-4} \cdot 2} = 2^{2^{i-3}}$$

Logarytmujemy stronami i podstawiamy $c = \frac{1}{\ln 2}$:

$$c \cdot \ln n - c \cdot \ln(3\alpha) - \log_2 \ln n < 2^{i-3}$$

Ponownie tworzymy mocniejszą nierówność przez zwiększenie lewej strony:

$$c \cdot \ln n < 2^{i-3}$$

Znowu logarytmujemy stronami:

$$\ln c + \log_2 \ln n < i - 3$$

Ostatecznie otrzymujemy

$$\log_2 \ln n + \mathcal{O}(1) < i \implies i^* \leq \log_2 \ln n + \mathcal{O}(1)$$

Znając i^* możemy dalej ograniczać prawdopodobieństwo, że $\nu_i(n)$ będzie duże dla $i \geq i^*$

$$\begin{aligned}
 P(\nu_{i^*+1}(n) > 9\alpha \ln n \mid \mathcal{E}_{i^*}) &\leq P(\mu_{i^*+1}(n) > 9\alpha \ln n \mid \mathcal{E}_{i^*}) \\
 &= P\left(\sum_{t \in [n]} Y_t > 9\alpha \ln n \mid \mathcal{E}_{i^*}\right) \\
 &\leq \frac{P(\text{Bin}(n, p_i) > 9\alpha \ln n)}{P(\mathcal{E}_{i^*})} \\
 &\leq \frac{P(\text{Bin}(n, \frac{3\alpha \ln n}{n}) > 9\alpha \ln n)}{P(\mathcal{E}_{i^*})} \\
 &\leq \frac{1}{n^\alpha P(\mathcal{E}_{i^*})}
 \end{aligned}$$

Pierwsze przekształcenia robimy tak samo jak wyżej i korzystając z $np_i < 3\alpha \ln n$. Ostatnia nierówność wynika z Czernowa:

$$P(X > (1+2) \cdot 3\alpha \ln n) \leq \left(\frac{e^2}{3^3}\right)^{3\alpha \ln n} < \left(e^{-\frac{1}{3}}\right)^{3\alpha \ln n} = \frac{1}{n^\alpha}$$

Z tego otrzymujemy:

$$\begin{aligned}
 P(\nu_{i^*+1}(n) > 9\alpha \ln n) &\leq P(\neg \mathcal{E}_{i^*}) P(\nu_{i^*+1}(n) > 9\alpha \ln n \mid \neg \mathcal{E}_{i^*}) + P(\mathcal{E}_{i^*}) P(\nu_{i^*+1}(n) > 9\alpha \ln n \mid \mathcal{E}_{i^*}) \\
 &\leq P(\neg \mathcal{E}_{i^*}) + \frac{1}{n^\alpha} \\
 &\leq \frac{i^* + 1}{n^\alpha}
 \end{aligned}$$

Zostało nam ostatnie warunkowe ograniczenie:

$$\begin{aligned}
 P(\mu_{i^*+2}(n) \geq \beta \mid \nu_{i^*+1}(n) \leq 9\alpha \ln n) &\leq \frac{P\left(\text{Bin}(n, \left(\frac{9\alpha \ln n}{n}\right)^2) \geq \beta\right)}{P(\nu_{i^*+1}(n) \leq 9\alpha \ln n)} \\
 &\leq \frac{n^\beta \left(\frac{9\alpha \ln n}{n}\right)^{2\beta}}{P(\nu_{i^*+1}(n) \leq 9\alpha \ln n)} \\
 &\leq \frac{(9\alpha \ln n)^{2\beta}}{n^\beta P(\nu_{i^*+1}(n) \leq 9\alpha \ln n)}
 \end{aligned}$$

Gdzie w pierwszej nierówności jako p dla w rozkładzie dwumianowym podstawiamy prawdopodobieństwo, że dwukrotnie trafimy do wystarczająco wypełnionych urn. Druga nierówność to union bound.

