Metody Probabilistyczne Informatyki

Wybrane Dowody

„Tak teraz na to patrzę i myślę, czy ta nierówność nie powinna być w drugą stronę...”

Popełnione przez

Załatany Ponton

V

Nahtamatu

Kraków

Anno Domini 2025

Spis treści

Licencja

Rozdział 1

Prawdopodobieństwo dyskretne

1.1 Definicje

1.1.1 Wartość oczekiwana

Definicja 1.1.1. Wartość oczekiwaną zmiennej losowej X definiujemy jako

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

Twierdzenie 1.1.1 (Lemat 2.9 P&C)**.** Niech X będzie zmienną losową przyjmującą jedynie wartości w liczbach naturalnych. Wtedy

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

Dowód.

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

□

Definicja 1.1.2. Warunkową wartość oczekiwaną $\mathbb{E}[X \mid Y = y]$ definiujemy jako

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

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

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

Lemat 1.1.1.

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

Dowód.

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

□

Lemat 1.1.2 (Lemat Syntaktyczny)**.**

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

Dowód.

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

□

1.1.2 Wariancja

Definicja 1.1.3. Wariancję zmiennej losowej X definiujemy jako

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

Definicja 1.1.4. Kowariancję zmiennych losowych X oraz Y definiujemy jako

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

Twierdzenie 1.1.2.

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

Dowód.

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

□

Twierdzenie 1.1.3. Dla dowolnych zmiennych losowych X, Y zachodzi

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

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

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

□

Twierdzenie 1.1.4. Dla niezależnych zmiennych losowych X, Y

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

a co za tym idzie

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

Dowód.

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

□

Twierdzenie 1.1.5. Niech $X_1,\ldots,X_n$ będą parami niezależne. Wtedy

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

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

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

□

1.1.3 Wyższe momenty

Definicja 1.1.5. n-tym momentem zmiennej losowej X nazywamy $\mathbb{E}[X^n]$

1.2 Indykatory

Definicja 1.2.1. Indykator zdarzenia A to zmienna losowa $Y = \begin{cases} 1 & A \text{ zaszło} \\ 0 & \text{wpp} \end{cases}$. Mamy

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

1.3 Funkcje tworzące momentów

Definicja 1.3.1. Funkcję tworzącą momenty zmiennej losowej X definiujemy jako

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

Twierdzenie 1.3.1 (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.

$$M_X^{(n)}(t) = \mathbb{E}\left[e^{tX}\right]^{(n)} = \mathbb{E}\left[X^n e^{tX}\right]$$
$$M_X^{(n)}(0) = \mathbb{E}[X^n]$$

□

Twierdzenie 1.3.2 (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.

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

□

1.4 Rozkład dwumianowy

Definicja 1.4.1. Mówimy, że zmienna losowa X ma **rozkład dwumianowy** z parametrami n, p (Oznaczana poprzez $B(n,p)$), jeśli dla $j = 0, 1, ..., n$:

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

Twierdzenie 1.4.1. Niech X ma rozkład dwumianowy z parametrami n, p. Wtedy

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

Dowód.

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

□

Twierdzenie 1.4.2. Niech X ma rozkład dwumianowy z parametrami n, p. Wtedy

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

Dowód.

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

Pozostaje nam tylko policzyć $E[X^2]$

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

W takim razie

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

□

Twierdzenie 1.4.3. MGF:

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

Dowód.

$$M_X(t) = \mathbb{E}\big[e^{tX}\big] = \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} e^{tj} =$$

$$= \sum_{j=0}^{n} \binom{n}{j} (pe^t)^j (1-p)^{n-j} =$$

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

□

1.5 Rozkład geometryczny

Definicja 1.5.1. Mówimy, że zmienna losowa X ma **rozkład geometryczny** z parametrem p jeśli dla $n > 0$

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

Twierdzenie 1.5.1 (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 1.5.2. Niech X ma rozkład geometryczny z parametrem p. Wtedy

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

Dowód. Skorzystamy z twierdzenia 1.1.1

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

□

Twierdzenie 1.5.3. Niech X ma rozkład geometryczny z parametrem p. Wtedy

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

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

Zaczniemy 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

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

Teraz możemy przejść do głównych obliczeń:

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

Dowód kończymy obliczając wariancję

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

□

Twierdzenie 1.5.4. 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.

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

Skorzystaliśmy tutaj z faktu, że szereg

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

jest zbieżny. Dzieje się tak gdy

$$\begin{aligned}
(1-p)e^t &< 1 \\
e^t &< \frac{1}{1-p} \\
t &< -\ln(1-p)
\end{aligned}$$

□

1.6 Problem kolekcjonera kuponów

Wyobraźmy sobie problem, który jest bliski wielu osobom. Próbujemy 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 $\mathbb{E}[X_i] = \frac{1}{p_i} = \frac{n}{n-i+1}$

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

1.7 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 1.7.1 (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 $x_1, \ldots, x_n$ będzie wejściowym ciągiem n różnych liczb. Niech $y_1, \ldots y_n$ 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 $Y^{i,j} = \{y_i, \ldots, y_j\}$.

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 znajdywać 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_1 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

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

□

1.8 Kule i urny

Zanim zaczniemy, zaprezentujemy dwa proste lematy potrzebne w oszacowaniu

Lemat 1.8.1. Dla dowolnych $n \geq M$

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

Dowód.

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

□

Lemat 1.8.2. 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 1.8.1 (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 a egzaminie raczej nie będziecie potrzebowali obliczeń.

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

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

Korzystamy teraz z obu lematów i ograniczamy prawdopodobieństwo na co najmniej M kul w urnie przez co najwyżej

$$n\Big(\frac{e}{M}\Big)^M$$

Teraz wstawiamy magiczne M z tezy i dostajemy:

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

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

$$n\left(\frac{\ln\ln n}{\ln n}\right)^{(3\ln n)/(\ln\ln n)} \leq \frac{1}{n}$$

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

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.

□

Rozdział 2

Nierówności na zmiennych losowych

2.1 Nierówność Markowa

2.1.1 Definicja

Twierdzenie 2.1.1. 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

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

□

2.1.2 Kolekcjoner kuponów

Spróbujmy użyć nierówności Markowa do oszacowania jakoś czasu zebrania wszystkich kuponów. Niech X będzie czasem zebrania wszystkich kuponów. W sekcji 1.6 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 $\frac{1}{2}$:(

2.2 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.

2.2.1 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 2.2.1 (Twierdzenie 3.6 P&C)**.** Dla dowolnego $a > 0$

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

Dowód. Korzystamy z nierówności Markowa

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

□

2.2.2 Kolekcjoner kuponów

Niech $X_1, \ldots, X_n$ 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 $\mathrm{Var}[X]$.

