



METODY PROBABILISTYCZNE INFORMATYKI - WYKŁADY 2025

„Mam nadzieję, że jesteście ostatnim rocznikiem, na którym całki idą nie tak”

POPEŁNIONE PRZEZ

ZAŁATANY PONTON

V

NAHTAMATU

Kraków
Anno Domini 2025

Spis treści

1	Wykład 1 (2025-10-03)	1
1.1	Przestrzeń probabilistyczna	1
1.2	Przykładowe algorytmy randomizowane	3
1.2.1	Weryfikacja równości wielomianów	3
1.2.2	Weryfikacja wyniku mnożenia macierzy	4
2	Nagranie 1 (2025-10-04)	5
2.1	Zmienna losowa	5
2.2	Wartość oczekiwana	7
2.2.1	Definicja	7
2.2.2	Liniowość wartości oczekiwanej	7
3	Wykład 2 (2025-10-10)	10
3.1	Złożenie wartości oczekiwanej	10
3.2	Rozkład dwumianowy	12
3.3	Warunkowa wartość oczekiwana	13
3.4	Proces gałęzkowy	15
3.5	Rozkład geometryczny	16
4	Nagranie 2 (2025-10-10)	18
4.1	Przykładowe problemy z wartości oczekiwanej	18
4.1.1	Problem kolekcjonera kuponów	18
4.1.2	Oczekiwany czas Quicksorta	19
4.2	Nierówność Markowa	21
4.2.1	Definicja	21
4.2.2	Rzuty monetą	21
4.3	Wariancja i momenty zmiennej losowej	22
5	Wykład 3 (2025-10-17)	26
5.1	Wstęp	26
5.2	Algebra i σ -algebra	28
5.3	Miara	29

5.4	Całka Lebesgue'a	31
5.4.1	Definicje	31
5.4.2	Przykłady	32
6	Wykład 4 (2025-10-24)	35
6.1	Zastosowania nierówności na zmiennych losowych	35
6.2	Nierówność Czebyszewa	37
6.2.1	Definicja	37
6.2.2	Rzuty monetą	37
6.2.3	Kolekcjoner kuponów	38
6.3	Funkcje tworzące momentów	41
6.4	Nierówność Chernoffa	44
6.4.1	Definicja	44
6.4.2	Rzuty monetą	44
7	Wykład 5 (2025-11-07)	46
7.1	Paradoks dnia urodzin	46
7.2	Kule i urny	48
7.3	Kule i urny a rozkład Poissona	50
7.4	Rozkład Poissona	52
7.5	Ograniczenia Chernoffa zmiennej o rozkładzie Poissona	55
7.6	Granica rozkładu dwumianowego	55
7.7	Aproksymacja Poissona	56
8	Wykład 6 (2025-11-14)	58
8.1	Kule i urny - lower bound	58
8.2	Kule i urny - kolekcjoner kuponów	60
8.3	Ograniczenia Chernoffa prób Poissona	63
9	Wykład 7 (2025-11-19)	65
9.1	Łańcuchy Markowa - definicje	65
9.2	Stany - definicje	67
9.3	Łańcuchy nieprzywiedlne	68
9.4	Stan powracający i czas pierwszego spotkania	70
9.5	Rozkład stacjonarny	73
10	Wykład 8 (2025-11-21)	74
10.1	Rozkład stacjonarny - kontynuacja	74
10.2	Norma całkowitego wahanja	76
10.3	Sprzęganie	79
11	Wykład 9 (2025-11-28)	83

11.1	Prawdopodobieństwo ciągłe - definicje	83
11.2	Rozkład jednostajny	86
12	Wykład 10 (2025-12-05)	89
12.1	Rozkład wykładniczy	89
12.1.1	Właściwości	89
12.1.2	Kantory	93
12.2	Kule i urny z feedbackiem	94
12.3	Rozkład Gamma	96

Licencja



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

Wykład 1 (2025-10-03)

1.1 Przestrzeń probabilistyczna

Definicja 1.1.1. Przestrzeń probabilistyczna to trójka $(\Omega, \mathcal{F}, P)$, gdzie

- Ω to zbiór zdarzeń elementarnych
- $\mathcal{F} \subseteq 2^\Omega$ to rodzina zdarzeń mierzalnych
W przypadku dyskretnym $\mathcal{F} = 2^\Omega$
W przypadku ogólnym, $\mathcal{F}$ musi być σ -algebrą (definicja σ -algebry: 5.2.2)
- $P : \mathcal{F} \rightarrow \mathbb{R}$ taka, że
 - $\forall_{E \in \mathcal{F}} 0 \leq P(E) \leq 1$
 - $P(\Omega) = 1$
 - $\forall \begin{matrix} \{E_i\}_{i \in \mathbb{N}} \\ \forall_{i \in \mathbb{N}} E_i \in \mathcal{F} \\ \forall_{i,j, i \neq j} E_i \cap E_j = \emptyset \end{matrix} P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$

Lemat 1.1.1. Dla dowolnych zdarzeń E_1, E_2 zachodzi

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

Lemat 1.1.2. Prawdopodobieństwo sumy / union bound

Dla dowolnej skończonej lub przeliczalnej rodziny zdarzeń $E_1, E_2, \dots$ zachodzi

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} P(E_i)$$

Lemat 1.1.3. Zasada włączeń i wyłączeń

Dla dowolnych zdarzeń $E_1, \dots, E_n$

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i) - \sum_{\{i,j\} \in \binom{[n]}{2}} P(E_i \cap E_j) + \sum_{\{i,j,k\} \in \binom{[n]}{3}} P(E_i \cap E_j \cap E_k) - \dots$$

Definicja 1.1.2. Zdarzenia $E_1, E_2 \in \mathcal{F}$ są **niezależne** gdy

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$$

Definicja 1.1.3. Zdarzenia $E_1, \dots, E_k \in \mathcal{F}$ są **niezależne** gdy

$$\forall_{\substack{I \subseteq [k] \\ I \neq \emptyset}} P\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} P(E_i)$$

Definicja 1.1.4. Prawdopodobieństwo warunkowe

Zdarzenie E pod warunkiem zdarzenia F

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, P(F) > 0$$

Jeżeli E i F są niezależne:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E) \cdot P(F)}{P(F)} = P(E)$$

1.2 Przykładowe algorytmy randomizowane

1.2.1 Weryfikacja równości wielomianów

Mamy stworzyć algorytm sprawdzający równość wielomianów F i G . Przykładowo

$$(x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \stackrel{?}{=} x^6 - 7x^3 + 25$$

Dla stopnia wielomianu d , złożoność to $O(d^2)$ (wymnażamy wszystko i sprawdzamy współczynniki). Chcemy sprawdzić poprawność algorytmu w lepszym czasie.

Nasz algorytm:

- Wybierz r losowo (każdy z równym prawdopodobieństwem - *rozkład jednostajny / uniform distribution*) ze zbioru $\{1, \dots, 100d\}$
- Oblicz $F(r)$ i $G(r)$
- Jeśli $F(r) \neq G(r)$, to zwróć **false**
- Jeśli $F(r) = G(r)$, to zwróć **true**

Jeżeli algorytm zwróci **false**, to na pewno F i G są różne. Jeżeli zwróci **true**, to pomyli się tylko wtedy, gdy r jest pierwiastkiem $F(x) - G(x)$, a tych jest co najwyżej d . Możemy się więc spodziewać, że prawdopodobieństwo, że algorytm się pomyli jest równe

$$p = \frac{d}{100d} = \frac{1}{100}$$

Dowód. $\Omega = \{1, \dots, 100d\}$, E - zdarzenie takie, że algorytm się pomylił

$$\begin{aligned} P(E) &= P(\text{Algorytm wylosował pierwiastek } F(x) - G(x) \in \{1, \dots, 100d\}) \\ &\leq \frac{\#\text{pierwiastków } F(x) - G(x)}{100d} \\ &\leq \frac{d}{100d} = \frac{1}{100} \end{aligned}$$

□

Jeżeli puścimy ten algorytm kilka razy:

E_i - zdarzenie, że algorytm się pomyli za i -tym razem (te zdarzenia są niezależne)

$$P(E_1 \cap \dots \cap E_k) = \prod_{i \in [k]} P(E_i) \leq \left(\frac{1}{100}\right)^k$$

1.2.2 Weryfikacja wyniku mnożenia macierzy

Mamy trzy macierze A, B, C o rozmiarze $n \times n$ nad zbiorem $\{0, 1\}$ (arytmetyka mod 2). Chcemy napisać algorytm sprawdzający

$$AB \stackrel{?}{=} C$$

Normalne mnożenie działa w $O(n^3)$

Nasz algorytm:

- $\bar{r} = (r_1, \dots, r_n) \in \{0, 1\}^n$, wybrane losowo jednostajnie
- oblicz $A(B\bar{r})$ i $C\bar{r}$ (w $O(n^2)$)
- jeżeli wyniki są różne, to zwróć **false**
- jeżeli są równe, to zwróć **true**

Jeżeli $AB \neq C$ i $\bar{r}$ jest wybrane losowo i jednostajnie, to

$$P(A(B\bar{r}) = C\bar{r}) \leq \frac{1}{2}$$

Dowód. Niech $D = AB - C \neq 0$. Zakładamy bez straty ogólności $d_{1,1} \neq 0$

$$D\bar{r} = 0 \implies \sum_{j=1}^n d_{1,j}r_j = 0 \implies r_1 = \frac{-\sum_{j=2}^n d_{1,j}r_j}{d_{1,1}}$$

Widzimy, że r_1 musi mieć ustaloną wartość w oparciu na $r_2, \dots$, a więc w przynajmniej połowie przypadków r_1 nie ma odpowiedniej wartości i $A(B\bar{r}) \neq C\bar{r}$. $\square$

Nagranie 1 (2025-10-04)

2.1 Zmienna losowa

Definicja 2.1.1. Zmienna losowa przestrzeni $(\Omega, \mathcal{F}, P)$ to $X : \Omega \rightarrow \mathbb{R}$ taka, że

$$\forall F \in \mathcal{F} X(F) \in B(\mathbb{R})^1$$

Definicja 2.1.2. Dyskretna zmienna losowa

Dyskretna zmienna losowa to zmienna losowa, która przyjmuje w swoim obrazie przeliczalnie wiele wartości.

Przykład 2.1.1. Rzut dwa razy symetryczną kostką

Rozważmy równie prawdopodobne zdarzenia elementarne $\Omega = \{(1, 1), \dots, (6, 6)\}$ oraz zmienną losową $X((i, j)) := i + j$ (suma oczek w dwóch rzutach).

Naturalnie możemy chcieć policzyć prawdopodobieństwo, że dostaniemy sumę oczek równą a

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

Na przykład $P(X = 4) = P(\{1, 3\}) + P(\{2, 2\}) + P(\{3, 1\}) = \frac{3}{36}$

Definicja 2.1.3. Zmienne losowe X, Y są **niezależne** gdy

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

Definicja 2.1.4. Zmienne losowe $X_1, X_2, \dots, X_k$ są **niezależne** gdy

$$\forall I \subseteq [k] \forall_{\substack{\{x_i\}_{i \in I} \\ x_i \in \mathbb{R}}} P\left(\bigcap_{i \in I} (X_i = x_i)\right) = \prod_{i \in I} P(X_i = x_i)$$

Definicja 2.1.5. **Indykator** zdarzenia A to zmienna losowa Y spełniająca

$$Y = \begin{cases} 1 & \text{gdy } A \text{ zaszło} \\ 0 & \text{wpp.} \end{cases}$$

¹Zbiór borelowski: 5.2.2.