Teraz mamy wszystko, żeby pokazać górne ograniczenie.

$$\begin{aligned}
P(\nu_{i^*+2+2\alpha}(n) \geq 1) &\leq P(\mu_{i^*+2+2\alpha}(n) \geq 1) \\
&\leq P(\mu_{i^*+2}(n) \geq 2\alpha) \\
&\leq P(\mu_{i^*+2}(n) \geq 2\alpha \mid \nu_{i^*+1}(n) \leq 9\alpha \ln n) P(\nu_{i^*+1}(n) \leq 9\alpha \ln n) \\
&\quad + P(\nu_{i^*+1}(n) > 9\alpha \ln n) \\
&\leq \frac{(9\alpha \ln n)^{4\alpha}}{n^{2\alpha}} + \frac{i^* + 1}{n^\alpha} \\
&< \frac{1}{n^{\alpha-1}}
\end{aligned}$$

Gdzie ostatnia nierówność zachodzi dla odpowiednio dużego n .

Pytanie 25

(★). Siła dwu wyborów. Model eksperymentu, wypowiedzi twierdzeń. Dowód ograniczenia górnego przez drzewa świadczące.

W tej sekcji znajduje się wyłącznie dowód ograniczenia górnego przez drzewa świadczące. Model eksperymentu oraz wypowiedzi twierdzeń znajdują się w rozdziale 24.

26.1 Ograniczenie górne przez drzewa świadczące

Dowód ograniczenia górnego przez drzewa świadczące wzorowany artykułem.

Definicje drzew świadczących

Jak poprzednio, rzucając kulę losujemy jednostajnie $d \geq 2$ urn i wrzucamy kulę do najmniejszej. Będziemy dalej nazywać i -tą wylosowaną urnę dla kuli b i -tą lokacją kuli b .

Definicja 26.1.1. Drzewo świadczące rzędu L to pełne, ukorzenione drzewo d -arne wysokości L . Każdy wierzchołek w drzewie reprezentuje pewną kulę, niekoniecznie unikalną - kula może być reprezentowana przez wiele wierzchołków. Dodatkowo, dla każdego wierzchołka v niebędącego liściem, kule odpowiadające dzieciom v muszą już znajdować się w urnach w momencie rzucania kuli v .

Ogólna idea dowodu będzie następująca; rozważmy wszystkie drzewa świadczące. Powiemy za moment, co oznacza że jakieś konkretne drzewo świadczące jest *aktywne*, a następnie pokażemy, że jeśli w którejś z urn jest dużo kul, to któreś z drzew świadczących jest *aktywne*. Tym samym prawdopodobieństwo, że któraś urna ma dużo kul jest ograniczone przez prawdopodobieństwo, że jest jakieś *aktywne* drzewo świadczące.

Definicja 26.1.2. Dodajmy sobie dodatkową interpretację dla struktury dowolnego drzewa świadczącego.

- **Krawędziozdarzenie** ("*edge event*") - dla krawędzi $e = (u, v)$, gdzie v jest i -tym dzieckiem u , krawędź e określa zdarzenie, że i -ta lokacja kuli u jest taka sama, co któraś z lokacji kuli v .

- **Liściozdarzenie** (*"leaf event"*) - liść drzewa świadczącego v określa zdarzenie, że każda z d lokacji kuli v wskazuje na urnę, gdzie znajdują się już co najmniej 3 inne kule, które nie są zareprezentowane wierzchołkami drzewa.

Definicja 26.1.3. Krawędź lub liść drzewa świadczącego są aktywne, jeśli podczas rzucania kul do urn zaszyły odpowiadające im krawędzio- lub liściozdarzenia. Drzewo świadczące jest aktywne, jeśli wszystkie jego krawędzie i liście są aktywne.

Konstrukcja drzew świadczących

Z początku będziemy zakładać, że zdarzenia opisywane przez drzewa świadczące są niezależne od siebie. W szczególności, rozważamy tylko drzewa, gdzie wierzchołki reprezentują parami różne kule - ta prostsza wersja posłuży później do pełnego dowodu.

Lemat 26.1.1. Jeśli któraś z urn posiada więcej niż $L + 3$ kule, to istnieje aktywne drzewo świadczące rzędu L .

Dowód. Niech urna x ma co najmniej $L + 4$ kule. Skonstruujemy aktywne drzewo następująco.

Niech korzeń reprezentuje ostatnio wrzuconą kulę z x . Zauważmy, że każda z d lokacji tej kuli wskazuje na urnę, gdzie były już $L + 3$ kule. Przypisujemy dzieciom korzenia kule, które znajdowały się najwyżej w tych urnach w momencie wrzucania kuli korzenia. Dalej postępujemy tak samo z dziećmi aż całe drzewo zostanie skonstruowane.