Skorzystamy tutaj z bardzo wygodnego twierdzenia 1.1.5 a następnie z 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:

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

2.3 Nierówność Chernoffa

2.3.1 Definicja

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

Twierdzenie 2.3.1.

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

oraz

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

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}. □

2.3.2 Rzuty monetą

Definicja 2.3.1. Próbami Poissona nazywany ciąg zmiennych losowych $X_1, \ldots, X_n$, dla których

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

Ponadto definiujemy

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

Lemat 2.3.1. Niech $X_1, \ldots X_n$ będą niezależnymi próbami Poissona oraz $X = \sum X_i, \mu = \mathbb{E}[X]$.

Wtedy dla $0 < \delta < 1$

$$P(|X - \mu| \geq \delta\mu) \leq 2\exp\left(\frac{-\mu\delta^2}{3}\right)$$

Dowód.

$$P(|X - \mu| \geq \delta\mu) \leq P(X \geq (1+\delta)\mu) + P(X \leq (1-\delta)\mu)$$

□

2.3.3 Przypadki specjalne

Twierdzenie 2.3.2. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi takimi, że $P(X_i = 1) = p_i$ oraz $P(X_i = 0) = 1 - p_i$. Niech $X = \sum_{i=1}^{n} X_i$ oraz $\mu = \mathbb{E}[X]$. Wtedy:

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

Dowód. Liczymy funkcję tworzącą

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

Zatem

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

Ustalmy $t > 0$, mamy

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

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

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

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

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

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

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

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

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

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

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

□

Twierdzenie 2.3.3. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi takimi, że $P(X_i = 1) = p_i$ oraz $P(X_i = 0) = 1 - p_i$. Niech $X = \sum_{i=1}^{n} X_i$, $\mu = \mathbb{E}[X]$ Wtedy dla każdego $\delta \in (0,1)$ zachodzi:

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

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

Twierdzenie 2.3.4. Niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o rozkładzie prawdopodobieństwa $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$. Niech $X = \sum_{i=1}^{n} X_i$. Dla każdego $a > 0$ mamy $P(X \geq a) \leq e^{-\frac{a^2}{2n}}$. Zauważmy, że nie ma sensu rozważać tu odchyleń multiplikatywnych, bo $\mathbb{E}[X] = 0$.

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

Rozwijamy w szereg Taylora:

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

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

Z tego wynika

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

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

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

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

Twierdzenie 2.3.5. Niech $Y_1, \ldots, Y_n$ będą niezależnymi indykatorami $P(Y_i = 0) = P(Y_i = 1) = \frac{1}{2}$. Niech $Y = \sum_{i=1}^{n} Y_i$, $\mu = \mathbb{E}[Y] = \frac{n}{2}$. Wtedy

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

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

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

oraz

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

□

2.3.4 Estymacja parametru

Definicja 2.3.2. Mówimy, że $[\hat{p} - \delta, \hat{p} + \delta]$ jest $(1 - \gamma)$ przedziałem ufności dla parametru p, jeśli $P(p \in [\hat{p} - \delta, \hat{p} + \delta]) \geq 1 - \gamma$. Chcemy, żeby n, γ, δ były małe, ale musi być między nimi jakiś balans.

Bierzemy z dużej populacji próbkę wielkości n wybraną w sposób jednostajny. p to nieznana wartość – szukane prawdopodobieństwo, które chcemy szacować (np. prawdopodobieństwo jakiejś mutacji genetycznej). Niech zmienna losowa $X = \hat{p}n$ oznacza liczbę wystąpień tej mutacji w naszej próbce. Spodziewamy się, że jak n rośnie, to $\hat{p} \to p$.

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

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

Mamy $\mathbb{E}[X] = np$, a więc Czernow daje

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

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

2.3.5 Problem Set Balancing

Mamy macierz $n \times m$ wypełnioną wartościami z $\{0, 1\}$.

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

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

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

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

Twierdzenie 2.3.6. Dla losowego $\overline{b}$ (każda współrzędna niezależnie, jednostajnie z $\{-1, 1\}$) zachodzi

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

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

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

Rozdział 3

Rozkład Poissona

3.1 Definicja

Definicja 3.1.1. Mówimy, że zmienna losowa X ma rozkład **Poissona** z parametrem λ jeśli

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

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 3.1.1. Niech X ma rozkład Poissona z parametrem λ. Wtedy

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

Dowód.

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

□

Twierdzenie 3.1.2 (Lemat 5.3 P&C)**.** Jeśli zmienna X ma rozkład Poissona z parametrem λ to

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

Dowód.

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

W przedostatnim przejściu korzystamy z faktu, że $\sum_0^\infty \frac{x^n}{n!} = \exp(x)$ □

Twierdzenie 3.1.3 (Lemat 5.2 P&C)**.** Jeśli X ma rozkład Poissona z parametrem λ_X a Y rozkład Poissona z parametrem λ_Y, a ponadto obie zmienne są niezależne to $X+Y$ ma rozkład Poissona z parametrem $\lambda_X + \lambda_Y$

Dowód. Są dwie ścieżki aby to pokazać – jedna polega na przeliczeniu explicite $P(X+Y=n)$ – zostawiamy to jako ćwiczenie dla Czytelnika.

Pokażemy bardziej elegancki dowód korzystający z funkcji tworzących momentów.

Ponieważ X i Y są niezależne to

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

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

Twierdzenie 3.1.4. Niech X ma rozkład Poissona z parametrem λ. Wtedy

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

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

3.2 Granica rozkładu dwumianowego

Twierdzenie 3.2.1. Ustalmy λ. Niech X_n będzie rozkładem dwumianowym z parametrami $n, p = \frac{\lambda}{n}$

Wtedy

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

Dowód.