Mamy

$$\mathbb{E}[Y] = 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) = P(Y = 1) = P(A)$$

2.2 Wartość oczekiwana

2.2.1 Definicja

Definicja 2.2.1. Wartość oczekiwaną zmiennej losowej X definiujemy jako

$$\mathbb{E}[X] = \sum_{x \in \text{im } X} x \cdot P(X = x)$$

2.2.2 Liniowość wartości oczekiwanej

Twierdzenie 2.2.1. Liniowość wartości oczekiwanej (wariant z sumą)

Dla dowolnych $X_1, \dots, X_n$ z ograniczonymi wartościami oczekiwanymi

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

Dowód. ($n = 2$)

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

□

Twierdzenie 2.2.2. Liniowość wartości oczekiwanej (wariant ze współczynnikiem)

Dla danej zmiennej losowej X oraz $c \in \mathbb{R}$

$$\mathbb{E}[cX] = c \cdot \mathbb{E}[X].$$

Dowód. Dla $c = 0$ oczywiste. Zakładamy, że $c \neq 0$.

$$\begin{aligned} \mathbb{E}[cX] &= \sum_j j \cdot P(cX = j) \\ &= c \cdot \sum_j \frac{j}{c} \cdot P(X = \frac{j}{c}) \\ &= c \cdot \sum_k k \cdot P(X = k) \\ &= c \cdot \mathbb{E}[X]. \end{aligned}$$

□

Przykład 2.2.1. Niech X to zmienna losowa reprezentująca sumę oczek na dwóch kostkach. Ile wynosi $\mathbb{E}[X]$? Z definicji trudno jest to policzyć:

$$\mathbb{E}[X] = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \dots + 12 \cdot \frac{1}{36} = 7.$$

Z tw. o liniowości wynik dostajemy natychmiast. Niech X_1 i X_2 to odpowiednio wartość rzutu pierwszej i drugiej kostki. Wtedy $X = X_1 + X_2$.

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 3.5 + 3.5 = 7.$$

Przykład 2.2.2. Urna ma w sobie 4 kulki, każda z których jest innego koloru. 4 razy wyjmujemy kulę, obserwujemy kolor i wrzucamy kulę z powrotem. Jaka jest oczekiwana liczba zaobserwowanych kolorów?

Niech X to liczba zaobserwowanych kolorów oraz $X_i =: \begin{cases} 1, & \text{jeśli } i\text{-ty kolor zaobserwowany} \\ 0 & \end{cases}$

Wtedy

$$X = X_1 + X_2 + X_3 + X_4$$

X_i to indykatory, a więc $\mathbb{E}[X_i] = P(X_i)$. Zastanówmy się więc jakie jest prawdopodobieństwo $P(X_i = 1)$. Aby to zrobić, łatwiej będzie policzyć prawdopodobieństwo, że kolor i -ty *nie* wystąpi, a więc 4 razy dostaniemy jeden z pozostałych 3 kolorów: $P(X_i = 1) = 1 - \left(\frac{3}{4}\right)^4$. Dostajemy zatem

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X_1 + X_2 + X_3 + X_4] = \sum_{i=1}^4 \mathbb{E}[X_i] \\ &= \sum_{i=1}^4 P(X_i = 1) \\ &= 4 \cdot \left(1 - \left(\frac{3}{4}\right)^4\right). \end{aligned}$$

Przykład 2.2.3. Mrówki chodzące po kiju

Zacznijmy od zagadki:

Rozważamy kij o długości 1m. Losowo kładziemy na niego 25 mrówek i losowo ustalamy zwrot w lewo lub w prawo. Każda mrówka porusza się z szybkością 1cm/s. Gdy mrówki się spotkają, odbijają się od siebie, a więc zmienia się ich zwrot. Ilu sekund potrzebujemy, aby mieć pewność, że każda mrówka spadnie z kija.

Można zauważyć, że mrówki są nierozróżnialne, a więc gdy dwie mrówki się spotykają, możemy przyjąć, iż się nie odbijają, a po prostu przenikają się wzajemnie i idą dalej. Zatem odpowiedź

to 100s.

W modelu takim jak w zagadce zastanawiamy się, jaka jest wartość oczekiwana wszystkich odbić. Problematyczne w zadaniu może okazać się, że jest continuum możliwych pozycji dla każdej mrówki. W rozwiązaniu jednak nie będzie nam potrzebna ta informacja.

Niech X to liczba odbić, lub równoważnie liczba przeniknięć. Ponadto niech $X_{i,j}$ to liczba przeniknięć i -tej i j -tej mrówki $X_{i,j} \in \{0, 1\}$. Zakładamy, że mrówki są posortowane i dla $i < j$ mrówka i -ta jest przed j -tą. Zauważmy, że

$$\mathbb{E}[X_{i,j}] = P(X_{i,j} = 1) = \frac{1}{4}$$

a więc

$$E(X) = \mathbb{E}\left[\sum_{1 \leq i < j \leq 25} X_{i,j}\right] = \sum_{1 \leq i < j \leq 25} \mathbb{E}[X_{i,j}] = \sum_{1 \leq i < j \leq 25} P(X_{i,j} = 1) = \sum_{1 \leq i < j \leq 25} \frac{1}{4} = \binom{25}{2} \frac{1}{4}.$$

Wykład 2 (2025-10-10)

3.1 Złożenie wartości oczekiwanej

Twierdzenie 3.1.1. Dla dowolnej zmiennej losowej X zachodzi

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$$

Dowód.

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

□

Przykład 3.1.1. Niech X to zmienna losowa, $P(X = n) = \frac{1}{99}, n \in \{1, \dots, 99\}$

$$\mathbb{E}[X]^2 = \left(\sum_{n=1}^{99} \frac{1}{99} \cdot n \right)^2 = \left(\frac{1}{99} \cdot \frac{99 \cdot 100}{2} \right)^2 = 50^2 = 2500$$

$$\mathbb{E}[X^2] = \sum_{n=1}^{99} \frac{1}{99} \cdot n^2 = \frac{1}{99} \cdot \frac{99 \cdot 100 \cdot 199}{6} = \frac{9950}{3} > 2500$$

Fakt 3.1.1.

$$\mathbb{E}[f(x)] = \sum_{z \in \text{im } f(X)} P(f(X) = z) \cdot z = \sum_{z \in \text{im } X} P(X = z) \cdot f(z)$$

Definicja 3.1.1. Funkcja $f : \mathbb{R} \rightarrow \mathbb{R}$ jest **wypukła**, jeśli:

$$\forall_{x_1, x_2 \in \mathbb{R}, \lambda \in [0, 1]} f(\lambda \cdot x_1 + (1 - \lambda) \cdot x_2) \leq \lambda \cdot f(x_1) + (1 - \lambda) \cdot f(x_2)$$

Twierdzenie 3.1.2. Nierówność Jensena

Jeśli funkcja f jest wypukła i $\mathbb{E}[X] < \infty$ to

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

Dowód. Zakładamy, że f ma rozwinięcie Taylora. Niech $\mu = \mathbb{E}[X]$. Z Taylora wiemy, że

$$\begin{aligned}\exists_{c \in \mathbb{R}} f(x) &= f(\mu) + f'(\mu) \cdot (x - \mu) + \frac{f''(c) \cdot (x - \mu)^2}{2} \\ &\geq f(\mu) + f'(\mu) \cdot (x - \mu)\end{aligned}$$

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

□

3.2 Rozkład dwumianowy

Definicja 3.2.1. 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 3.2.1. Rozkład dwumianowy jest poprawnie zdefiniowany

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

Twierdzenie 3.2.1. Niech X ma rozkład dwumianowy z parametrami n, p . Wtedy

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

Dowód.

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

Można to też zrobić prościej:

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

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

□

3.3 Warunkowa wartość oczekiwana

Definicja 3.3.1. Warunkową wartość oczekiwaną $\mathbb{E}[X | Y = y]$ definiujemy jako

$$\mathbb{E}[X | Y = y] = \sum_{x \in \text{im } X} x P(X = x | Y = y)$$

Ponadto $\mathbb{E}[X | Y]$ definiujemy jako zmienną losową taką, że

$$\mathbb{E}[X | Y](y) = \mathbb{E}[X | Y = y]$$

Przykład 3.3.1. X_1, X_2 - wyniki niezależnych rzutów kostką 6-ścienną, $X = X_1 + X_2$

$$\begin{aligned} \mathbb{E}[X | X_1 = 2] &= \sum_{x \in [12]} x \cdot P(X = x | X_1 = 2) \\ &= \sum_{x=3}^8 x \cdot \frac{1}{6} \\ &= 5,5 \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X_1 | X = 5] &= \sum_{x=1}^4 x \cdot P(X_1 = x | X = 5) \\ &= \sum_{x=1}^4 x \cdot \frac{P(X_1 = x \cap X = 5)}{P(X = 5)} \\ &= \sum_{x=1}^4 x \cdot \frac{\frac{1}{36}}{\frac{4}{36}} \\ &= 2,5 \end{aligned}$$

Przykład 3.3.2. X_1, X_2 - wyniki niezależnych rzutów kostką 6-ścienną, $X = X_1 + X_2$

$$\begin{aligned} \mathbb{E}[X | X_1](\omega) &= \mathbb{E}[X | X_1 = X_1(\omega)] \\ &= \sum_{y=X_1(\omega)+1}^{X_1(\omega)+6} y P(X = y | X_1 = X_1(\omega)) \\ &= \sum_{y=X_1(\omega)+1}^{X_1(\omega)+6} y \frac{1}{6} \\ &= X_1(\omega) + 3,5 \end{aligned}$$

A więc $\mathbb{E}[X | X_1] = X_1 + 3,5$

Lemat 3.3.1.

$$\mathbb{E}[X] = \sum_{y \in \text{im } Y} \mathbb{E}[X \mid Y = y] \cdot P(Y = y)$$

Dowód.

$$\begin{aligned} \sum_{y \in \text{im } Y} \mathbb{E}[X \mid Y = y] \cdot P(Y = y) &= \sum_{y \in \text{im } Y} \left(P(Y = y) \sum_{x \in \text{im } X} x \cdot P(X = x \mid Y = y) \right) \\ &= \sum_{x \in \text{im } X} \left(x \cdot \sum_{y \in \text{im } Y} P(Y = y) \cdot P(X = x \mid Y = y) \right) \\ &= \sum_{x \in \text{im } X} \left(x \cdot \sum_{y \in \text{im } Y} P(Y = y) \cdot \frac{P(X = x \wedge Y = y)}{P(Y = y)} \right) \\ &= \sum_{x \in \text{im } X} x \cdot P(X = x) \\ &= \mathbb{E}[X] \end{aligned}$$

□

Fakt.

$$\mathbb{E} \left[\sum_i X_i \mid Y = y \right] = \sum_i \mathbb{E}[X_i \mid Y = y]$$

Czyli liniowość zachodzi też dla warunkowej wartości oczekiwanej

Lemat 3.3.2 (Lemat Syntaktyczny).

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

Dowód.

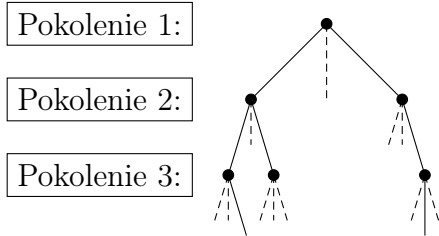
$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \sum_{y \in \text{im } Y} \mathbb{E}[X \mid Y = y] \cdot P(Y = y) = \mathbb{E}[X]$$

□

3.4 Proces gałązkowy

Zaczynamy od programu A, który n razy próbuje stworzyć kopię siebie z prawdopodobieństwem p . Następnie, każda utworzona kopia powtarza ten sam proces.

Przykład dla $n = 3$



Jaka jest oczekiwana liczba wszystkich odpalonych programów?