Można zauważyć, że kula korzenia została wrzucona do swojej urny po co najmniej $L + 3$ innych kulach. Podobnie w momencie wrzucania kul każdego z dzieci korzenia, w ich urnach były już co najmniej $L + 2$ kule itd.

Zauważmy też, że każda kula reprezentowana przez liść miała już co najmniej 3 inne kule w urnie w momencie jej wrzucania.

Otrzymaliśmy drzewo świadczące o wszystkich krawędziach i liściach aktywnych, więc jest to drzewo aktywne. $\square$

Jeśli udałoby się ograniczyć z góry prawdopodobieństwo na istnienie aktywnego drzewa świadczącego rzędu L , to byłoby to także ograniczenie na prawdopodobieństwo, że istnieje urna o co najmniej $L + 4$ kulach.

Jeśli w drzewie świadczącym jest m wierzchołków, to możemy przypisać im kule na n^m sposobów. Prawdopodobieństwo, że krawędź (u, v) jest aktywna jest co najwyżej $\frac{d}{n}$, ponieważ prawdopodobieństwo, że i -ta lokacja u trafi w wybraną z lokacji v jest co najwyżej $\frac{1}{n}$. Z niezależności prawdopodobieństwo, że wszystkie krawędzie są aktywne jest co najwyżej $\left(\frac{d}{n}\right)^{m-1}$. Prawdopodobieństwo aktywacji konkretnego liścia jest co najwyżej $\frac{1}{3^d}$, ponieważ każda z lokacji kuli liścia musi trafić w urnę o co najmniej 3 innych kulach - a takich urn jest co najwyżej $\frac{n}{3}$. Tym samym gdy oznaczymy przez q liczbę liści w drzewie, to prawdopodobieństwo aktywacji wszystkich liści jest co najwyżej 3^{-dq} .

Łącząc to wszystko ze sobą i nakładając union bounda dostajemy ograniczenie na istnienie aktywnego drzewa świadczącego:

$$n^m \cdot \left(\frac{d}{n}\right)^{m-1} \cdot 3^{-dq}$$

Zachodzą także następujące ograniczenia: $m \leq 2q$ oraz $2d^2 \leq 3^d$. Gdy podstawimy je wraz z $q = d^L$ dostaniemy ograniczenie:

$$\begin{aligned} n^m \cdot \left(\frac{d}{n}\right)^{m-1} \cdot 3^{-dq} &= n \cdot d^{m-1} \cdot 3^{-dq} \leq n \cdot d^{2q} \cdot 3^{-dq} \\ &\leq n \cdot 2^{-q} \cdot 3^{dq} \cdot 3^{-dq} = n \cdot 2^{-d^L} \end{aligned}$$

A gdy weźmiemy $L \geq \log_d \log_2 n + \log_d(1 + \alpha)$ dla $\alpha > 0$, ograniczymy prawdopodobieństwo istnienia odpowiedniego aktywnego drzewa świadczącego przez $n^{-\alpha}$.

Konstrukcja pełnych drzew świadczących

W poprzedniej części dowodu rozważaliśmy tylko prawdopodobieństwo aktywacji drzew świadczących o parami różnych kulach. W rzeczywistości jednak ta sama kula może wystąpić w drzewie wielokrotnie - i wtedy zdarzenia konstruujące drzewo nie są już niezależne. W tym celu możemy poodcinać niektóre wierzchołki, które nam się nie będą podobać, aby przywrócić niezależność. Odcinanie wierzchołków zwiększa jednak prawdopodobieństwo aktywacji takiego drzewa, więc zaczniemy od większej struktury.

Definicja 26.1.4. Pełne drzewo świadczące rzędu L dla pewnej stałej $\kappa \in \mathbb{N}_2$ ma następującą konstrukcję: Korzeń drzewa ma κ dzieci, każde z których ma dokładnie jedno dziecko - tym samym korzeń ma także κ wnuków. Każdy z wnuków jest korzeniem standardowego drzewa świadczącego rzędu L . W dodatku, w pełnym drzewie świadczącym kule przypisane do dzieci korzenia muszą być parami różne, a korzeń jako jedyny nie ma przypisanej kuli.

Lemat 26.1.2. Jeśli któraś z urn posiada więcej niż $L + 3 + \kappa$ kul, to istnieje aktywne pełne drzewo świadczące.