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &= \frac{(n-k+1)\cdot\dots\cdot n}{n^k} \cdot \frac{n^k}{k!} \cdot p^k \cdot (1-p)^n \cdot (1-p)^{-k} \\ &= \frac{(np)^k}{k!} \cdot (1-p)^n \cdot \frac{(n-k+1)\cdot\dots\cdot n}{n^k} \cdot (1-p)^{-k} \\ &= \frac{\lambda^k}{k!} \cdot \left(1-\frac{\lambda}{n}\right)^n \cdot \frac{(n-k+1)\cdot\dots\cdot n}{n^k} \cdot \left(1-\frac{\lambda}{n}\right)^{-k} \end{aligned}$$

Biorąc granicę dostajemy:

$$\begin{aligned} \lim_{n\to\infty} P(X_n = k) &= \lim_{n\to\infty} \left(\frac{\lambda^k}{k!} \cdot \left(1-\frac{\lambda}{n}\right)^n \cdot \frac{(n-k+1)\cdot\dots\cdot n}{n^k} \cdot \left(\frac{n-\lambda}{n}\right)^{-k} \right) \\ &= \left(\lim_{n\to\infty} \frac{\lambda^k}{k!} \cdot \left(1-\frac{\lambda}{n}\right)^n \right) \cdot \left(\lim_{n\to\infty} \frac{(n-k+1)\cdot\dots\cdot n}{n^k} \cdot \left(\frac{n}{n-\lambda}\right)^k \right) \\ &= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \end{aligned}$$

□

Pokażemy teraz trudniejszy dowód, który jest w książce

Szanujemy swoich Czytelników i udowodnimy prosty lemat, z którego skorzystamy.

Lemat 3.2.1. Dla dowolnego $|x| \leq 1$

$$e^x(1-x^2) \leq 1 + x \leq e^x$$

Dowód. Korzystamy z rozwinięcia Taylora dla e^x

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

1. Zacznijmy od górnej nierówności. W oczywisty sposób

$$1 + x \leq 1 + x + \frac{x^2}{2!} + \dots$$

2. Drugą udowadniamy korzystając z faktu, że

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \leq \sum_{n=0}^{\infty} x^n = \frac{1}{1-x} = \frac{1+x}{1-x^2}$$

□

Lemat 3.2.2. Dla $n \geq 1$ oraz $x \geq -1$ zachodzi

$$(1+x)^n \geq 1 + nx$$

Twierdzenie 3.2.2 (Twierdzenie 5.5 P&C)**.** Ustalmy λ. Niech X_n będzie rozkładem dwumianowym z parametrami $n, p = \frac{\lambda}{n}$

Wtedy

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

Dowód. Korzystamy z obu poprzednich lematów aby dostać kolejne oszacowania górne:

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(1-p)^n}{(1-p)^k} \\ &\leq \frac{n^k}{k!} \cdot p^k \cdot \frac{(e^{-p})^n}{1-pk} \\ &= e^{-pn} \cdot \frac{(np)^k}{k!} \cdot \frac{1}{1-pk} \\ &= e^{-\lambda} \cdot \frac{\lambda^k}{k!} \cdot \frac{1}{1-pk} \end{aligned}$$

Wygląda dość znajomo. Widzimy, że w granicy to wyrażenie zbiega do

$$e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

To teraz oszacowanie dolne

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ &\geq \frac{(n-k+1)^k}{k!} \cdot p^k \cdot (1-p)^{n-k} \\ &\geq \frac{(n-k+1)^k}{k!} \cdot p^k \cdot (1-p)^n \\ &\geq \frac{((n-k+1)p)^k}{k!} \cdot \left(e^{-p} \cdot (1-p^2)\right)^n \\ &\geq e^{-pn} \cdot \frac{(np - p(k+1))^k}{k!} \cdot (1-np^2) \\ &= e^{-\lambda} \cdot \frac{(\lambda - p(k+1))^k}{k!} \cdot (1-p\lambda)^2 \end{aligned}$$

Tutaj ponownie widzimy, że skoro $p \to 0$ to to wyrażenie zbiega do

$$e^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

Z twierdzenia o trzech ciągach wychodzi że granica rozkładu dwumianowego to Poisson, niesamowite. □

3.3 Ograniczenia Chernoffa

Twierdzenie 3.3.1. 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

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

gdzie podstawiliśmy $t = \ln\left(\frac{x}{\mu}\right) > 0$. Drugi punkt robi się identycznie, wtedy mamy $\ln\left(\frac{x}{\mu}\right) < 0$. Trzeci i czwarty punkt są po prostu podstawieniem do poprzednich. □

3.4 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 3.4.1 (Twierdzenie 5.6 P&C)**.** Niech

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

opisują (faktyczne) rozmieszczenie k kul w n urnach.

Ponadto, niech

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

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

Wtedy

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

Dowód. Policzmy najpierw lewą stronę równości

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

Policzmy teraz prawą stronę

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

Korzystamy z faktu, że nasze zmienne są niezależne, oraz suma n Poissonów z parametrem

$\lambda = \frac{m}{n}$ ma rozkład Poissona z parametrem m

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

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

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

Dowód.

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

□

3.4.1 Kule i urny

W twierdzeniu 1.8.1 pokazaliśmy, że górne ograniczenie na liczbę kul w najcięższej urnie to z dużym prawdopodobieństwem $O\big(\frac{\ln n}{\ln \ln n}\big)$.

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

Twierdzenie 3.4.3 (Lemat 5.12 P&C)**.** 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 więcej niż 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

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

Nie wiem skąd wynika to oszacowanie, ale najwyraźniej tak jest.

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 3.4.2 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 $\ln\ln n \in o\big(\frac{\ln n}{\ln\ln n}\big)$. Możemy więc zamienić $\frac{\ln n}{\ln n\ln}$ na $\ln\ln n + \ln 2e$

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

czyli nasze M działa. Uff. □

3.4.2 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 3.4.4. Niech X będzie liczbą zebranych kuponów aż do zebrania wszystkich n rodzajów. Wtedy dla dowolnej stałej c

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

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

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

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

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

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

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

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

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

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

   W takim razie z nierówności Czebyszewa

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

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

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

A to jest dokładnie to co chcieliśmy pokazać. □

Rozdział 4

Łańcuchy Markowa

4.1 Definicje

Definicja 4.1.1. 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 i zapisujemy X_t
Mówimy, że proces jest **skończony** jeśli zmienne X_t przyjmują skończenie wiele wartości.

Definicja 4.1.2. Procesem Markowa (czasu homogenicznego) nazywamy taki proces stochastyczny $X_0, X_1, X_2, \ldots$ w którym dla dowolnego t zachodzi

$$P(X_t = a_t \mid X_{t-1} = a_{t-1}, X_{t-2} = a_{t-2}, \ldots X_0 = a_0) = P(X_t = a_t \mid X_{t-1} = a_{t-1})$$

Innymi słowy 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 $X_{t-2}, X_{t-3}, \ldots$ – jest, ale cała ta zależność jest zawarta w zależności od stanu X_{t-1}.

Definicja 4.1.3. Łańcuch Markowa jest **czasu homogenicznego** jeśli $P(X_t = a_t \mid X_{t-1} = a_{t-1} \wedge t = t_0) = P(X_t = a_t \mid X_{t-1} = a_{t-1})$

Mniej formalnie - na rozwój wydarzeń ma jedynie wpływ stan łańcucha, a nie czas w którym ten stan ma miejsce.