Y_i - liczba programów w i -tym pokoleniu

$Y_0 = 0$, Y_1 ma rozkład dwumianowy z parametrami $(n, p) \implies \mathbb{E}[Y_1] = np$

$$\begin{aligned}
 \mathbb{E}[Y_i | Y_{i-1} = y_{i-1}] &= \mathbb{E}\left[\sum_{j=1}^{y_{i-1}} Z_j | Y_{i-1} = y_{i-1}\right] \\
 &= \sum_{\ell=0}^{\infty} \ell P\left(\sum_{j=1}^{y_{i-1}} Z_j = \ell | Y_{i-1} = y_{i-1}\right) \\
 &= \sum_{\ell=0}^{\infty} \ell P\left(\sum_{j=1}^{y_{i-1}} Z_j = \ell\right) \\
 &= \mathbb{E}\left[\sum_{j=1}^{y_{i-1}} Z_j\right] = \sum_{j=1}^{y_{i-1}} \mathbb{E}[Z_j] = \sum_{j=1}^{y_{i-1}} np = y_{i-1} np \\
 \mathbb{E}[Y_i | Y_{i-1}] &= Y_{i-1} np
 \end{aligned}$$

Teraz, dowodzimy indukcyjnie, że $\mathbb{E}[Y_i] = (np)^i$

$$\mathbb{E}[Y_i] = \mathbb{E}[\mathbb{E}[Y_i | Y_{i-1}]] = \mathbb{E}[Y_{i-1} np] = np \mathbb{E}[Y_{i-1}] = np (np)^{i-1} = (np)^i$$

A następnie obliczamy wynik:

$$\mathbb{E}\left[\sum_{i=0}^{\infty} Y_i\right] \stackrel{1}{=} \sum_{i=0}^{\infty} \mathbb{E}[Y_i] = \sum_{i=0}^{\infty} (np)^i$$

Widzimy, że dla $np < 1$ wartość oczekiwana wynosi $\frac{1}{1-np}$, a w przeciwnym przypadku nie ma wartości oczekiwanej, ponieważ liczba programów rośnie do nieskończoności.

¹Według Micka można to zrobić, o ile prawa strona równości jest zbieżna

3.5 Rozkład geometryczny

Definicja 3.5.1. 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 3.5.1 (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. Zauważmy najpierw, że dla dowolnego $0 < p < 1$ mamy

$$\sum_{i=k}^{\infty} (1-p)^i = (1-p)^k \cdot \frac{1}{1-(1-p)} = \frac{(1-p)^k}{p}$$

W takim razie

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

□

Twierdzenie 3.5.2 (Lemat 2.9 P&C). Niech X będzie zmienną losową przyjmującą jedynie wartości w liczbach naturalnych. Wtedy

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

Dowód.

$$\begin{aligned}
 \sum_{n=1}^{\infty} P(X \geq n) &= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(X = k) \\
 &= \sum_{k=1}^{\infty} \sum_{n=1}^k P(X = k) \\
 &= \sum_{k=1}^{\infty} k \cdot P(X = k) \\
 &= \mathbb{E}[X]
 \end{aligned}$$

□

Twierdzenie 3.5.3. Niech X ma rozkład geometryczny z parametrem p . Wtedy

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

Dowód. Skorzystamy z twierdzenia 3.5.2

$$\begin{aligned}
 \mathbb{E}[X] &= \sum_{n=1}^{\infty} P(X \geq n) \\
 &= \sum_{n=1}^{\infty} \sum_{i=n}^{\infty} (1-p)^{i-1} \cdot p \\
 &= \sum_{n=1}^{\infty} (1-p)^n \\
 &= \frac{1}{1 - (1-p)} \\
 &= \frac{1}{p}
 \end{aligned}$$

Alternatywna metoda:

$$Y = \begin{cases} 1 & \text{sukces w pierwszym rzucie} \\ 0 & \text{wpp.} \end{cases}$$

$$\begin{aligned}
 \mathbb{E}[X] &= \mathbb{E}[X|Y=0] P(Y=0) + \mathbb{E}[X|Y=1] P(Y=1) \\
 &= (1 + \mathbb{E}[X])(1-p) + 1 \cdot p \\
 &= 1 + \mathbb{E}[X] - p\mathbb{E}[X]
 \end{aligned}$$

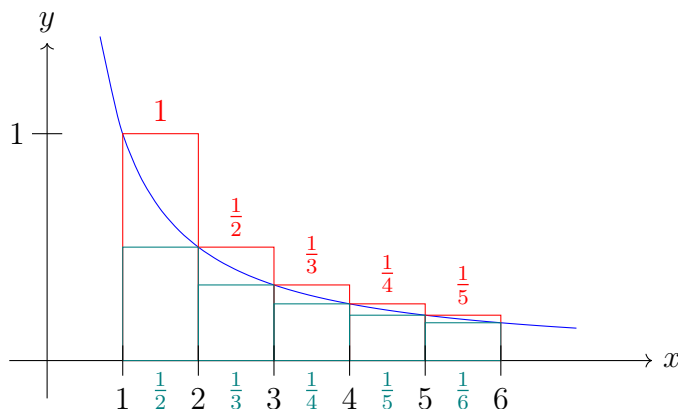
$$p\mathbb{E}[X] = 1 \implies \mathbb{E}[X] = \frac{1}{p}$$

□

Nagranie 2 (2025-10-10)

4.1 Przykładowe problemy z wartości oczekiwanej

Lemat 4.1.1. $H_n = \ln(n) + \Theta(1)$



$$\left. \begin{array}{l} \sum_{i=1}^{n-1} \frac{1}{i} = H_{n-1} \geq \int_1^n \frac{1}{x} dx = \ln(n) \\ \sum_{i=2}^n \frac{1}{i} = H_n - 1 \leq \int_1^n \frac{1}{x} dx = \ln(n) \end{array} \right\} \implies \ln(n) \leq H_n \leq \ln(n) + 1$$

4.1.1 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 $\mathbb{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$

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

4.1.2 Oczekiwany czas 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 4.1.1 (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 faktu, że $\sum_{k=1}^n \frac{1}{k} = H_n = \ln n + \Theta(1)$ i dostajemy

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

□

4.2 Nierówność Markowa

4.2.1 Definicja

Twierdzenie 4.2.1. Jeśli X jest zmienną losową, która przyjmuje nieujemne wartości to

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

Dowód. Niech I będzie indykatorem

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

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

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

□

4.2.2 Rzuty monetą

Rzucamy n razy monetą - zliczamy liczbę orłów.

Jak *bardzo* możemy ograniczyć prawdopodobieństwo otrzymania *dużej* liczby orłów? (Jest to trochę nieformalne, nie wiemy co dokładnie znaczy *bardzo* ani *duża liczba orłów*, ale jakieś przykładowe ograniczenie możemy podać)

Niech $X_i = \begin{cases} 1, & \text{orzeł za } i\text{-tym razem} \\ 0 & \text{wpp} \end{cases}$

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_i X_i\right] = \sum_i \mathbb{E}[X_i] = \frac{n}{2}$$

Chcemy ograniczyć prawdopodobieństwo tego, że w więcej niż $\frac{3}{4}$ rzutów otrzymamy orła.

Z Markowa:

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

Intuicyjnie ta wartość powinna być dużo mniejsza (i coraz mniejsza dla większych n) ale Markow daje formalne ograniczenie (choć jak widać dość słabe).

4.3 Wariancja i momenty zmiennej losowej

Definicja 4.3.1. n -tym momentem ($n \geq 1$) zmiennej losowej X nazywamy

$$\mathbb{E}[X^n]$$

W szczególności, wartość oczekiwana jest pierwszym momentem zmiennej losowej.

Definicja 4.3.2. 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 4.3.3. Odchylenie standardowe zmiennej losowej X definiujemy jako

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

Przykład 4.3.1. Niech X - stała, $X = c$ (z prawdopodobieństwem 1 przyjmuje wartość c).

Wartość oczekiwana X :

$$\mathbb{E}[X] = c \cdot P(X = c) = c \cdot 1 = c$$

Wariancja:

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[(c - c)^2] = 0$$

Wariancja wynosi 0 i rzeczywiście, dla stałej nie spodziewamy się żadnego odchylenia od wartości oczekiwanej.

Niech $Y = \begin{cases} k \cdot c, & \text{z prawdopodobieństwem } \frac{1}{k} \\ 0, & \text{z prawdopodobieństwem } \frac{k-1}{k} \end{cases}$

Wartość oczekiwana Y :

$$\mathbb{E}[Y] = 0 \cdot \frac{k-1}{k} + k \cdot c \cdot \frac{1}{k} = c$$

Wariancja Y :

$$\text{Var}[Y] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = (kc)^2 \cdot \frac{1}{k} - c^2 = kc^2 - c^2 = c^2 \cdot (k - 1)$$

Więc dwie zmienne mające tą samą wartość oczekiwaną mogą mieć różne wariancje.

Definicja 4.3.4. Kowariancję zmiennych losowych X oraz Y definiujemy jako

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

Twierdzenie 4.3.1. 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 4.3.2. Dla dowolnych niezależnych zmiennych losowych X i Y :

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Dowód.

$$\begin{aligned} \mathbb{E}[X \cdot Y] &= \sum_i \sum_j (ij) \cdot \text{P}(X = i \cap Y = j) \\ &= \sum_i \sum_j ij \cdot \text{P}(X = i) \cdot \text{P}(Y = j) \\ &= \sum_i i \cdot \text{P}(X = i) \cdot \sum_j j \cdot \text{P}(Y = j) \\ &= \mathbb{E}[X] \cdot \mathbb{E}[Y] \end{aligned}$$

□

Twierdzenie 4.3.3. 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.

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

□

Twierdzenie 4.3.4. Niech X ma rozkład dwumianowy z parametrami n, p . Wtedy

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

Dowód.

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

Pozostaje nam tylko policzyć $\mathbb{E}[X^2]$

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{j=0}^n \binom{n}{j} p^j (1-p)^{n-j} j^2 \\ &= \sum_{j=0}^n \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j} (j^2 - j + j) \\ &= \sum_{j=0}^n \frac{n!(j^2 - j)}{j!(n-j)!} p^j (1-p)^{n-j} + \sum_{j=0}^n \frac{n!j}{j!(n-j)!} p^j (1-p)^{n-j} \\ &= n(n-1)p^2 \sum_{j=0}^n \frac{(n-2)!}{(j-2)!(n-j)!} p^{j-2} (1-p)^{n-j} \\ &\quad + np \sum_{j=1}^n \frac{(n-1)!}{(j-1)!(n-j)!} p^{j-1} (1-p)^{n-j} \\ &= n(n-1)p^2 + np \end{aligned}$$

W takim razie

$$\text{Var}[X] = n(n-1)p^2 + np - (np)^2 = np - np^2 = np(1-p)$$

Ponieważ kolejne próby są niezależne, można obliczyć to prościej:

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

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

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

¹Możemy to zrobić przez 4.3.2

Sumując po i dostajemy:

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

□

Twierdzenie 4.3.5. 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

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

□

Wykład 3 (2025-10-17)

5.1 Wstęp

Na początek skupmy się na prostym problemie: mamy Ω - uniwersum, oraz $A \subseteq \Omega$. Losujemy jednostajnie element $x \in \Omega$ i chcemy poznać $P(x \in A)$.

Naturalnie nasuwa się definicja

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

Jak możemy jednak zdefiniować wielkość? Dla $|\Omega| < \infty$ jest to proste: $\text{wielkość}(X) = |X|$, lecz dla Ω nieskończonego nie ma już oczywistej odpowiedzi. Na ratunek przychodzi pojęcie miary.

Przykład 5.1.1. Przykładowe miary

- Długość odcinka
- Pole powierzchni na płaszczyźnie
- Objętość w przestrzeni trójwymiarowej

Okazuje się jednak, że nie wszystkie zbiory można mierzyć.

Przykład 5.1.2. Paradoks Banacha-Tarskiego

Niech S to będzie sfera w przestrzeni trójwymiarowej. Możemy podzielić S na rozłączne części w następujący sposób

$$S = P_1 \cup P_2 \cup P_3 \cup P_4 \cup P_5$$

a następnie za pomocą ustalonych przesunięć $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5$ dojść do

$$S = \tau_1(P_1) \cup \tau_2(P_2)$$

$$S = \tau_3(P_3) \cup \tau_4(P_4) \cup \tau_5(P_5)$$

Gdybyśmy chcieli mierzyć objętość naszej sfery i pozwalalibyśmy na podzielenie jej na dowolne części i wykonanie przesunięć doszlibyśmy więc do oczywiście sprzecznego wniosku

$$\text{Objętość}(S) = 2 \cdot \text{Objętość}(S)$$

Widzimy, że nie we wszystkich przypadkach intuicyjne pojęcie miary jest poprawne i nie wszystkie zbiory są mierzalne. W następnej sekcji skupimy się na znalezieniu zbiorów, które możemy mierzyć i dokładnie zdefiniujemy miarę.