Dowód. Niech urna x posiada co najmniej $L+4+\kappa$ kul. Wybieramy z niej κ ostatnio wrzuconych kul i przyporządkowujemy je dzieciom korzenia. Następnie przyporządkujemy kule wnukom. Rozważmy kulę b przypisaną do v - dziecka korzenia. Przynajmniej jedna z lokacji b wskazuje na urnę x - jeśli jest ich więcej, należy wybrać jedną z nich. Niech i to indeks tej lokacji. Weźmy następną lokację, czyli $i+1 \bmod d$. W momencie wrzucania b , ta lokacja wskazuje na urnę o co najmniej $L+3$ kulach. Ostatnio wrzuconą do tej urny kulę przed wrzuceniem b przypisujemy do dziecka v , czyli wnuka korzenia. $\square$

Przycinanie pełnych drzew świadczących

Definicja 26.1.5. Zdefiniujemy **przycięte pełne drzewo świadczące** konstrukcyjnie. Zaczynając od pełnego drzewa świadczącego, będziemy odcinać jego krawędzie w odpowiedni sposób. Począwszy od korzenia, przeglądamy drzewo w kolejności BFS. Za każdym razem gdy przeglądamy wierzchołek v , który reprezentuje kulę, która była już widziana wcześniej, przecinamy krawędź łączącą v z jego rodzicem. Odcinamy w ten sposób v wraz z całym jego poddrzewem. Zauważmy, że procedura ta nie odetnie nigdy dzieci korzenia, gdyż reprezentują kule parami różne. Przecięte w ten sposób krawędzie będziemy dalej nazywać *odciętymi* ("cutoff edge"). Kontynuujemy odcinanie do momentu, aż przeglądnijemy całe drzewo albo odetniemy κ krawędzi. Do przyciętego drzewa świadczącego trafiają jednak tylko przeglądnięte wierzchołki i krawędzie, wraz z krawędziami odciętymi, jako świadectwo miejsc odcięcia.

Wyróżniamy dwa przypadki.

Przypadek 1 - mniej niż κ odciętych

Jeśli odcięliśmy mniej niż κ krawędzi, to znaczy, że jeden z wnuków korzenia wraz z poddrzewem przetrwał przycinanie. To oznacza, że mamy aktywne drzewo świadczące o parami różnych kulach, a prawdopodobieństwo na to już ograniczyliśmy przez $n^{-\alpha}$.

Przypadek 2 - κ odciętych

Pozostaje jedynie ograniczyć prawdopodobieństwo wystąpienia aktywnego przyciętego drzewa świadczącego o κ odciętych krawędziach. Ograniczymy to prawdopodobieństwo przy założeniu, że liczba kul reprezentowanych przez pełne drzewo świadczące jest co najwyżej $M = 2\kappa(\alpha + 1) \log_2 n$. Jest co najwyżej M^κ sposobów na przycięcie pełnego drzewa świadczącego - zamianę pełnego drzewa świadczącego w przycięte. Niech m będzie liczbą kul reprezentowanych przez przycięte drzewo, a q to liczba liści. Liczba sposobów na dopasowanie kul do wierzchołków to co najwyżej n^m . Przyjmijmy, że korzeń przyciętego drzewa (który nie ma przypisanej kuli) zamiast tego ma przypisaną urnę, z której wyjęliśmy kule dla jego dzieci. Jest n sposobów na wybranie tej urny, i mając tę urnę prawdopodobieństwo aktywacji wszystkich krawędzi to co najwyżej $(\frac{d}{n})^m$, a prawdopodobieństwo, że wszystkie liście są aktywne jest ograniczone przez 3^{-dq} .

Okazuje się, że dostaliśmy niemal to samo ograniczenie. Mimo to, q i m mogą być mniejsze niż poprzednio. Możemy jednak wciąż wyciągnąć lepsze ograniczenie patrząc na odcięte krawędzie.

Każda odcięta krawędź jest świadkiem, że jakaś kula b reprezentowana przez nieprzycięty wierzchołek u dzieli lokację z jakąś kulą b' innego wierzchołka u' - jest to powód, dlaczego ta krawędź została odcięta. Wierzchołek u' został przeglądnięty przed u , więc z pewnością jest częścią przyciętego drzewa świadczącego. Przycięta krawędź opisuje, która lokacja kuli b trafiła w lokację dzieloną z kulą b' . Liczba możliwości na wybranie u' i tym samym b' jest ograniczona z góry przez $m \leq M$. Prawdopodobieństwo, że wybrana lokacja b trafi w tę samą urnę co któraś z lo-

kacji b' jest co najwyżej $\frac{d}{n}$. Tym samym prawdopodobieństwo posiadania κ odciętych krawędzi ograniczamy przez $\left(M \frac{d}{n}\right)^\kappa$.