Definicja 4.1.4. Macierzą przejścia nazywamy macierz **P** taką, że

$$P_{i,j} = P(X_t = j \mid X_{t-1} = i)$$

oraz

$$\forall_i : \sum_j P_{i,j} = 1$$

Definicja 4.1.5. Rozkładem stacjonarnym nazywamy wektor $\bar{\pi}$ taki, że $\bar{\pi} = \bar{\pi}\mathbf{P}$ 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.

Ćwiczenie 4.1.1. Niech dana będzie macierz Q wymiaru $n \times n$, taka, że suma wartości w każdym wierszu wynosi 1. Rozważmy łańcuch Markowa na n stanach zadany tą macierzą.

1. Czy ma on rozkład stacjonarny?
2. Co jeśli wszystkie wartości macierzy są dodatnie?
3. Co jeśli dodatkowo suma wartości w każdej kolumnie wynosi 1?

Dowód. 1. Nie. Łańcuch zadany macierzą

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

nie ma jednoznacznego rozkładu stacjonarnego.

2. Tak
3. Tak, a ponadto $\bar{\pi} = \left[\frac{1}{n}, \ldots, \frac{1}{n}\right]$

□

4.2 Klasyfikacja stanów

Definicja 4.2.1. Stan j jest **osiągalny** ze stanu i jeśli istnieje $n \geq 0$ takie, że $(P^n)_{i,j} > 0$.

Definicja 4.2.2. Stany i oraz j są **skomunikowane** jeśli i jest osiągalne z j oraz j jest osiągalne z i. Zapisujemy $i \leftrightarrow j$.

Definicja 4.2.3. Łańcuch jest **nieredukowalny (nieprzywiedlny)** jeśli wszystkie stany są parami skomunikowane.

Definicja 4.2.4. Definiujemy **prawdopodobieństwo pierwszego spotkania w zadanym momencie** $r_{i,j}^t$ jako

$$r_{i,j}^t = P(X_t = j \wedge X_{t-1} \neq j \dots X_1 \neq j \mid X_0 = i)$$

Definicja 4.2.5. Definiujemy **prawdopodobieństwo pierwszego spotkania** $r_{i,j}$ jako

$$r_{i,j} = \sum_{t=1}^{\infty} r_{i,j}^t$$

Definicja 4.2.6. Stan i jest **rekurencyjny (powracający)** jeśli $\sum_{t\geq 1} r_{i,i}^t = 1$ a **chwilowy** jeśli $\sum_{t\geq 1} r_{i,i}^t < 1$.
Mówimy, że łańcuch jest rekurencyjny jeśli każdy jego stan jest rekurencyjny.

Definicja 4.2.7. Definiujemy **oczekiwany czas pierwszego spotkania** $h_{i,j} = \sum_{t\geq 1} t \cdot r_{i,j}^t$

Definicja 4.2.8. Rekurencyjny stan i jest **pozytywnie rekurencyjny** jeśli $h_{i,i} < \infty$, w przeciwnym wypadku jest **null rekurencyjny**.

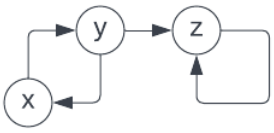
Definicja 4.2.9. Stan i jest **okresowy** jeśli istnieje $\Delta \in \mathbb{N}, \Delta > 1$ takie, że

$$\forall s : P(X_s = i \mid X_0 = i) \neq 0 \Rightarrow \Delta \mid s$$

oraz

$$\exists s : P(X_s = i \mid X_0 = i) \neq 0$$

Łańcuch jest okresowy jeśli posiada co najmniej jeden stan okresowy. Stan lub łańcuch, które nie są okresowe nazywamy nieokresowymi.



Rysunek 4.1: Przykład okresowości – stany x i y są okresowe, stan z nie jest

Definicja 4.2.10. Stan jest **ergodyczny** jeśli jest nieokresowy i pozytywnie rekurencyjny. Łańcuch jest ergodyczny jeśli każdy jego stan jest ergodyczny.

Lemat 4.2.1. 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 – $(P^0)_{i,j} = 1$
2. $i \leftrightarrow j \implies j \leftrightarrow i$
   Koniunkcja jest przemienna, możemy zatem zamienić kolejność warunków w definicji.
3. $i \leftrightarrow j \land j \leftrightarrow k \implies i \leftrightarrow k$
   Skoro $i \leftrightarrow j$ to mamy n dla którego $(P^n)_{i,j} > 0$.

   Podobnie mamy m dla którego $(P^m)_{j,k} > 0$.

   W takim razie $(P^{n+m}) \geq (P^n)_{i,j} \cdot (P^m)_{j,k} > 0$ zatem k jest osiągalne z i.

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

□

Lemat 4.2.2. W skończonym procesie Markowa:

1. istnieje co najmniej jeden stan rekurencyjny
2. każdy stan rekurencyjny jest pozytywnie rekurencyjny

Dowód.

1. Załóżmy nie wprost, że wszystkie stany są chwilowe.

   Niech $Y_1, \ldots, Y_n$ będą zmiennymi losowymi liczącymi liczbę powrotów do każdego ze stanów.

   Zmienne te mają rozkład geometryczny – jeśli wychodzimy z i-tego stanu to z prawdopodobieństwem $p = \sum_{i=1}^{\infty} r_{i,i}^t$ wracamy kiedyś do i (co liczmy jako porażkę) a z prawdopodobieństwem $1 - p$ nigdy już nie wracamy do tego stanu (co liczymy jako sukces).

   Ponieważ z definicji $p < 1$ to $1 - p > 0$ zatem

   $$\forall_i : \mathbb{E}[Y_i] \in \mathbb{R}$$

   Z drugiej jednak strony łańcuch trwa nieskończenie długo, czyli

   $$\infty = \mathbb{E}\left[\sum_i Y_i\right] = \sum_i \mathbb{E}[Y_i]$$

co nam daje sprzeczność.

2.

□

Z lematu tego wysuwamy poniższy wniosek:

Lemat 4.2.3. Skończony, nieredukowalny, nieokresowy łańcuch Markowa jest ergodyczny

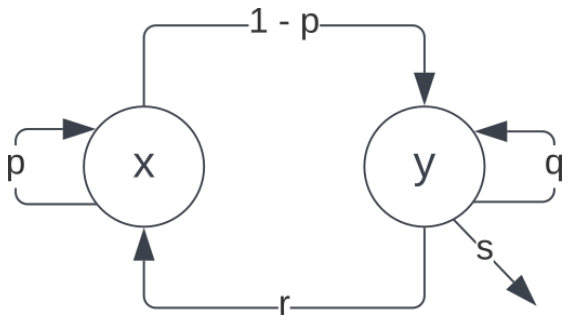
Lemat 4.2.4. Jeśli stan x jest rekurencyjny to wszystkie stany w tej samej klasie skomunikowania również są rekurencyjne.