5.2 Algebra i σ -algebra

Definicja 5.2.1. Algebra to takie $\Sigma \subseteq 2^\Omega$ dla danego uniwersum Ω , że spełnione są następujące warunki

- $\emptyset, \Omega \in \Sigma$
- $e \in \Sigma \implies \Omega \setminus e \in \Sigma$
- $E_1, \dots, E_n \in \Sigma \implies \bigcup_{i=1}^n E_i \in \Sigma$

Definicja 5.2.2. σ -algebra to algebra spełniająca dodatkowy warunek

$$E_1, \dots \in \Sigma \implies \bigcup_{i=1}^{\infty} E_i \in \Sigma$$

Przykład 5.2.1. Proste przykłady algebr i σ -algebr

- $\Sigma = 2^\Omega$ - σ -algebra
- $\Sigma = \{\emptyset, \Omega\}$ - σ -algebra
- $\Omega = \mathbb{R}, \Sigma = \{\bigcup_{i=1}^n (a_i, b_i] \mid a_i, b_i \in \mathbb{R} \cup \{-\infty, +\infty\}, a_i < b_i\}$ - algebra, ale nie σ -algebra, bo np. $\bigcup_{i=1}^{\infty} (0, 1 - \frac{1}{n}] = (0, 1) \notin \Sigma$

Lemat 5.2.1. Niech Ω to uniwersum, a $(\Sigma_\alpha)_{\alpha \in I}$ to σ -algebry na Ω . Wtedy

$$\bigcap_{\alpha \in I} \Sigma_\alpha \text{ jest } \sigma\text{-algebrą}$$

Definicja 5.2.3. σ -algebra generowana przez $F \subseteq 2^\Omega$ to

$$\sigma(F) = \bigcap \{ \Sigma \subseteq 2^\Omega : F \subseteq \Sigma \text{ i } \Sigma \text{ jest } \sigma\text{-algebrą} \}$$

Intuicyjnie, jest to najmniejsza możliwa σ -algebra zawierająca w sobie F .

Przykład 5.2.2. Przykład σ -algebry generowanej

$$\sigma(\{[a, b] \mid a, b \in \mathbb{R}, a \leq b\}) =: B(\mathbb{R})$$

$$\sigma(d\text{-wymiarowe kostki}) =: B(\mathbb{R}^d)$$

B to oznaczenie na zbiory borelowskie, które są przykładem zbiorów mierzalnych, nie mając problemów takich jak Banach-Tarski.

Uwaga. Od teraz, gdy będziemy mówić o $\mathbb{R}, \mathbb{C}$ w kontekście miary, mamy na myśli przestrzenie mierzalne $(\mathbb{R}, B(\mathbb{R})), (\mathbb{C}, B(\mathbb{C}))$, o ile nie jest powiedziane inaczej.

5.3 Miara

Definicja 5.3.1. Przestrzeń mierzalna to para (Ω, Σ) dla uniwersum Ω i Σ będącego σ -algebrą na Ω .

Definicja 5.3.2. Miara dla danej przestrzeni mierzalnej (Ω, Σ) to funkcja $\mu : \Sigma \rightarrow [0, +\infty]$ spełniająca

- $\mu(\emptyset) = 0$
- $E_1, E_2, \dots \in \Sigma$ parami rozłączne $\implies \mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$

Definicja 5.3.3. Przestrzeń z miarą to trójka (Ω, Σ, μ) gdzie (Ω, Σ) tworzą przestrzeń mierzalną, a μ jest miarą na tej przestrzeni mierzalnej.

Definicja 5.3.4. Funkcja $f : \Omega_1 \rightarrow \Omega_2$ jest **mierzalna** dla przestrzeni probabilistycznych $(\Omega_1, \Sigma_1), (\Omega_2, \Sigma_2)$ gdy

$$\forall B \in \Sigma_2 f^{-1}(B) \in \Sigma_1$$

Definicja 5.3.5. Miara probabilistyczna to miara dla której $\mu(\Omega) = 1$

Lemat 5.3.1. Niech $A_1, A_2, \dots \in \Sigma, \forall_{i \in \mathbb{N}_+} A_i \subseteq A_{i+1}$. Wtedy

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{x \rightarrow \infty} \mu(A_i)$$

Lemat 5.3.2. Niech $A_1, A_2, \dots \in \Sigma, \forall_{i \in \mathbb{N}_+} A_i \supseteq A_{i+1}$ oraz $\mu(A_1) < \infty$. Wtedy

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{x \rightarrow \infty} \mu(A_i)$$

Przykład 5.3.1. Miara licząca

Miara licząca dla $\Omega < \infty$ to $\mu(A) = |A|$

Przykład 5.3.2. Miara Lebesgue'a

Miara Lebesgue'a to uogólnienie pojęcia długości, pola powierzchni i objętości na większą liczbę wymiarów. Przykładowo, w jednym wymiarze:

$$\mu([a, b]) = b - a$$

a więc

$$\mu(\{a\}) = \mu\left(\bigcap_{n=1}^{\infty} [a, a + \frac{1}{n}]\right) = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

$$\mu(\mathbb{Q}) = \mu(\{q_1, q_2, \dots\}) = \mu\left(\bigcup_{n=1}^{\infty} \{q_1, \dots, q_n\}\right) = \lim_{n \rightarrow \infty} \mu(\{q_1, \dots, q_n\}) = 0$$

5.4 Całka Lebesgue'a

5.4.1 Definicje

Definicja 5.4.1. Funkcja $S : \Omega \rightarrow [0, +\infty]$ jest **prosta** (schodkowa), jeśli:

$$S = \sum_{j=1}^n \alpha_j \chi_{A_j}$$

gdzie $\forall_{j \in [n]} A_j \in \Sigma, \alpha_j \in \mathbb{R}$ a χ_{A_j} jest funkcją charakterystyczną A_j

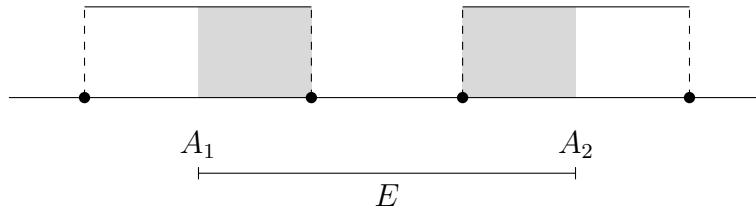
Od teraz niech $E \in \Sigma$

Definicja 5.4.2. Niech $S : \Omega \rightarrow [0, +\infty]$ - funkcja prosta.

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

Gdzie μ jest miarą Lebesgue'a.

Mozna sobie to wyobrazić jako suma pól prostokątów pod funkcją S :



Definicja 5.4.3. Niech $f : \Omega \rightarrow [0, +\infty]$ - funkcja mierzalna.

$$\int_E f \, d\mu = \sup \left\{ \int_E s \, d\mu : s : \Omega \rightarrow [0, \infty], s \text{ jest prosta}, s \leq f \right\}$$

Definicja 5.4.4. Niech $f : \Omega \rightarrow \mathbb{R}$ mierzalna

$$f_+ := \max\{0, f\}$$

$$f_- := -\min\{0, f\}$$

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

Definicja 5.4.5. Dla liczb zespolonych:

$$f : \Omega \rightarrow \mathbb{C}$$

Dzielimy f na część rzeczywistą i urojoną:

$$f = u + i \cdot v, \text{ gdzie } u(z) = \operatorname{Re}(f(z)), v(z) = \operatorname{Im}(f(z))$$

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

Co tutaj w ogóle definiujemy? Zaczynamy od definicji funkcji prostej i całki z funkcji prostej. Potem na jej podstawie definiujemy całkę z funkcji nieujemnej mierzalnej i ostatecznie dochodzimy do dowolnej funkcji mierzalnej.

Twierdzenie 5.4.1. (bez dowodu)

Dla:

$f : (\Omega, \Sigma) \rightarrow [0, +\infty]$ mierzalna

$\{S_n\}_{n=1}^{+\infty}$

$$S_n : \Omega \rightarrow [0, +\infty], S_1 \leq S_2 \leq \dots \leq f, \forall_{x \in \Omega} f(x) = \lim_{n \rightarrow \infty} S_n(x) \quad (*)$$

Zachodzi:

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

Twierdzenie 5.4.2.

$$f : (\Omega, \Sigma) \rightarrow [0, +\infty] \text{ mierzalna} \implies \exists \{S_n\}_{n=1}^{+\infty} \text{ takie, że } (*)$$

Wniosek 5.4.1. Jak to się łączy z prawdopodobieństwem?

Prawdopodobieństwo definiujemy jako miarę probabilistyczną na (Ω, Σ)

Zmienną losową definiujemy jako funkcję mierzalną $f : (\Omega, \Sigma) \rightarrow \mathbb{R}$

Wartość oczekiwana definiujemy:

$$\mathbb{E}[X] = \int_{\Omega} X \, dP$$

Wariancję definiujemy:

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

Dzięki temu otrzymujemy odpowiednie definicje poza przypadkiem dyskretnym.

5.4.2 Przykłady

Przykład 5.4.1. Niech $|\Omega| = n < +\infty$, $\Sigma = 2^{\Omega}$

Niech X - zmienna losowa przyjmująca wartości od 1 do n z równym prawdopodobieństwem.

$\mathbb{E}[X]$ możemy obliczyć z miary liczącej:

$$P(A) = \frac{|A|}{|\Omega|} \implies P(X = i) = \frac{1}{n}$$

$$\mathbb{E}[X] = \sum_{i=1}^n P(X = i) \cdot i$$

Alternatywnie, możemy zauważyć, że X jest funkcją prostą.

$$X(x) = \sum_{i=1}^n \chi_{\{i\}}(x) \frac{1}{n}$$

$$\mathbb{E}[X] = \int_{\Omega} X \, dP = \sum_{i=1}^n i \cdot \frac{1}{n}$$

Przykład 5.4.2. Losowanie z odcinka.

$$\Omega = [0, 1], \Sigma = B([0, 1])$$

Wtedy $P(A) = \lambda(A)$, gdzie λ jest miarą Lebesgue'a (długością).

Przykład 5.4.3. Losowanie zgodnie z funkcją

$$\Omega = \mathbb{R}, \Sigma = B(\mathbb{R}), A \in \Sigma$$

Niech $\int_{-\infty}^{+\infty} f(x) \, dx = 1$, oraz chcemy losować zgodnie z f .

$$\text{Wtedy } P(A) = \int_A f \, d\lambda$$

Analogicznie dla $\Omega = \mathbb{R}^2, \Sigma = B(\mathbb{R}^2)$ otrzymamy:

$$P(A) = \int_A f \, d\lambda^2$$

Przykład 5.4.4. Możemy także łączyć prawdopodobieństwo dyskretne z ciągłym. Rozważmy następujący problem:

Mamy odcinek $[0, 1]$. Przy punkcie 0 znajduje się magnes, który przyciąga wszystko z przedziału $[0, \frac{1}{3}]$. W punkcie 1 stoi osoba, która rzuca piłką i z jednostajnym prawdopodobieństwem trafia w dowolny punkt na naszym odcinku. Jakie jest prawdopodobieństwo, że trafi w dany przedział A ?

$$\Omega = [0, 1]$$

Dla $A \subseteq [0, 1]$ mamy: $P(\{0\}) = \frac{1}{3}$ - to, co przyciągnie magnes.

$$\text{Dla } A \subseteq [\frac{1}{3}, 1]: P(A) = \lambda(A)$$

Przykład 5.4.5. Nieskończony ciąg rzutów monetą symetryczną.

$$\Omega = \{(X_n)_{n=1}^{+\infty} : X_n \in \{0, 1\}\}$$

Pojawia się problem - jak zdefiniować Σ , aby był mierzalny?