Możemy teraz zauważyć, że skoro pozbyliśmy się zbędnych kul, a wszystkie liściozdarzenia, krawędziozdarzenia i odcięciozdarzenia są niezależne, to pozostaje przejść do ostatecznego ograniczenia, co następuje:

$$M^k \cdot n^{m+1} \cdot \left(\frac{d}{n}\right)^m \cdot 3^{-dq} \cdot \left(\frac{Md}{n}\right)^\kappa = n \cdot d^m \cdot 3^{-dq} \cdot \left(\frac{M^2d}{n}\right)^\kappa$$

Stosujemy teraz podobne ograniczenia jak poprzednio: $m \leq 2q$, $d^2 \leq 3^d$ oraz $M \leq 2\kappa(\alpha + 1) \log_2 n$.

$$\begin{aligned} n \cdot d^m \cdot 3^{-dq} \cdot \left(\frac{M^2d}{n}\right)^\kappa &\leq n \cdot \left(\frac{M^2d}{n}\right)^\kappa \\ &\leq n \left(\frac{d(2\kappa(\alpha + 1) \log_2 n)}{n}\right)^\kappa = n^{-\kappa+1+o(1)} \end{aligned}$$

To zamyka ograniczenie drugiego przypadku, a tym samym cały dowód.

Pytanie 26

(★). Siła dwu wyborów. Model eksperymentu, wypowiedzi twierdzeń. Dowód ograniczenia dolnego.

W tej sekcji znajduje się wyłącznie dowód ograniczenia dolnego. Model eksperymentu oraz wypowiedzi twierdzeń znajdują się w rozdziale 24.

27.1 Ograniczenie dolne

Teraz będziemy dowodzić, że:

$$P(\log_2 \ln n - \mathcal{O}(1) \leq \max(X_1, \dots, X_n)) \geq 1 - \frac{1}{n}$$

Czyli dowodzimy ograniczenie dolne twierdzenia 24.1.1. Co ważne, nie robimy tego w pełnym wariancie tego twierdzenia, czyli dla n^α dla dowolnego α , a wyłącznie dla $\alpha = 1$. Tak samo jest jednak w książce oraz w materiałach które otrzymaliśmy i na egzaminie wystarczy znajomość poniższego dowodu. Sam dowód przebiega bardzo podobnie do iteracyjnego ograniczenia górnego.

Niech

$$\begin{aligned}\gamma_0 &= n \\ \gamma_{i+1} &= \frac{n}{2^{i+3}} \cdot \left(\frac{\gamma_i}{n}\right)^2\end{aligned}$$

Lemat 27.1.1.

$$\gamma_i = \frac{n}{2^{\sum_{k=0}^{i-1} (i+2-k)2^k}}$$

W szczególności:

$$\gamma_i \geq \frac{n}{2^{4 \cdot 2^i}}$$

Dowód. Można to pokazać indukcyjnie. Zauważmy, że dla $i = 0$ suma w potęgze mianownika wyniesie 0, zatem faktycznie $\gamma_0 = n$. Zakładamy, że powyższy wzór jest prawdziwy dla γ_i .

Pokazujemy prawdziwość dla γ_{i+1} :

$$\begin{aligned}
 \gamma_{i+1} &= \frac{n}{2^{i+3}} \left(\frac{\gamma_i}{n} \right)^2 \\
 &= \frac{n}{2^{i+3}} \cdot \left(\frac{1}{2^{\sum_{k=0}^{i-1} (-+2-k)2^k}} \right)^2 \\
 &= \frac{n}{2^{i+3} \cdot 2^{2 \cdot \sum_{k=0}^{i-1} (i+2-k)2^k}} \\
 &= \frac{n}{2^{i+3} \cdot 2^{\sum_{k=0}^{i-1} (i+2-k)2^{k+1}}} \\
 &= \frac{n}{2^{i+3} \cdot 2^{\sum_{k=1}^i (i+3-k)2^k}} \\
 &= \frac{n}{2^{\sum_{k=0}^i (i+3-k)2^k}}
 \end{aligned}$$

□

Zdefiniujmy teraz $\mathcal{F}_i$ jako zdarzenie zachodzące jeśli $\nu_i(n(1 - \frac{1}{2^i})) \geq \gamma_i$. Widzimy, że $P(\mathcal{F}_0) = 1$. Będziemy wykazywać, że $P(\neg \mathcal{F}_{i+1} \mid \mathcal{F}_i)$ jest małe.