Dowód. Wybierzmy dowolny stan y w tej samej klasie skomunikowania co x. Pokażemy, że jest on rekurencyjny.

Rozważmy zachowanie naszego łańcucha na stanach x oraz y. Możliwe jest kilka zdarzeń:

1. Wychodząc ze stanu x wracamy do x zanim napotkamy y – z prawdopodobieństwem p
2. Wychodząc ze stanu x zanim wrócimy do x to napotykamy y
   Ponieważ x jest rekurencyjny to to zdarzenie ma szansę $1 - p$
3. Wychodząc ze stanu y wracamy do y zanim napotkamy x – z prawdopodobieństwem q
4. Wychodząc ze stanu y napotkamy x zanim wrócimy do y – z prawdopodobieństwem r
5. Wychodząc ze stanu y nigdy nie wracamy ani do x ani do y – z prawdopodobieństwem s

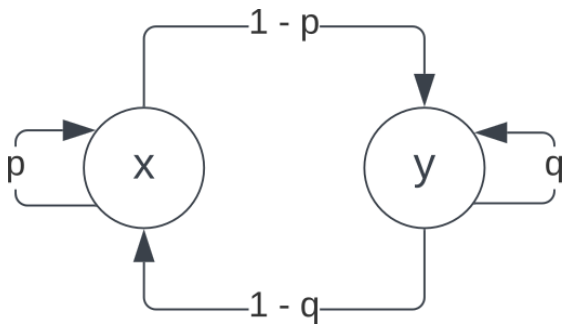
Oczywiście $q + r + s = 1$ bo zdarzenia te wyczerpują wszystkie możliwe zachowania łańcucha. Ponadto ze skomunikowania mamy $1 - p > 0, r > 0$ Poniżej przedstawiamy graficzną reprezentację opisanych zdarzeń.



Rysunek 4.2: Ilustracja przejść między możliwymi sytuacjami

Jak się dobrze przyjrzymy to dojdziemy do wniosku, że $s = 0$. Dlaczego ? Bo inaczej z prawdopodobieństwem co najmniej $(1 - p) \cdot s$ wychodząc z x nigdy już do niego nie wrócimy, co jest sprzeczne z założeniem że jest on rekurencyjny.

W takim razie $s = 0$, a co za tym idzie $r = 1 - q$, możemy zatem uprościć nieco nasz rysunek:



Rysunek 4.3: Ilustracja przejść między możliwymi sytuacjami po uproszczeniu

Aby pokazać, że y jest rekurencyjny pokażemy, że prawdopodobieństwo na to, że wychodząc z y od pewnego momentu nigdy już nie wrócimy do y jest zerowe. Do y nigdy nie wracamy, jeśli po skończonej liczbie kroków trafiamy do x a następnie nigdy już nie wracamy do y.

Innymi słowy przechodzimy nieskończenie wiele razy po pętli $x \to x$ za każdym razem z prawdopodobieństem p, a szansa na takie zdarzenie wynosi $\lim_{n\to\infty} p^n = 0$

□

4.3 Spacery losowe

Definicja 4.3.1. Spacerem losowym na nieskierowanym grafie G nazywamy łańcuch Markowa, którego stany odpowiadają wierzchołkom grafu. Prawdopodobieństwo przejścia ze stanu v do stanu u wynosi $\frac{1}{\deg(v)}$ gdy $(v,u) \in \mathbb{E}$ i 0 w przeciwnym przypadku.

Twierdzenie 4.3.1 (Lemat 7.12 P&C)**.** Spacer losowy na grafie G jest nieokresowy wtedy i tylko wtedy gdy G nie jest dwudzielny

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

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

Twierdzenie 4.3.2 (Lemat 7.13 P&C)**.** Spacer losowy na spójnym, niedwudzielnym grafie G posiada rozkład stacjonarny $\bar{\pi}$ w którym $\pi_v = \frac{\deg(v)}{2|E|}$

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

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

a więc faktycznie jest to rozkład. Mamy też

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

gdzie drugie przejście to zastosowanie określenia macierzy P dla spaceru. □

Definicja 4.3.2. Czas pokrycia grafu G to chwila (indeks łańcucha Markowa), w której spacer odwiedził już każdy wierzchołek. Taką zmienną losową oznaczamy C_G.

Lemat 4.3.1. Dla każdej krawędzi uv w grafie G zachodzi $\mathbb{E}[r_{uv}] + \mathbb{E}[r_{vu}] \leq 2\mathbb{E}(G)$.

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

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

$\sum_{uv\in\mathbb{E}(D)} \frac{1}{2\mathbb{E}(G)} = 1$, więc jest to rozkład. Mamy też

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

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

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

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

□

Twierdzenie 4.3.3 (Twierdzenie 7.15 P&C)**.** Czas pokrycia grafu $G = (V, E)$ jest ograniczony od góry przez $2|E|\left(|V| - 1\right)$.

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

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

□

4.3.1 Ćwiczenia

Ćwiczenie 4.3.1. Mysz i kot niezależnie od siebie w sposób losowy biegają po spójnym, nieskierowanym, niedwudzielnym grafie G. Pokaż, że oczekiwany czas spotkania jest $O(m^2 n)$ gdzie n to liczba wierzchołków, a m liczbą krawędzi.

Rozważ graf stanów G' złożony z par wierzchołków.

Dowód. Wyznaczamy ścieżkę złożoną z $O(n)$ wierzchołków w grafie G' takich, że wierzchołek początkowy to wyjściowe pozycje myszy oraz kota, a wierzchołek końcowy to stan, w którym oba zwierzęta są w tym samym wierzchołku. Oczekiwany czas przejścia między kolejnymi stanami jest ograniczony przez $2|E(G')|$ zatem oczekiwany czas przejścia to $O(n \cdot |E(G')|)$

Pozostaje pokazać, że $|E(G')| \leq m^2$ oraz jak konstruujemy tę ścieżkę.

Zacznijmy od pokazanie, że $|E(G')| \leq m^2$. Rozważmy krawędź $((a,b),(a',b'))$ w grafie G'. Widzimy, że odpowiada ona istnieniu pary krawędzi $(a,a'),(b,b')$ w grafie G Ponieważ każda krawędź z G' jest tworzona przez unikatową parę krawędzi z G to $|E(G')| \leq m^2$.