Definiujemy **cylindry**:

$$a_1, \dots, a_k \in \{0, 1\}$$

$$[a_1, \dots, a_k] := \{(x_n)_{n=1}^{+\infty} \mid a_1 = x_1, \dots, a_k = x_k\}$$

$$\text{Cyl} := \{[a_1, \dots, a_k] \mid a_1, \dots, a_k \in \{0, 1\}, k \in \mathbb{N}\}$$

$$\Sigma = \sigma(\text{Cyl})$$

czyli Σ jest σ -algebrą generowaną przez zbiór cylindrów.

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

Wykład 4 (2025-10-24)

6.1 Zastosowania nierówności na zmiennych losowych

Pokażmy przykład zastosowania nierówności na zmiennych losowych poprzez wypełnienie tabelki pokazującej rezultaty jakie dają nam poszczególne nierówności.

Dla wprowadzenia, gdy mówimy o problemie kolekcjonera kuponów odwołujemy się do 4.1.1. Za to gdy mówimy o rzutach monetą, to X_i to zmienna losowa dla której

$$P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$$

oraz

$$X = \sum_{i=1}^n X_i, \mathbb{E}[X] = \frac{n}{2}$$

	Rzuty monetą Ograniczenie $P(X \geq \frac{3}{4}n)$	Kolekcjoner kuponów Ograniczenie $P(X \geq 2nH_n)$
Nierówność Markowa (4.2.1)	$\frac{2}{3}$ (4.2.2)	$\frac{1}{2}$ (6.1.2)
Nierówność Czebyszewa (6.2.1)	$\frac{4}{n}$ (6.2.2)	$O(\frac{1}{\ln^2 n})$ (6.2.3)
Nierówność Chernoffa (6.4.1)	$e^{-\frac{n}{20}}$ (6.4.2)	e^{-n}
Union bound (1.1.2)		$\frac{1}{n}$ (6.1.1)

W następnych sekcjach pokażemy odpowiednie definicje tych nierówności oraz dowody wartości podanych w tabelce.

Przykład 6.1.1. Union bound dla kolekcjonera kuponów

Niech A_i to zdarzenie polegające na niezyskaniu i -tego kuponu po $n \ln(n) + cn$ krokach

$$P(A_i) = \left(1 - \frac{1}{n}\right)^{n \ln(n) + cn} = \left(1 + \frac{1}{-n}\right)^{-n(-\ln(n) - c)} \leq e^{-(\ln(n) + c)} = \frac{1}{e^c n}$$

Z 1.1.2 wiemy, że

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) = \frac{1}{e^c}$$

Niech $c = \ln(n)$, a więc A_i to zdarzenie polegające na niezyskaniu i -tego kuponu po $2n \ln(n)$

krokach, a $2n \ln(n) \leq 2nH_n$, a więc

$$P(X \geq 2nH_n) \leq P(X \geq 2n \ln(n)) = P\left(\bigcup_{i=1}^n A_i\right) \leq \frac{1}{e^{\ln(n)}} = \frac{1}{n}$$

Przykład 6.1.2. Nierówność Markowa dla kolekcjonera kuponów

Spróbujmy użyć nierówności Markowa do oszacowania jakoś czasu zebrania wszystkich kuponów. Niech X będzie czasem zebrania wszystkich kuponów. W sekcji 4.1.1 pokazaliśmy, że $\mathbb{E}[X] = nH_n = \Theta(n \ln n)$

Możemy zatem skorzystać z nierówności Markowa aby otrzymać

$$P(X \geq 2nH_n) \leq \frac{\mathbb{E}[X]}{2nH_n} = \frac{1}{2}$$

Nie jest to jakieś szczególnie satysfakcjonujące oszacowanie – prawdopodobieństwo, że musimy czekać dwa razy dłużej niż tego oczekujemy może wynosić aż $\frac{1}{2}$:(

6.2 Nierówność Czebyszewa

Nierówność Markowa nie daje nam zbyt dobrych ograniczeń, ale jeśli jedyne co wiemy o zmiennej X to jej wartość oczekiwana, to to jest i tak dobry wynik.

6.2.1 Definicja

Spodziewamy się, że jeśli wiemy coś więcej o rozkładzie X to możemy lepiej szacować pewne prawdopodobieństwa.

Faktycznie, tak jest – w nierówności Czebyszewa przyjmujemy że znamy wariancję.

Twierdzenie 6.2.1 (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

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

□

6.2.2 Rzuty monetą

Rzucamy n razy monetą - zliczamy liczbę orłów.

Niech $X_i = \begin{cases} 1, & \text{orzeł za } i\text{-tym razem} \\ 0 & \text{wpp} \end{cases}$

$$\forall_{i \in [n]} \mathbb{E}[X_i] = \frac{1}{2} \wedge \mathbb{E}[X_i^2] = \frac{1}{2}$$

$$\text{Var}[X_i] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

Ponieważ zmienne losowe X_i są od siebie niezależne, to z 4.3.5 mamy

$$\text{Var}[X] = \sum_{i=1}^n \text{Var}[X_i] = \frac{n}{4}$$

a więc, z nierówności Chebysheva

$$P\left(X \geq \frac{3}{4}n\right) = P\left(X - \frac{n}{2} \geq \frac{n}{4}\right) \leq P\left(\left|X - \frac{n}{2}\right| \geq \frac{n}{4}\right) \leq \frac{\frac{n}{4}}{16} = \frac{4}{n}$$

6.2.3 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 $\text{Var}[X]$.
Potrzebne nam do tego będzie następujące twierdzenie

Twierdzenie 6.2.2. Niech X ma rozkład geometryczny z parametrem p . Wtedy

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

Dowód. Pokazaliśmy w 3.5.3, że $\mathbb{E}[X] = \frac{1}{p}$. Pozostaje nam zatem policzyć $\mathbb{E}[X^2]$

Zacniemy od pokazania pomocniczych równości.

Wiemy, że dla $0 < x < 1$

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i$$

Różniczkujemy obustronnie

$$\frac{1}{(1-x)^2} = \sum_{i=0}^{\infty} i x^{i-1} = \sum_{i=0}^{\infty} (i+1) x^i$$

I jeszcze raz

$$\frac{2}{(1-x)^3} = \sum_{i=0}^{\infty} i(i+1) x^{i-1} = \sum_{i=0}^{\infty} (i+1)(i+2) x^i$$

W takim razie

$$\begin{aligned} \sum_{i=1}^{\infty} i^2 x^i &= \sum_{i=0}^{\infty} i^2 x^i \\ &= \sum x^i \cdot (i^2 + 3i + 2 - 3(i+1) + 1) \\ &= \sum (i+1)(i+2) x^i - 3 \sum_{i=1}^{\infty} (i+1) x^i + \sum_{i=1}^{\infty} x_i \\ &= \frac{2}{(1-x)^3} - 3 \cdot \frac{1}{(1-x)^2} + \frac{1}{1-x} \\ &= \frac{x^2 + x}{(1-x)^3} \end{aligned}$$

Teraz możemy przejść do głównych obliczeń:

$$\begin{aligned}
 \mathbb{E}[X^2] &= \sum_{i=1}^{\infty} i^2 \cdot (1-p)^{i-1} p \\
 &= \frac{p}{1-p} \cdot \sum_{i=1}^{\infty} i^2 \cdot (1-p)^i \\
 &= \frac{p}{1-p} \cdot \frac{(1-p)^2 + (1-p)}{p^3} \\
 &= \frac{(1-p) + 1}{p^2} = \frac{2-p}{p^2}
 \end{aligned}$$

Dowód kończymy obliczając wariancję

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

□

Alternatywny dowód

Dowód. Niech Y ma rozkład geometryczny z parametrem p

Niech $X = \begin{cases} 1 & \text{w pierwszym rzucie orzeł} \\ 0 & \text{w pierwszym rzucie reszka} \end{cases}$

Z 3.3.1 mamy:

$$\begin{aligned}
 \mathbb{E}[Y^2] &= P(X=0)\mathbb{E}[Y^2 | X=0] + P(X=1)\mathbb{E}[Y^2 | X=1] \\
 &= (1-p)\mathbb{E}[Y^2 | X=0] + p\mathbb{E}[Y^2 | X=1]
 \end{aligned}$$

Wiemy, że

- Dla $X=1$, $Y=1$, a więc $\mathbb{E}[Y^2 | X=1] = 1$
- Dla $X=0$, $Y > 1$. Niech Z to liczba pozostałych rzutów do otrzymania pierwszego orła

W takim razie

$$\mathbb{E}[Y^2] = (1-p)\mathbb{E}[(Z+1)^2] + p = (1-p)\mathbb{E}[Z^2] + 2(1-p)\mathbb{E}[Z] + 1$$

Wiemy z 3.5.1, że rozkład geometryczny jest bez pamięci, a więc Z też ma rozkład geometryczny z parametrem p , a więc $\mathbb{E}[Z] = \frac{1}{p}$, $\mathbb{E}[Z^2] = \mathbb{E}[Y^2]$, co znaczy, że

$$\mathbb{E}[Y^2] = (1-p)\mathbb{E}[Y^2] + 2(1-p)\frac{1}{p} + 1 \implies p\mathbb{E}[Y^2] = \frac{2-p}{p} \implies \mathbb{E}[Y^2] = \frac{2-p}{p^2}$$

□

Wracamy do zastosowania nierówności Czebyszewa dla kolekcjonera kuponów. Skorzystamy tutaj z bardzo wygodnego twierdzenia 4.3.5 a następnie z 6.2.2 aby dostać

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

Teraz wkładamy to do nierówności Czebyszewa:

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

6.3 Funkcje tworzące momentów

Definicja 6.3.1. Funkcję tworzącą momenty zmiennej losowej X definiujemy jako

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

Twierdzenie 6.3.1 (Twierdzenie 4.1 P&C). Jeśli $M_X(t)$ tworzy momenty zmiennej X to

$$\mathbb{E}[X^n] = M_X^{(n)}(0)$$

Dowód. Zakładamy tutaj, że możemy zamieniać kolejnością operatory różniczkowania i wartości oczekiwanej. To założenie działa jeśli tworząca istnieje blisko zera i okazuje się, że zachodzi dla rozkładów, którymi się będziemy zajmować.

$$\begin{aligned} M_X^{(n)}(t) &= \mathbb{E}[e^{tX}]^{(n)} = \mathbb{E}[X^n e^{tX}] \\ M_X^{(n)}(0) &= \mathbb{E}[X^n] \end{aligned}$$

□

Twierdzenie 6.3.2. (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.

Twierdzenie 6.3.3 (Twierdzenie 4.3 P&C). Dla niezależnych zmiennych losowych X oraz Y zachodzi

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

Dowód.

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

□

Twierdzenie 6.3.4. 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.

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

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 6.3.1. Wcześniej obliczyliśmy już $\mathbb{E}[X]$ oraz $\mathbb{E}[X^2]$. Możemy to też zrobić z funkcji tworzącej momentów.

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:

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

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

6.4 Nierówność Chernoffa

6.4.1 Definicja

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

Twierdzenie 6.4.1.

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

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} . $\square$

6.4.2 Rzuty monetą

Definicja 6.4.1. 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**.

Przykład 6.4.1. Chcemy ograniczyć z góry prawdopodobieństwo, że przy n rzutach monetą wyrzucimy orła więcej niż $\frac{3}{4}n$ razy.

Niech $X_1, \dots, X_n$ - ciąg niezależnych prób Poissona oraz $P(X_i = 1) = p_i$. Niech $\mu = \mathbb{E}[X] = \sum_{i=1}^n p_i$. Niech $\delta > 0$ będzie ustalone.