Ustalmy i oraz $t \in R = [n(1 - \frac{1}{2^i}), n(1 - \frac{1}{2^{i+1}})]$. Zdefiniujmy zmienną losową Z_t :

$$Z_t = 1 \Leftrightarrow h(t) = i + 1 \vee \nu_{i+1}(t - 1) \geq \gamma_{i+1}$$

Dodatkowo:

$$P(h(t) = i + 1) = \left(\frac{\nu_i(t - 1)}{n} \right)^2 - \left(\frac{\nu_{i+1}(t - 1)}{n} \right)^2$$

Ponieważ aby t -ta kula była na $(i + 1)$ -szym poziomie, to musimy dwukrotnie wylosować urnę mającą przynajmniej i kul (lecz nie wylosować dwa razy urny mającej przynajmniej $i + 1$ kul). Ograniczamy z dołu $P(Z_t = 1)$, znając wyniki poprzednich rzutów:

$$\begin{aligned}
 P(Z_t = 1 \mid (K_1, \dots, K_{t-1}) = (\omega_1, \dots, \omega_{t-1}), \mathcal{F}_i) &\geq \left(\frac{\gamma_i}{n} \right)^2 - \left(\frac{\gamma_{i+1}}{n} \right)^2 \\
 &\geq \left(\frac{\gamma_i}{n} \right)^2 - \left(\frac{\frac{n}{2^{i+3}} \left(\frac{\gamma_i}{n} \right)^2}{n} \right)^2 \\
 &\geq \frac{1}{2} \left(\frac{\gamma_i}{n} \right)^2
 \end{aligned}$$

Pierwsza nierówność wynika z tego, że albo $\nu_{i+1}(t - 1) \geq \gamma_{i+1}$, co samo daje nam $Z_t = 1$, albo $\nu_{i+1}(t - 1) < \gamma_{i+1}$ oraz $\nu_i(t - 1) \geq \gamma_i$ (z warunkowania po $\mathcal{F}_i$)

$$P(Z_t = 1 \mid (K_1, \dots, K_{t-1}) = (\omega_1, \dots, \omega_{t-1}), \mathcal{F}_i) \geq p_i := \frac{1}{2} \left(\frac{\gamma_i}{n} \right)^2$$

Z lematu 24.2.3 otrzymujemy:

$$\begin{aligned} \mathbb{P}\left(\sum_{t \in R} Z_t < \gamma_{i+1} \mid \mathcal{F}_i\right) &\leq \mathbb{P}\left(\text{Bin}\left(\frac{n}{2^{i+1}}, p_i\right) < \gamma_{i+1}\right) \\ &\leq e^{-\frac{np_i}{2^{i+1}} \frac{1}{8}} \\ &< \frac{1}{n^2} \end{aligned}$$

Gdzie druga nierówność wynika z 24.2.1, a trzecia działa dla $\frac{np_i}{2^{i+1}} \geq 17 \ln n$.

Jeżeli zaszło $\neg \mathcal{F}_{i+1}$ to bezpośrednio z definicji $\mathcal{F}_i$ mamy

$$\forall_{t \in R} \nu_{i+1}(t-1) \leq \nu_{i+1}\left(n\left(1 - \frac{1}{2^{i+1}}\right)\right) < \gamma_{i+1}$$

W takim razie, dalej dla $\neg \mathcal{F}_{i+1}$, otrzymujemy

$$\begin{aligned} \sum_{t \in R} Z_t &= \sum_{t \in R} [h(t) = i+1] \\ &\leq \nu_{i+1}\left(n\left(1 - \frac{1}{2^{i+1}}\right)\right) \\ &< \gamma_{i+1} \end{aligned}$$