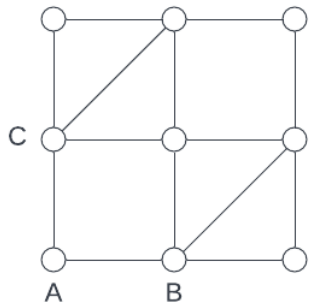
Jak natomiast konstruujemy naszą ścieżkę? Korzystamy z faktu, że graf nie jest dwudzielny i wybieramy w nim dowolny cykl nieparzysty. Prowadzimy mysz i kota najkrótszymi ścieżkami do momentu aż oba wylądują na cyklu. Jeśli któreś zwierzę znajdzie się tam wcześniej to po prostu każemy mu chodzić dookoła. Pesymistycznie musimy przejść po n stanach w ten sposób.

Następnie kierujemy zwierzęta w przeciwnych kierunkach i wykonujemy $2n$ kroków po cyklu. Ponieważ cykl ten jest nieparzysty i ma długość co najwyżej n to na pewno po drodze będzie moment w którym oba zwierzęta są w jednym wierzchołku.

Łącznie zatem wykonujemy $O(n)$ kroków, a ograniczenie na przejście do kolejnego kroku to $O(|E(G')|) \subseteq O(m^2)$ zatem w $O(m^2 n)$ zwierzęta się spotkają.

□

Ćwiczenie 4.3.2. Rozważ spacer losowy po grafie przedstawionym na rysunku (4.4). Spacer startuje z wierzchołka A. Rozważ stowarzyszony z nim łańcuch Markowa $(X_n)_{n\in\mathbb{N}}$, gdzie stanami są wierzchołki grafu oraz $X_0 = A$.



Rysunek 4.4: Śmieszny graf G

(i) Wyznacz $\lim_{n\to\infty} \mathrm{P}(X_n = A)$ oraz $\lim_{n\to\infty} \mathrm{P}(X_n = B)$.

(ii) Jaka jest oczekiwana liczba odwiedzeń wierzchołka B przed pierwszym powrotem do A?

Dowód. Zauważmy, że spacer po naszym grafie G ma rozkład stacjonarny. Dlaczego? Możemy odhaczyć wszystkie wymagania:

- skończony nieskierowany? — jak najbardziej
- spójny? — również widać na pierwszy rzut oka
- niedwudzielny? — trójkąty, aż miło patrzeć

(i) Wiemy, że z własności rozkładu w skończonym, nieredukowalnym i ergodycznym łańcuchu Markowa

$$\pi_v = \lim_{t\to\infty} P^t_{Av} = \lim_{t\to\infty} \mathrm{P}(X_t = v)$$

Czyli dokładnie to, czego szukamy. Spacerujemy po grafie, więc

$$\pi_v = \frac{\deg(v)}{2|E|}$$

Liczymy więc wszystkie krawędzie (jest ich 14) oraz stopnie interesujących nas wierzchołków i mamy odpowiedź:

$$\lim_{t\to\infty} \mathrm{P}(X_t = A) = \pi_A = \frac{\deg(A)}{2|E|} = \frac{2}{2\cdot 14} = \frac{1}{14}$$
$$\lim_{t\to\infty} \mathrm{P}(X_t = B) = \pi_B = \frac{\deg(B)}{2|E|} = \frac{4}{2\cdot 14} = \frac{1}{7}$$

(ii) Ułatwimy sobie trochę życie i policzymy oczekiwaną liczbę odwiedzeń B *lub* C przed pierwszym powrotem do A. Odpowiedź do zadania będzie równa połowie z tego wyniku, ponieważ sytuacja jest symetryczna.

Zauważmy najpierw, że, będąc w B lub C, możemy dojść do A tylko na jeden sposób — przechodząc bezpośrednio odpowiednią krawędzią (z prawdopodobieństwem $\frac{1}{4}$). Jest to cenna obserwacja, ponieważ gwarantuje nam, że nie przejdziemy do A jakąś okrężną drogą. Możemy też być pewni, że natychmiast po wyjściu z A znajdziemy się w B lub C i dzięki temu pracować pod założeniem, że będziemy w którymś z nich przynajmniej raz przed powrotem.

Ponieważ graf jest skończony i spójny, to wszystkie stany są rekurencyjne. Możemy więc powiedzieć, że mamy rozkład geometryczny z parametrem $\frac{1}{4}$: za każdym razem, gdy jesteśmy w B lub C mamy prawdopodobieństwo sukcesu — powrotu do A — równe $\frac{1}{4}$, a porażka, ze względu na rekurencyjność, ostatecznie zaprowadzi nas z powrotem do stanu „B lub C" (być może długą drogą, ale w łańcuchu Markowa trasa, którą obraliśmy, nie ma znaczenia) i postawi przed takim samym eksperymentem. Zatem tak naprawdę oczekujemy po prostu na sukces, który wypada z szansą $\frac{1}{4}$. Oczekiwana liczba prób wynosi zatem 4, a odpowiedź do naszego zadania 2.

□

4.4 Ruina gracza

4.5 2-SAT

2-SATa nikomu nie trzeba przedstawiać – mamy n zmiennych i k klauzul postaci $a \vee b$, gdzie $a = x_i$ albo $a = \bar{x}_i$.

Przykładem instancji problemu 2-SAT jest na przykład taka formuła:

$$(x_1 \vee x_3) \wedge (\neg x_1 \vee x_2) \wedge (x_3 \vee \neg x_1)$$

Znamy algorytm rozwiązujący ten problem w czasie $O(n + k)$ za pomocą silnie spójnych składowych, ale tutaj pokażemy **wolniejszy** algorytm.

4.5.1 Algorytm

Niech $\lambda \in \mathbb{N}$ będzie parametrem algorytmu (stałą czasu działania).

1. Wylosuj dowolne wartościowanie zmniennych $x_1, \ldots, x_n$
2. Powtarzaj maksymalnie $2\lambda n^2$ razy lub do znalezienia rozwiązania
   (a) Wylosuj niespełnioną klazulę
   (b) Wylosuj literał z tej klauzuli i odwróć wartość zmiennej tego literału
3. Jeśli mamy wartościowanie to je zwracamy
4. W przeciwnym razie orzekamy, że formuła jest niespełnialna

4.5.2 Własności algorytmu

Twierdzenie 4.5.1 (Lemat 7.1 P&C)**.** Jeśli dana jest formuła spełnialna, oraz pozwalamy działać algorytmowi dowolnie długo to oczekiwana liczba kroków wynosi co najwyżej n^2

Dowód. Będziemy modelować zachowanie algorytmu jako łańcuch Markowa (a jakże). Wybierzmy sobie dowolne wartościowanie S, które spełnia formułę. Nazwijmy wartościowanie stworzone przez algorytm w i-tym kroku przez A_i

Niech X_i oznacza liczbę zmiennych, które mają to samo wartościowanie w S oraz w A_i

Mamy zatem taki proces $X_0, X_1, \ldots$, który niestety nie jest łańcuchem Markowa, bo tracimy informacje o tym, które zmienne mają jakie wartościowanie, a prawdopodobieństwo przejścia z X_i do X_{i+1} jest zadane wartościowaniem A_i, które nie jest częścią łańcucha.