Liczymy z funkcji tworzącej:

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

Z niezależności:

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

Teraz stosujemy to dla monet:

$$M_{X_i}(t) \leq e^{\frac{1}{2}(e^t-1)} \implies M_X(t) \leq e^{\frac{1}{2}n(e^t-1)}$$

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

Następnie minimalizujemy wykładnik:

$$f(t) = e^t - 1 - \frac{3}{2}t \implies f'(t) = e^t - \frac{3}{2}$$

Stąd mamy wartość minimalną f dla $t = \ln(\frac{3}{2})$ równą $-\frac{1}{10}$

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

Wykład 5 (2025-11-07)

7.1 Paradoks dnia urodzin

Zadajmy sobie pytanie - ile osób w sali wystarczy, aby prawdopodobieństwo tego, że dwie osoby mają urodziny w tym samym dniu było większe od $\frac{1}{2}$?

Zakładamy, że każdy dzień urodzin jest równie prawdopodobny, oraz że rok ma 365 dni.

Odpowiedź: wystarczy 23 osoby.

Przykład 7.1.1. Sprawdźmy najpierw co się stanie dla 30 osób.

$$P(\text{Żaden dzień się nie powtórzy}) = \frac{\binom{365}{30} \cdot 30!}{365^{30}} = \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{29}{365}\right) \approx 0,29$$

Przykład 7.1.2. Spróbujmy teraz trochę bardziej ogólnie.

Niech n - liczba dni w roku.

m - wymagana liczba osób w pokoju.

Skorzystamy tutaj z takiego przybliżenia (przy założeniu, że k jest małe względem n):

$$1 - \frac{k}{n} \approx e^{-\frac{k}{n}}$$

$$\prod_{j=1}^{m-1} \left(1 - \frac{j}{n}\right) \approx \prod_{j=1}^{m-1} e^{-\frac{j}{n}} = e^{-\sum_{j=1}^{m-1} \frac{j}{n}} = e^{-\frac{m(m-1)}{2n}} \approx e^{-\frac{m^2}{2n}}$$

Następnie szukamy m :

$$e^{-\frac{m^2}{2n}} = \frac{1}{2}$$

Logarytmujemy stronami:

$$-\frac{m^2}{2n} = \ln\left(\frac{1}{2}\right)$$
$$m = \sqrt{2n \ln 2}$$

Podstawiając $n = 365$:

$$m \approx 22,49$$

Użyliśmy tutaj kilku przybliżeń, ale wynik wciąż jest bardzo blisko poprawnego 23

Przykład 7.1.3. Spróbujmy jeszcze bardziej ogólnie.

Niech E_i - i -ta osoba ma dzień urodzin różny od wszystkich osób przed nią.

$\overline{E_i}$ - i -ta osoba powtórzyła dzień urodzin.

Wtedy prawdopodobieństwo tego, że jakiś dzień się powtórzy to:

$$P(\overline{E_1} \cap \dots \cap \overline{E_m}) = P(\overline{E_1} \cup \dots \cup \overline{E_m}) \leq \sum_{j=1}^m P(\overline{E_j}) \leq \sum_{i=1}^m \frac{i-1}{n} = \frac{m(m-1)}{2n}$$

Teraz możemy zauważyć, że dla $m \leq \sqrt{n}$ to prawdopodobieństwo będzie mniejsze od $\frac{1}{2}$ (dla $n = 365, m \approx 19,1$).

Możemy też oszacować co się stanie dla $m \geq 2\sqrt{n}$:

$$P(\text{dzień urodzin się nie powtórzy}) < \left(1 - \frac{\sqrt{n}}{n}\right)^{\sqrt{n}} < \frac{1}{e} < \frac{1}{2}$$

Czyli dzielimy nasze osoby na dwie grupy po $\sqrt{n}$ osób. Zakładamy, że w pierwszej nie będzie żadnych powtórzeń. Wtedy każda osoba z drugiej grupy musi ominąć przynajmniej $\sqrt{n}$ miejsc i stąd nasze oszacowanie.

7.2 Kule i urny

Zanim zaczniemy, zaprezentujemy dwa proste lematy potrzebne w oszacowaniu

Lemat 7.2.1. 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 7.2.2. 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 7.2.1 (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 podzbiórów 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.

□

7.3 Kule i urny a rozkład Poissona

Rozważmy ponownie model z urnami i kulami - rzucamy m kul do n urn. Zbadajmy indykatory X_i zdarzeń polegających na tym, że i -ta urna jest pusta.

$$\mathbb{E}[X_i] = \left(1 - \frac{1}{n}\right)^m$$

Zanim przejdziemy dalej warto przyjąć następujące przybliżenie:

Lemat 7.3.1.

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

Dowód. Jeśli przyjmiemy najlepiej że m jest duże i podstawimy $\lambda = \frac{m}{n}$ dostajemy:

$$\left(1 + \frac{-\lambda}{m}\right)^m \approx e^{-\lambda}$$

□

Niech $X = \sum X_i$ zlicza liczbę pustych urn. Przyjmując powyższe przybliżenie dostajemy:

Twierdzenie 7.3.1.

$$\mathbb{E}[X] \approx ne^{-\frac{m}{n}}$$

Dowód.

$$\mathbb{E}[X] = \sum \mathbb{E}[X_i] = n \left(1 - \frac{1}{n}\right)^m \approx ne^{-\frac{m}{n}}$$

□

Podobnie postąpimy szacując prawdopodobieństwo, że w konkretnej urnie jest r kul.

Twierdzenie 7.3.2.

$$P(\text{W danej urnie jest } r \text{ kul}) \approx \frac{e^{-\frac{m}{n}} \left(\frac{m}{n}\right)^r}{r!}$$

Dowód.

$$\begin{aligned} P(\text{W danej urnie jest } r \text{ kul}) &= \binom{m}{r} \left(\frac{1}{n}\right)^r \left(1 - \frac{1}{n}\right)^{m-r} \\ &= \frac{1}{r!} \cdot \frac{m(m-1) \dots (m-r+1)}{n^r} \left(1 - \frac{1}{n}\right)^{m-r} \\ &\approx \frac{1}{r!} \left(\frac{m}{n}\right)^r \left(1 - \frac{1}{n}\right)^{m-r} \\ &\approx \frac{\left(\frac{m}{n}\right)^r e^{-\frac{m}{n}}}{r!} \end{aligned}$$

□

7.4 Rozkład Poissona

Definicja 7.4.1. 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 7.4.1. 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 7.4.2 (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 $\sum_0^{\infty} \frac{x^n}{n!} = \exp(x)$ □

Twierdzenie 7.4.3 (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. Są dwie ścieżki aby to pokazać – pierwszy polega na przeliczeniu explicite $P(X + Y = n)$.

$$\begin{aligned}
 P(X + Y = n) &= \sum_{k=0}^{\infty} P(X = k \cap Y = n - k) \\
 &= \sum_{k=0}^{\infty} P(X = k) P(Y = n - k) \\
 &= \sum_{k=0}^{\infty} \frac{e^{-\lambda_X} \lambda_X^k}{k!} \cdot \frac{e^{-\lambda_Y} \lambda_Y^{n-k}}{(n-k)!} \\
 &= \frac{e^{-\lambda_X - \lambda_Y}}{n!} \sum_{k=0}^{\infty} \frac{\lambda_X^k \lambda_Y^{n-k} n!}{k!(n-k)!} \\
 &= \frac{e^{-\lambda_X - \lambda_Y}}{n!} \sum_{k=0}^{\infty} \lambda_X^k \lambda_Y^{n-k} \binom{n}{k} \\
 &= \frac{e^{-\lambda_X - \lambda_Y}}{n!} (\lambda_X + \lambda_Y)^n
 \end{aligned}$$

Co daje nam rozkład Poissona z parametrem $\lambda_X + \lambda_Y$. □

Dowód. Pokażemy teraz bardziej elegancki dowód korzystający z funkcji tworzących momentów.

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 $\lambda_X + \lambda_Y$ □

7.5 Ograniczenia Chernoffa zmiennej o rozkładzie Poissona

Twierdzenie 7.5.1. 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

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

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

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

7.6 Granica rozkładu dwumianowego

Twierdzenie 7.6.1. Niech $n, \lambda \in \mathbb{R}$, p jest funkcją n . X_n jest zmienną losową o o rozkładzie dwumianowym z parametrami n, p . Dodatkowo

$$\lim_{n \rightarrow \infty} np = \lambda$$

Wtedy

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

Dowód. Bez dowodu, ponoć idzie z twierdzenia o trzech funkcjach $\square$

7.7 Aproksymacja Poissona

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 7.7.1 (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 7.7.2. 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.

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

□

Wykład 6 (2025-11-14)

8.1 Kule i urny - lower bound

W twierdzeniu 7.2.1 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 8.1.1 (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 7.7.2 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. □

8.2 Kule i urny - kolekcjoner kuponów

Myśleliście, że poprzedni dowód był brzydki i miał dużo obliczeń? No to teraz wyrzuci Was ze skarpetek, bo ten dowód będzie jeszcze gorszy.

Twierdzenie 8.2.1. 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 7.7.1, że wolno nam jedynie przejść równością warunkową tj. $P(\mathcal{E} \mid X = m)$, a tego nie znamy.

Aby sobie z tym poradzić rozbijamy 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

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

A to jest dokładnie to co chcieliśmy pokazać.

□

8.3 Ograniczenia Chernoffa prób Poissona

Twierdzenie 8.3.1. Niech $X_1, \dots, X_n$ to niezależne próby Poissona, czyli $P(X_i = 1) = p_i, P(X_i = 0) = 1 - p_i$. Dodatkowo oznaczamy $X = \sum X_i$ i $\mu = \mathbb{E}[X]$. Wtedy:

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

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

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

Dowód. Zaczniemy od punktu 1, a pozostałe wynikną z niego.

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

Podstawiamy tutaj $t = \ln(1 + \delta)$

$$\begin{aligned} \frac{e^{(e^t - 1)\mu}}{e^{t(1 + \delta)\mu}} &= \frac{e^{\delta\mu}}{e^{(1 + \delta)\mu \ln 1 + \delta}} \\ &= \frac{e^{\delta\mu}}{(1 + \delta)^{(1 + \delta)\mu}} \\ &= \left(\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \right)^\mu \end{aligned}$$

□

Dowód. Do drugiego punktu pokażemy, że $\frac{e^\delta}{(1 + \delta)^{1 + \delta}} \leq e^{-\frac{\delta^2}{3}}$ dla $\delta \in (0, 1)$. Weźmiemy obustronnie logarytm i po odrobinie analizy wyjdzie:

$$\begin{aligned} f(\delta) &= \delta - (1 + \delta) \ln 1 + \delta + \frac{\delta^2}{3} \leq 0 \\ \frac{d}{d\delta} f(\delta) &= 1 - \ln 1 + \delta = \frac{1 + \delta}{1 + \delta} + \frac{2\delta}{3} \\ \frac{d^2}{d\delta^2} f(\delta) &= \frac{2}{3} - \frac{1}{1 + \delta} \end{aligned}$$

Zauważmy, że druga pochodna jest równa 0 tylko dla $\delta = \frac{1}{2}$: dla mniejszych wartości jest ujemna, a dla większych dodatnia. Pierwsza pochodna jest w zerze równa 0 i w jedynce ujemna

- wraz z poprzednim implikuje to, że dla $\delta \in (0, 1)$ mamy ujemną pochodną. Dodając do tego jeszcze $f(0) = 0$ dostajemy zadaną na początku nierówność. $\square$

Dowód. Do trzeciego punktu położymy $R = (1 + \delta)\mu$. I dalej:

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

$\square$

Wykład 7 (2025-11-19)

9.1 Łańcuchy Markowa - definicje

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 .

Mówimy, że proces jest **skończony**, jeśli zmienne X_t przyjmują skończenie wiele wartości. Proces jest także **dyskretny**, jeśli zmienne te przyjmują wartości ze zbioru przeliczalnego.

Definicja 9.1.2. Proces jest **z czasem dyskretnym**, jeśli T jest przeliczalne (najczęściej $T = \mathbb{N}$).

Definicja 9.1.3. 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 9.1.4. Oznaczamy dalej $p_{ij} = P(X_t = j \mid X_{t-1} = i)$.