Skoro $\neg \mathcal{F}_{i+1} \implies \sum_{t \in R} Z_t < \gamma_{i+1}$ to

$$\mathbb{P}(\neg \mathcal{F}_{i+1} \mid \mathcal{F}_i) \leq \mathbb{P}\left(\sum_{i \in R} Z_t < \gamma_{i+1} \mid \mathcal{F}_i\right) < \frac{1}{n^2}$$

Gdzie ostatnia nierówność ponownie jest prawdziwa dla takich i , że $\frac{np_i}{2^{i+1}} \geq 17 \ln n$. Niech i^* będzie największym i spełniającym tę nierówność. Wtedy:

$$\mathbb{P}(\mathcal{F}_{i^*}) \geq \mathbb{P}(\mathcal{F}_{i^*} \mid \mathcal{F}_{i^*-1}) \dots \mathbb{P}(\mathcal{F}_1 \mid \mathcal{F}_0) \mathbb{P}(\mathcal{F}_0) \geq \left(1 - \frac{1}{n^2}\right)^{i^*}$$

Teraz chcielibyśmy poznać i^* . Podobnie jak dla poprzedniego dowodu, tworzymy sobie ciąg coraz mocniejszych nierówności. Przypomnijmy, że $p_i = \frac{1}{2} \left(\frac{\gamma_i}{n}\right)^2$ oraz $\gamma_{i+1} = \frac{n}{2^{i+3}} \cdot \left(\frac{\gamma_i}{n}\right)^2$

$$\frac{np_i}{2^{i+1}} = \frac{n}{2^{i+2}} \left(\frac{\gamma_i}{n}\right)^2 = 2\gamma_{i+1}$$

$$2\gamma_{i+1} \geq 17 \ln n$$

Ponieważ $\gamma_i \geq \frac{n}{2^{4 \cdot 2^i}}$ (z 27.1.1):

$$2 \cdot \frac{n}{2^{4 \cdot 2^{i+1}}} \geq 17 \ln n$$

$$2^{4 \cdot 2^{i+1}} \leq \frac{2n}{17 \ln n}$$

$$4 \cdot 2^{i+1} \leq \log_2 2 + \log_2 n - \log_2(17 \ln n)$$

Dla wystarczająco dużego n , to jest mocniejsza nierówność:

$$4 \cdot 2^{i+1} \leq \frac{1}{2 \ln 2} \ln n$$

$$2 + i + 1 \leq \log_2 \ln n + \log_2 \left(\frac{1}{2 \ln 2} \right)$$

$$i \leq \log_2 \ln n - \mathcal{O}(1)$$

Zatem

$$i^* \geq \log_2 \ln n - \mathcal{O}(1) = \frac{\ln \ln n}{\ln 2} - \mathcal{O}(1)$$

oraz:

$$P(\mathcal{F}_{i^*}) \geq \left(1 - \frac{1}{n^2}\right)^{i^*}$$

Pozostało teraz tylko dowieść $\left(1 - \frac{1}{n^2}\right)^{i^*} \geq 1 - \frac{1}{n}$ dla odpowiednio dużego n , z czego będziemy już mieli $P(\mathcal{F}_{i^*}) \geq 1 - \frac{1}{n}$, a więc z wysokim prawdopodobieństwem istnieje urna z przynajmniej i^* kulami, co kończy dowód. Przejdźmy więc do tej ostatniej części

$$\left(1 - \frac{1}{n^2}\right)^{\frac{\ln \ln n}{\ln 2} - c} \geq \left(1 - \frac{1}{n^2}\right)^{\frac{\ln \ln n}{\ln 2}} = \left(1 - \frac{1}{n^2}\right)^{-n^2 \cdot \frac{-\ln \ln n}{n^2 \ln 2}}$$

$\left(1 - \frac{1}{n^2}\right)^{-n^2}$ dąży do e z góry, zatem:

$$e^2 > (4/3)^4 \geq \left(1 - \frac{1}{n^2}\right)^{-n^2} \geq e$$

$$e^{\frac{-2 \ln \ln n}{n^2 \ln 2}} \leq \left(1 - \frac{1}{n^2}\right)^{-n^2 \cdot \frac{-\ln \ln n}{n^2 \ln 2}} \leq e^{\frac{-\ln \ln n}{n^2 \ln 2}}$$