Zrobimy zatem sztuczkę i rozważymy proces $Y_0, Y_1, \ldots$, który będzie łańcuchem Markowa i jednocześnie będzie pesymistyczną sytuacją naszego procesu.

Zauważmy, że jeśli wybieramy klauzulę, która nie jest spełniona, to wartościowania A_i oraz S wartościują którąś ze zmiennych (być może obie) inaczej. W takim razie

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

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

Zatem w pesymistycznej sytuacji, którą modeluje nasz Y_i mamy:

$$Y_0 = X_0$$

$$P(Y_{i+1} = j + 1 \mid Y_i = j) = \frac{1}{2}$$

$$P(Y_{i+1} = j - 1 \mid Y_i = j) = \frac{1}{2}$$

Przy czym (z uwagi na to, że mamy tylko jedną opcję)

$$P(Y_{i+1} = 1 \mid Y_i = 0) = 1$$

Niech Z_i oznacza liczbę kroków potrzebną do pierwszego dotarcia do stanu n zaczynając w stanie i. Mamy

$$h_i = \mathbb{E}[Z_i] = \mathbb{E}\left[1 + \frac{Z_{i-1}}{2} + \frac{Z_{i+1}}{2}\right]$$

Dostajemy zatem układ $n + 1$ równań

$$\begin{cases} h_0 = h_1 + 1 \\ h_i = 1 + \frac{1}{2}(h_{i-1} + h_{i+1}) \\ h_n = 0 \end{cases}$$

Przekształcamy środkową zależność do postaci

$$h_{i+1} = 2h_i - h_{i-1} - 2$$

Rozwiązując indukcyjnie dostajemy

$$\begin{aligned} h_1 &= h_0 - 1 \\ h_2 &= 2h_1 - h_0 - 2 = h_1 - 3 = h_0 - 1 - 3 \\ h_3 &= 2h_2 - h_1 - 2 = h_2 - 5 = h_0 - 1 - 3 - 5 \\ h_i &= h_0 - i^2 \end{aligned}$$

W takim razie

$$h_n = 0 = h_0 - n^2$$

czyli

$$h_0 = n^2$$

Start w stanie 0 jest najbardziej pesymistyczny a pokazaliśmy, że w oczekiwaniu po n^2 krokach znajdziemy wartościowanie, co należało pokazać. □

Twierdzenie 4.5.2 (Lemat 7.2 P&C)**.** Jeśli formuła jest spełnialna, to algorytm 4.5.1 myli się z prawdopodobieństwem $2^{-\lambda}$

Dowód. Załóżmy na chwilę, że pozwalamy algorytmowi działać w nieskończoność.

Dzielimy wykonanie algorytmu na bloki długości $2n^2$. Niech Z_i oznacza liczbę kroków wykonaną od początku i-tego bloku (zakładając, że nie znaleźliśmy wcześniej wartościowania).

Pokazaliśmy przed chwilą, że $\mathbb{E}[Z_i] \leq n^2$ zatem z nierówności Markowa

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

W takim razie prawdopodobieństwo, że po wykonaniu λ bloków nie znaleźliśmy wartościowania wynosi

$$\left(\frac{1}{2}\right)^{\lambda}$$

□

4.6 3-SAT

2-sata rozwiązaliśmy wolniej, ale 3-sat jest już problemem NP-zupełnym, więc jest większe pole do usprawnień. Z drugiej strony, raczej nie spodziewamy się lepszego wyniku niż wykładniczego, nawet w randomizowanym przypadku – świat byłby wtedy zbyt idealny.

Pokażemy dwa algorytmy – pierwszy będzie działał tak jak 2-sat, drugi będzie z lekkim twistem.

4.6.1 Naiwny algorytm

Niech $\lambda \in \mathbb{N}$ będzie parametrem algorytmu (stałą czasu działania).

1. Wylosuj dowolne wartościowanie zmniennych $x_1, \ldots, x_n$
2. Powtarzaj maksymalnie λ razy lub do znalezienia rozwiązania
   (a) Wylosuj niespełnioną klazulę
   (b) Wylosuj literał z tej klauzuli i odwróć wartość zmiennej tego literału
3. Jeśli mamy wartościowanie to je zwracamy
4. W przeciwnym razie orzekamy, że formuła jest niespełnialna

4.6.2 Czas działania naiwnego algorytmu

Twierdzenie 4.6.1. Jeśli dana jest spełnialna formuła to algorytm 4.6.1 wykonuje w oczekiwaniu $O(2^n)$ kroków zanim ją znajdzie.

Dowód. Sytuacja ma się dość podobnie jak w 2-sacie – mamy ciąg $X_0, X_1, \ldots$ zmiennych liczących zgodność z ustalonym wartościowaniem S. Tylko oszacowania są nieco gorsze bo spośród trzech literałów, dwa mogą się zgadzać, a tylko jeden nie.

$$P(X_{i+1} = j + 1 \mid X_i = j) \geq \frac{1}{3}$$

$$P(X_{i+1} = j - 1 \mid X_i = j) \leq \frac{2}{3}$$

Zatem pesymistyczny przypadek modelujemy ciągiem $Y_0, Y_1, \ldots$

$$Y_0 = X_0$$

$$P(Y_{i+1} = 1 \mid Y_i = 0) = 1$$

$$P(Y_{i+1} = j + 1 \mid Y_i = j) = \frac{1}{3}$$

$$P(Y_{i+1} = j - 1 \mid Y_i = j) = \frac{2}{3}$$

Ponownie, niech h_i oznacza oczekiwaną liczbę kroków zaczynając w stanie i. Dostajemy układ równań

$$\begin{cases} h_0 = h_1 + 1 \\ h_i = 1 + \frac{2}{3}h_{i-1} + \frac{1}{3}h_{i+1} \\ h_n = 0 \end{cases}$$

Przekształcamy środek do

$$h_{i+1} = 3h_i - 2h_{i-1} - 3$$

a następnie rozwiązujemy indukcyjnie

$$\begin{aligned} h_1 &= h_0 - 1 \\ h_2 &= 3h_1 - 2h_0 - 3 = h_1 - 5 = h_1 - (8 - 3) \\ h_3 &= 3h_2 - 2h_1 - 3 = h_2 - 2 \cdot (8 - 3) - 3 = h_2 - (16 - 3) \end{aligned}$$