Macierzą przejścia nazywamy macierz $\mathbf{P}$ zadaną współczynnikami p_{ij} .

Definicja 9.1.5. $p_{ij}(n) = P(X_n = j \mid X_0 = i)$

9.2 Stany - definicje

Definicja 9.2.1. Stan i jest **pochłaniający** jeśli $p_{ii} = 1$.

Definicja 9.2.2. Stan j jest **osiągalny** ze stanu i jeśli istnieje $n \geq 0$ takie, że $p_{ij}(n) > 0$. Zapisujemy $i \rightarrow j$.

Definicja 9.2.3. 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 – $p_{ij}(0) = 1$

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 9.2.4. Stan i jest **nieistotny** jeśli $\exists j : i \rightarrow j \wedge j \nrightarrow i$

9.3 Łańcuchy nieprzywiedlne

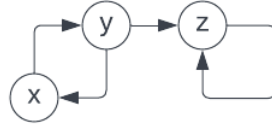
Definicja 9.3.1. Okres Stanu Niech $i \in S$.

Definiujemy:

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

Czyli zbiór takich t , że jesteśmy w stanie dojsć z i do i w t krokach.

Okres stanu i definiujemy jako $o(i) = \gcd(\mathcal{T}(i))$.



Rysunek 9.1: Stan z ma okres 1

Definicja 9.3.2. Łańcuch jest **nieprzywiedlny (nieredukowalny)** jeśli wszystkie stany są parami skomunikowane. Wtedy jego graf skierowany jest silnie spójny.

Definicja 9.3.3. 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, które nie są okresowe nazywamy nieokresowymi.

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 dojsć 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 .

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

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

$$\implies o(i) \mid n \implies o(i) \leq o(j)$$

Analogicznie dowodzimy w drugą stronę, otrzymując równość. □

Lemat 9.3.2. Jeśli $(X_t \mid t \in \mathbb{N})$ jest nieokresowym (okres jest równy 1) 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 $n_0 = m_0 + m_1$

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

□

9.4 Stan powracający i czas pierwszego spotkania

Definicja 9.4.1. Definiujemy **piewszy 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 9.4.2. 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 9.4.3. 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 9.4.4. 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 9.4.5. Definiujemy **czas pierwszego spotkania**

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

dodatkowo

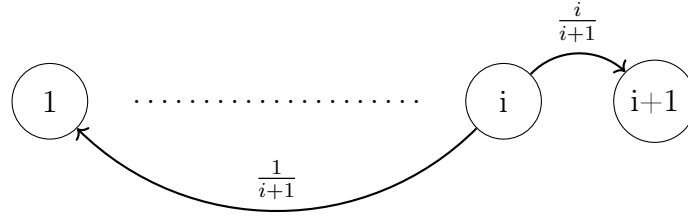
$$\mathbb{E}[T_{i,j}] = \mathbb{E}[T_j^+ \mid X_0 = i] = \sum_{n=1}^{\infty} n P(T_j^+ = n \mid X_0 = i)$$

Definicja 9.4.6. Stan powracający i jest **dodatni** jeśli $\mathbb{E}[T_{i,i}] < \infty$, w przeciwnym wypadku jest **zerowy**.

Przykład 9.4.1. Niech zbiór stanów $S = \mathbb{N}_1$ oraz niech

$$P(X_{j+1} = i + 1 \mid X_j = i) = \frac{i}{i + 1}$$

$$P(X_{j+1} = 1 \mid X_j = i) = \frac{1}{i + 1}$$



$$P(T_1^+ > t \mid X_0 = 1) = \frac{1}{2} \cdot \frac{2}{3} \cdots \frac{t}{t+1} = \frac{1}{t+1}$$

$$f_{1,1} = P(T_1^+ < \infty \mid X_0 = 1) = 1 - \lim_{t \rightarrow \infty} P(T_1^+ > t \mid X_0 = 1) = 1 - \lim_{t \rightarrow \infty} \frac{1}{t+1} = 1$$

$$\mathbb{E}[T_{1,1}] = \sum_{t=1}^{\infty} t P(T_i^+ = t \mid X_0 = i) = \sum_{t=1}^{\infty} t \left(\frac{1}{2} \cdots \frac{t-1}{t} \frac{1}{t+1} \right) = \sum_{t=1}^{\infty} \frac{1}{t+1} \rightarrow \infty$$

Z $f_{1,1} = 1$ stan 1 jest powracający, a z $\mathbb{E}[T_{1,1}] \rightarrow \infty$ jest zerowy.

Twierdzenie 9.4.1. W skończonym, nieprzywiedlnym łańcuchu Markowa zachodzi

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

Dowód. Nieprzywiedlność oraz skończoność dają nam

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

Mimo tego, że linia ta może początkowo być trudna do przetworzenia, jest całkiem prosta. Nieprzywiedlność mówi nam że dla każdego (x, y) , y jest osiągalne z x , a więc $\exists_n \quad p_{x,y}(n) > 0$. 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 $\exists_{j \in [r]} \quad p_{z,y}((k-1)r + j) > \varepsilon$, 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

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

□

¹Wynika to z 3.5.2

9.5 Rozkład stacjonarny

Definicja 9.5.1. Rozkładem stacjonarnym nazywamy wektor $\bar{\pi}$ taki, że $\bar{\pi} = \bar{\pi}P$ oraz $\sum_i \bar{\pi}_i = 1$

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 9.5.1. Istnienie rozkładu stacjonarnego Niech $z \in S$, $\bar{\pi}_z = (\bar{\pi}_{z,y})_{y \in S}$. Dodatkowo niech

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

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

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

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

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

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

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

Pozostaje nam więc tylko udowodnić część pierwszą

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

$P_z(X_0 = y) = P_z(X_{T_z^+} = y)$ ponieważ oba są indykatorami $x = y$. □

Wykład 8 (2025-11-21)

10.1 Rozkład stacjonarny - kontynuacja

Twierdzenie 10.1.1. 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 $2, 3, \dots$ 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 $\sum_{i \in S} \pi_i = \sum_{i \in S} \phi_i = 1$

□

Twierdzenie 10.1.2. 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 9.5.1 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 10.3.1. $\square$

10.2 Norma całkowitego wahania

Interesuje nas pytanie, jak długo należy prowadzić eksperyment aby rozkład aktualnego stanu był bliski rozkładowi stacjonarnemu. Oczywiście pytanie to ma sens tylko dla łańcuchów nieokresowych. Co jednak nawet znaczy "bliski" w tym kontekście?

Przykład 10.2.1. Tasujemy n kart. W każdym kroku:

- Wybieramy kartę losowo, jednostajnie i niezależnie od poprzednich kroków
- Kładziemy wybraną kartę na wierzchu talii

Niech X_0 to pewna ustalona permutacja początkowa, a X_n to permutacja kart po n krokach. Możemy zauważyć, że $(X_n)_{n \in \mathbb{N}}$ jest skończonym, nieprzywiedlnym oraz nieokresowym (bo możemy wziąć kartę z wierzchu) łańcuchem Markowa, a więc ma rozkład stacjonarny $(\pi_x)_{x \in S}$.

Niech $x \in S$. $N(x)$ to zbiór stanów osiągalnych z x w jednym kroku. Oczywiście $|N(x)| = n$ oraz

$$\pi_x = \frac{1}{n} \sum_{y \in N(x)} \pi_y$$

Możemy zauważyć, że rozkład jednostajny spełnia ten układ równań, a więc z unikalności rozkładu stacjonarnego

$$\pi_x = \frac{1}{n!}$$

Jeśli wierzymy w to, że odpowiednio długo tasując zbiegamy do rozkładu stacjonarnego, to jest to dobre tasowanie, bo w rozkładzie stacjonarnym każda permutacja jest równie prawdopodobna.

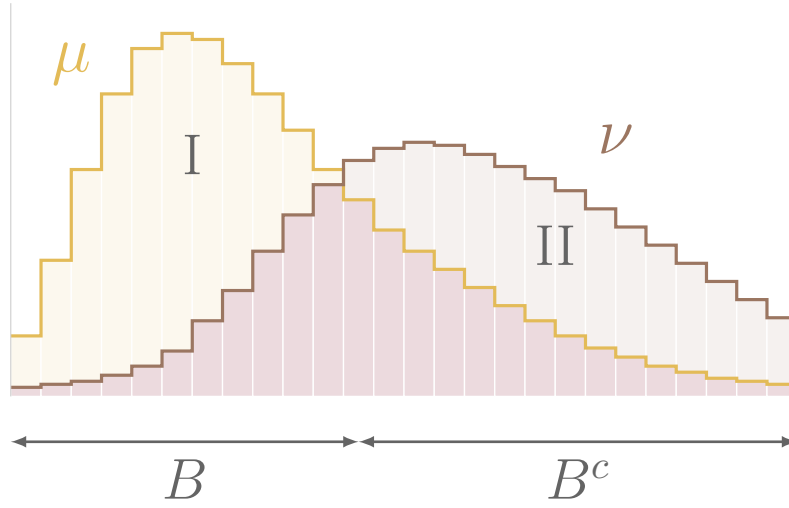
Definicja 10.2.1. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S . Normą całkowitego wahania (total variation distance) tych rozkładów nazywamy wartość

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

Lemat 10.2.1. 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 wahania. 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 wahania, 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 10.2.2. 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

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

□

Przykład 10.2.2. Powróćmy na moment do poprzedniego przykładu z tasowaniem kart. Ustalmy π jako rozkład stacjonarny oraz x jako pewną ustaloną permutację kart. Będziemy tasować, aż obecny stan ma rozkład D taki, że $\|D - \pi\|_{TV} > \epsilon$ dla pewnego $\epsilon > 0$.

$$\epsilon > \|D - \pi\|_{TV} = \max_{A \subseteq S} |D(A) - \pi(A)| \geq \left| D(x) - \frac{1}{n!} \right|$$

$$\frac{1}{n!} - \epsilon < D(x) < \frac{1}{n!} + \epsilon$$

Przykład 10.2.3. Dla kontrastu, gdy będziemy oszukiwać przy tasowaniu zostawiając asa pik na wierzchu talii. Ustalmy B jako zbiór wszystkich permutacji z asem pik na wierzchu.

$$\begin{aligned}\epsilon &> \|D - \pi\|_{TV} = \max_{A \subseteq S} |D(A) - \pi(A)| \\ &\geq |D(B) - \pi(B)| = \left|1 - \frac{1}{52}\right| = \frac{51}{52}\end{aligned}$$

10.3 Sprzęganie

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 10.3.1. 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 10.3.2. 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 10.3.1. 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 ustalamy $X = x$ i $Y = y$, gdzie x jest wybierany losowo z S z rozkładem $\left(\frac{1}{p_1} \max(\mu(x) - \nu(x), 0) : x \in S\right)$, a y z rozkładem $\left(\frac{1}{p_2} \max(\nu(y) - \mu(y), 0) : y \in S\right)$. W przypadku reszki jedna zmienna przyjmuje tylko te wartości, na których μ jest większe, a druga tylko te, na których ν jest większe.

Mamy więc $P(X \neq Y) = 1 - p_3 = \|\mu - \nu\|_{TV}$, a (X, Y) faktycznie jest sprzęganiem μ i ν – zmienne mają odpowiednie rozkłady. $\square$

Definicja 10.3.3. 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.