Teraz skorzystamy z rozwinięcia e^x w szereg Taylora

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

możemy zauważyć, że dla $x \in (-\varepsilon, 0)$

$$1 + x + x^2 \geq e^x \geq 1 + x - x^2$$

zatem:

$$1 - \frac{1}{n} \leq 1 - \frac{2 \ln \ln n}{n^2 \ln 2} - \left(\frac{2 \ln \ln n}{n^2 \ln 2}\right)^2 \leq \left(1 - \frac{1}{n^2}\right)^{-n^2 \cdot \frac{-\ln \ln n}{n^2 \ln 2}}$$

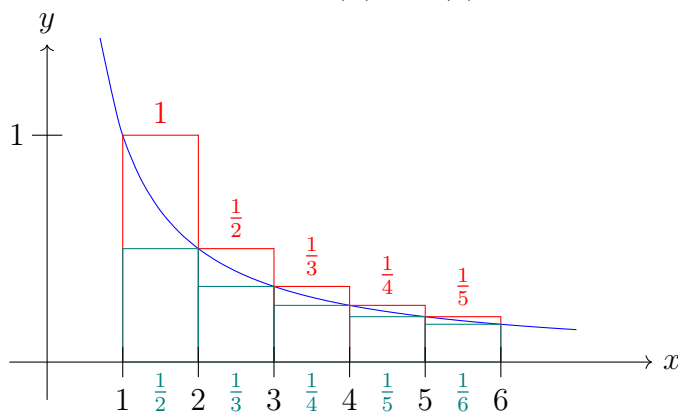
gdzie pierwsza nierówność zachodzi dla wystarczająco dużego n a druga wynika z dolnego ograniczenia na e^x .

Dodatkowe materiały

W tym rozdziale znajdują się twierdzenia i lematy, które nie są bezpośrednio częścią żadnego pytania, ale jakieś dowody do nich linkują.

28.1 Aproksymacja szeregu harmonicznego

Lemat 28.1.1. $H_n = \ln(n) + \Theta(1)$



$$\left. \begin{aligned} \sum_{i=1}^{n-1} \frac{1}{i} = H_{n-1} &\geq \int_1^n \frac{1}{x} dx = \ln(n) \\ \sum_{i=2}^n \frac{1}{i} = H_n - 1 &\leq \int_1^n \frac{1}{x} dx = \ln(n) \end{aligned} \right\} \implies \ln(n) \leq H_n \leq \ln(n) + 1$$

28.2 Wartość oczekiwana iloczynu

Twierdzenie 28.2.1. Dla dowolnych niezależnych zmiennych losowych X i Y :

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Dowód.

$$\begin{aligned}
 \mathbb{E}[X \cdot Y] &= \sum_i \sum_j (ij) \cdot \mathbb{P}(X = i \cap Y = j) \\
 &= \sum_i \sum_j ij \cdot \mathbb{P}(X = i) \cdot \mathbb{P}(Y = j) \\
 &= \sum_i i \cdot \mathbb{P}(X = i) \cdot \sum_j j \cdot \mathbb{P}(Y = j) \\
 &= \mathbb{E}[X] \cdot \mathbb{E}[Y]
 \end{aligned}$$

□

28.3 Wartość oczekiwana naturalnej zmiennej losowej

Twierdzenie 28.3.1 (Lemat 2.9 P&C). Niech X będzie zmienną losową przyjmującą jedynie wartości w liczbach naturalnych. Wtedy

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

Dowód.

$$\begin{aligned}
 \sum_{n=1}^{\infty} \mathbb{P}(X \geq n) &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathbb{P}(X = k) \\
 &= \sum_{k=1}^{\infty} \sum_{n=1}^k \mathbb{P}(X = k) \\
 &= \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) \\
 &= \mathbb{E}[X]
 \end{aligned}$$

□

28.4 Równość funkcji tworzących

Twierdzenie 28.4.1. (Bez dowodu)

Niech X, Y - zmienne losowe.

Jeżeli zachodzi:

$$\forall_{t \in (-\delta, \delta)} M_X(t) = M_Y(t)$$

gdzie $\delta > 0$ oraz $M_X(t)$ i $M_Y(t)$ istnieją w przedziale $(-\delta, \delta)$ to X i Y mają ten sam rozkład.

28.5 Ograniczenie górne Czernowa prób Poissona

Twierdzenie 28.5.1. Niech $X_1, \dots, X_n$ to niezależne próby Poissona. Dodatkowo oznaczamy $X = \sum_{i=1}^n X_i$ i $\mu = \mathbb{E}[X]$. 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 4.1.2, 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. $\square$