Lemat 10.3.2. Niech $((X_t, Y_t))_{t \in \mathbf{N}}$ będzie sprzęganiem łańcuchów (skończonych, nieprzywiedlnych, nieokresowych) z macierzą przejścia P i zbiorem stanów S . Niech $T \in \mathbf{N}$ 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

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

□

Lemat 10.3.3 (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 sprzężaniem rozkładów $P^t(x, \cdot)$ i π spełniającym $P(X_t \neq Y_t) = \Delta_x(t)$ (przedtem pokazaliśmy, że istnieje sprzężanie, dla którego ta równość zachodzi). Definiujemy (X_{t+1}, Y_{t+1}) w następujący sposób: jeśli $X_t = Y_t$, wykonujemy krok łańcucha zgodnie z macierzą P (na obu współrzędnych taki sam), a w przeciwnym wypadku wykonujemy dwa niezależne kroki. Zauważmy, że Y_{t+1} dalej ma rozkład π . Mamy

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

□

Twierdzenie 10.3.1 (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: mając zadane $X_t = x, Y_t = y$ (początkowe wartości wybieramy dowolnie) oznaczamy $\mu = P^r(x, \cdot), \nu = P^r(y, \cdot)$ i $B = \{x \in S : \mu(x) \geq \nu(x)\}$. Niech $p_1 = \mu(B) - \nu(B)$, $p_2 = \nu(B^c) - \mu(B^c)$ i $p_3 = 1 - p_1 = 1 - p_2$.

Rzucamy monetą z prawdopodobieństwem orła p_3 . Jeśli wypadnie orzeł, to ustalamy $X_{t+1} = Y_{t+1} = s$, gdzie s wybieramy z S z rozkładem $\left(\frac{1}{p_3} \min(\mu(s), \nu(s)) : s \in S\right)$. Jeśli wypadnie reszka ustalamy $X_{t+1} = x$ i $Y_{t+1} = y$, gdzie x wybieramy 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 ten sposób skonstruowaliśmy sprzężanie, dla którego zachodzi

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

Mamy więc

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

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

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

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

Wykład 9 (2025-11-28)

11.1 Prawdopodobieństwo ciągłe - definicje

Definicja 11.1.1. Zmienna losowa X jest ciągła, jeśli istnieje funkcja $f : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ taka, że

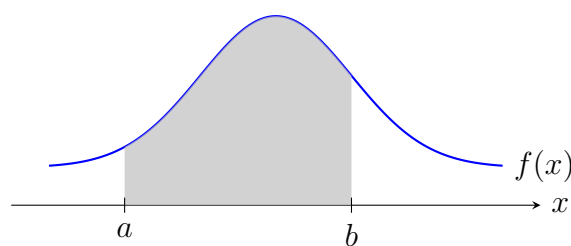
$$\forall B \in \mathcal{B}(\mathbb{R}) \quad P(X \in B) = \int_B f(x) dx.$$

Taką funkcję nazywamy funkcją gęstości lub gęstością zmiennej X .

Własności funkcji gęstości:

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

Wizualizujemy to jako pole pod wykresem:



Rysunek 11.1: Przykład dla $P(a < X < b)$

Mamy $\forall x \in \mathbb{R} \quad P(X = x) = 0$. W szczególności daje to $P(X \leq x) = P(X < x)$. Dlaczego tak jest? Można to pokazać na przykładzie jednostajnym. Powiedzmy, że losujemy wartości z przedziału $[0, 1)$. Wtedy $P(X = x) = 0$. Weźmy teraz jakiś zbiór S k punktów. Wtedy $P(X \in S) = kp$. Z tego, że jest to prawdopodobieństwo, mamy $\forall k \in \mathbb{N} \quad kp \leq 1$ a z tego dostajemy $p = 0$.

Definicja 11.1.2. Dystrybuanta zmiennej losowej X to funkcja $F(x) = P(X \leq x)$. Dla zmiennej ciągłej jest $F(x) = \int_{-\infty}^x f(y) dy$, a więc $f(x) = F'(x)$.

Definicja 11.1.3. Wartość oczekiwana ciągłej zmiennej X to

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

Definicja 11.1.4. Wariancja ciągłej zmiennej X :

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

Lemat 11.1.1 (Lemat 8.1 P&C). Jeśli zmienna losowa X przyjmuje wartości nieujemne to

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

Dowód. Niech f będzie gęstością X . Mamy

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

gdzie ostatnie przejście wynika z nieujemności X . □

Definicja 11.1.5. Wspólna dystrybuanta zmiennych losowych X, Y to

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

Definicja 11.1.6. Wspólna gęstość ciągłych zmiennych losowych X, Y to funkcja f taka, że

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

a więc

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

Definicja 11.1.7. Dla dwóch zmiennych X, Y o zadanej wspólnej dystrybuancie $F(x, y)$ brzegowa dystrybuanta zmiennej X to funkcja

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

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

Definicja 11.1.8. Zmienne X, Y są niezależne, jeśli

$$\forall_{x,y \in \mathbf{R}} P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y).$$

Niezależność jest równoważna odpowiednim równościom dystrybuant i gęstości:

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

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

Definicja 11.1.9. Prawdopodobieństwo warunkowe definiujemy jako całkę

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

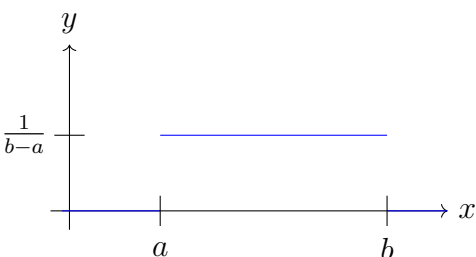
gdzie funkcję $f_{X|Y} = \frac{f(x,y)}{f_Y(y)}$ nazywamy warunkową gęstością.

Definicja 11.1.10. Warunkowa wartość oczekiwana to całka

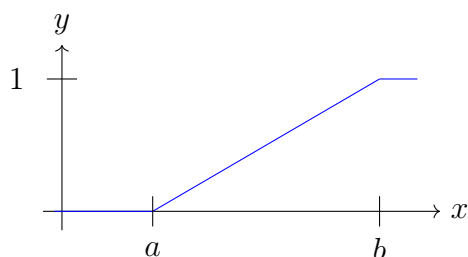
$$\mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x, y) dx.$$

11.2 Rozkład jednostajny

Definicja 11.2.1. 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 11.2.1. 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 11.2.2. 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 11.2.3. 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 11.2.4. 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 łuku $P_0, \dots, P_n$. W ten sposób X_i jest odległością zgodnie ze wskazówkami zegara punktów 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 $\frac{k}{n+1}$. □

Alteratywny dowód (wyłącznie dla $k = 1$)

Dowód.

$$Y_1 = \min((X_1, X_2, \dots))$$

$$\begin{aligned} P(Y_1 \geq y) &= P(\min((X_1, X_2, \dots)) \geq y) \\ &= P\left(\bigcap_{i=1}^n X_i \geq y\right) \\ &= \prod_{i=1}^n P(X_i \geq y) \\ &= \prod_{i=1}^n (1 - y) \\ &= (1 - y)^n \end{aligned}$$

$$\begin{aligned} \mathbb{E}[Y_1] &= \int_0^1 P(Y_1 \geq y) \, dy \\ &= \int_0^1 (1 - y)^n \, dy \\ &= \left[-\frac{(1 - y)^{n+1}}{n + 1} \right]_0^1 \\ &= -\frac{(1 - 1)^{n+1}}{n + 1} + \frac{(1 - 0)^{n+1}}{n + 1} \\ &= \frac{1}{n + 1} \end{aligned}$$

□

Według Micka można jakoś ten dowód rozszerzyć na $k > 1$, ale nie zrobił tego na wykładzie

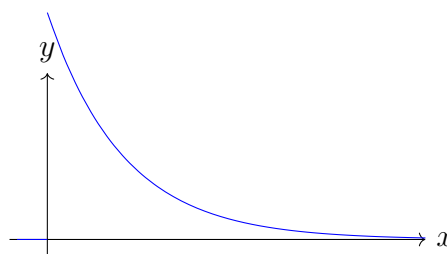
Wykład 10 (2025-12-05)

12.1 Rozkład wykładniczy

12.1.1 Właściwości

Definicja 12.1.1. Rozkładem wykładniczym z parametrem λ nazywamy rozkład zadany gęstością

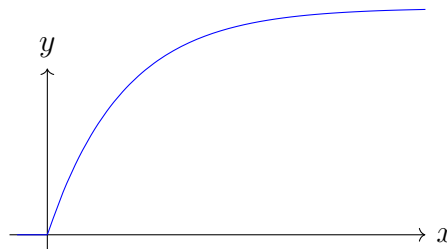
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla } x \geq 0 \\ 0 & \text{wpp.} \end{cases}$$



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 12.1.1. 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 12.1.1. 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 12.1.1 rozszerzyć na dowolne λ .

Twierdzenie 12.1.2 (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 12.1.3. Rozkład wykładniczy jest jedynym ciągłym rozkładem bez pamięci o dziedzinie $[0, \infty)$, czyli jeśli X jest ciągłą zmienną losową i zachodzi

$$\forall_{s,t>0} P(X > s + t \mid X > t) = P(X > s),$$

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

Dowód. Na razie bez dowodu

□

Twierdzenie 12.1.4 (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:

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

$$(-\lambda_2) \cdot \frac{-\lambda_1}{\lambda_2 \cdot (\lambda_1 + \lambda_2)} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

□

12.1.2 Kantory

Jesteśmy na lotnisku, na którym stoi n kantorów obsługujących klientów. W chwili t kiedy do nich podchodzimy, wszystkie z nich są zajęte. Przyjmijmy, że czas obsługi przez i -ty kantor ma rozkład $X_i \sim \text{Exp}(1)$. Jaki jest rozkład czasu oczekiwania na zwolnienie pierwszego kantoru?

W chwili w której podchodzimy do kantorów, każdy z nich jest zajęty od nieznanej nam chwili wcześniejszej od t . Wiemy jednak, że wszystkie z nich są obecnie zajęte. W takim razie, z faktu, że rozkład wykładniczy jest bez pamięci, możemy myśleć o nich jako o rozkładach wykładniczych rozpoczynających się w momencie t . A więc z 12.1.4, rozkład czasu po którym pierwszy kantor się zwolni to po prostu $\text{Exp}(n)$.

12.2 Kule i urny z feedbackiem

Jak zwykle, zanim zaczniemy to pokażemy pomocniczy lemat:

Lemat 12.2.1. 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 12.2.1. 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 12.2.1 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. $\square$

12.3 Rozkład Gamma

Definicja 12.3.1. 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 12.3.1. $\Gamma(1) = 1$

Dowód.

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

□

Fakt 12.3.2. $\forall_{a>0} \Gamma(a+1) = a\Gamma(a)$

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

Wynika to bezpośrednio z poprzedniego faktu.

Definicja 12.3.2. 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:

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

Zatem jest to poprawny rozkład.

Definicja 12.3.3. Mówimy, że ciągła zmienna losowa ma **rozkład Gamma** z parametrem (a, λ) , jeżeli jej funkcja gęstości jest równa:

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

Dowód. Niech X ma rozkład Gamma z parametrem $(a, 1)$. Niech $Y = \frac{X}{\lambda}$. Wtedy mówimy, że Y ma rozkład Gamma z parametrem (a, λ) . Chcemy pokazać, że to poprawny rozkład. Możemy

to sobie oczywiście przecałkować i zobaczyć, że faktycznie wyjdzie 1 ale można też skorzystać z takiego przekształcenia:

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

□

Fakt 12.3.4. Dla $X \sim \text{Gamma}(a, 1)$,

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

$$\mathbb{E}[X^2] = a(a+1)$$

Dowód.

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

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \frac{1}{\Gamma(a)} x^a e^{-x} \frac{dx}{x} = \frac{1}{\Gamma(a)} \int_0^\infty x^{a+2} e^{-x} \frac{dx}{x} = \frac{\Gamma(a+2)}{\Gamma(a)} = a(a+1)$$

□

Fakt 12.3.5. Dla $Y \sim \text{Gamma}(a, \lambda)$ (Czyli $Y = \frac{X}{\lambda}$, dla $X \sim \text{Gamma}(a, 1)$):

$$\mathbb{E}[Y] = \mathbb{E}\left[\frac{X}{\lambda}\right] = \frac{1}{\lambda} \mathbb{E}[X] = \frac{a}{\lambda}$$

Fakt 12.3.6. $\text{Gamma}(1, 1) \equiv \text{Exp}(1)$

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

Twierdzenie 12.3.1. 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 6.3.2. 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. □