Widzimy, że wyjdzie

$$\begin{aligned} h_{i+1} &= h_i - (2^{i+2} - 3) \\ h_i &= h_{i+1} + 2^{i+1} - 3 \end{aligned}$$

W takim razie mamy

$$\begin{aligned} h_n &= 0 \\ h_{n-1} &= h_n + 2^{n+1} - 3 = 2^{n+1} - 3 \\ h_{n-2} &= h_{n-1} + 2^n - 3 = 2^{n+1} + 2^n - 2 \cdot 3 \\ h_i &= 2^{n+1} + 2^n + \cdots + 2^{i+2} - 3(n - i) = 2^{n+2} - 2^{i+2} - 3(n - i) \end{aligned}$$

Trochę do bani – równie dobrze można sprawdzić wszystkie wartościowania w $O(n^2 \cdot 2^n)$ □

Naiwny algorytm ma pewną wadę – będzie się kręcił blisko wartościowań, które mają mało wspólnych zmiennych, z prostej przyczyny, że łatwiej popsuć niż poprawić.

Zauważmy jednak, że nic nie stoi na przeszkodzie aby zresetować algorytm tj. wylosować nowe wartościowanie i zacząć szukać od niego. Losowe wartościowanie ma średnio połowę zmiennych ustawionych poprawnie, więc z dużą szansą jest bliżej rozwiązania niż wartościowanie naiwnego algorytmu po długim czasie.

4.6.3 Sprytny algorytm

Niech $\lambda \in \mathbb{N}$ będzie parametrem algorytmu (stałą czasu działania).

1. Powtarzaj maksymalnie λ razy lub do znalezienia rozwiązania

(a) Wylosuj dowolne wartościowanie zmniennych $x_1, \ldots, x_n$

(b) Powtarzaj maksymalnie $3n$ razy lub do znalezienia rozwiązania

i. Wylosuj niespełnioną klazulę

ii. Wylosuj literał z tej klauzuli i odwróć wartość zmiennej tego literału

2. Jeśli mamy wartościowanie to je zwracamy

3. W przeciwnym razie orzekamy, że formuła jest niespełnialna

4.6.4 Czas działania sprytnego algorytmu

Zaczynamy od lematu, który przyjmujemy na wiarę (jego dowód to przykra analiza)

Lemat 4.6.1 (Wzór Stirlinga)**.**

$$\forall n > 0 : n! = \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \cdot (1 \pm o(1))$$

W szczególności

$$\forall n > 0 : \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n \leq n! \leq 2\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n$$

Używamy tego lematu aby pokazać

$$\begin{aligned}
\binom{3i}{i} &= \frac{(3i)!}{i!(2i)!} \\
&\geq \frac{\sqrt{2\pi 3i}}{\left(2\sqrt{2\pi i}\right) \cdot \left(2\sqrt{2\pi 2i}\right)} \cdot \left(\frac{3i}{e}\right)^{3i} \cdot \left(\frac{e}{i}\right)^{i} \cdot \left(\frac{e}{2i}\right)^{2i} \\
&= \frac{\sqrt{3}}{8\sqrt{\pi}\sqrt{i}} \cdot \left(\frac{27}{4}\right)^{i} \\
&= \frac{c}{\sqrt{i}} \cdot \left(\frac{27}{4}\right)^{i}
\end{aligned}$$

Gdzie c jest stałą odsyfiającą wyrażenie.

Twierdzenie 4.6.2. Algorytm 4.6.3 wykonuje w oczekiwaniu $O\left(n^{3/2} \cdot \left(\frac{4}{3}\right)^n\right)$ kroków zanim znajdzie poprawne wartościowanie.

Dowód. Chcemy oszacować od dołu prawdopodobieństwo q, że po co najwyżej 3n krokach, z losowego wartościowania znajdziemy poprawne.

Niech q_i oznacza prawdopodobieństwo, że startując z wartościowania, któremu brakuje dokładnie i zmiennych to po 3n krokach dojdziemy do poprawnego wartościowania.

q_i szacujemy od dołu pesymistycznym przypadkiem, który szacujemy przez możliwe rozłożenia

kroków „w górę” i „w dół”

$$\begin{aligned} q_i &\geq \max_{k=0,\ldots,i} \binom{i+2k}{k} \cdot \left(\frac{2}{3}\right)^k \cdot \left(\frac{1}{3}\right)^{i+k} \\ &\geq \binom{3i}{i} \cdot \left(\frac{2}{3}\right)^i \cdot \left(\frac{1}{3}\right)^{2i} \\ &\geq \frac{c}{\sqrt{i}} \cdot \left(\frac{27}{4}\right)^i \cdot \left(\frac{2}{3}\right)^i \cdot \left(\frac{1}{3}\right)^{2i} \\ &= \frac{c}{\sqrt{i}} \cdot 2^{-i} \end{aligned}$$

Niech $\mathcal{E}_i$ oznacza zdarzenie, w którym wylosowane wartościowanie jest **różne** od zadanego S na dokładnie i pozycjach.

Mamy zatem

$$\begin{aligned} q &= \sum_{i=0}^{n} P(\mathcal{E}_i) \cdot q_i \\ &= 2^{-n} + \sum_{i=1}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n \cdot q_i \\ &\geq \sum_{i=1}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n \cdot \frac{c}{\sqrt{i}} \cdot 2^{-i} \\ &= \sum_{i=1}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^{n+i} \cdot \frac{c}{\sqrt{i}} \\ &\geq \frac{c}{\sqrt{n}} \cdot \left(\frac{1}{2}\right)^n \sum_{i=1}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^i \cdot 1^{n-i} \\ &= \frac{c}{\sqrt{n}} \cdot \left(\frac{1}{2}\right)^n \cdot \left(\frac{3}{2}\right)^n \\ &= \frac{c}{\sqrt{n}} \cdot \left(\frac{3}{4}\right)^n \end{aligned}$$

Najgorsze za nami.

q to jest prawdopodobieństwo, że w konkretnym bloku (dla konkretnego wartościowania startowego) nam się udało. Bloki są niezależne, więc całość ma rozkład geometryczny z parametrem q, zatem w oczekiwaniu musimy przejść przez $\frac{1}{q} = O\left(\sqrt{n} \cdot \left(\frac{4}{3}\right)^n\right)$ bloków, a każdy blok zajmuje czas $O(n)$ co daje tezę. □

Rozdział 5

Prawdopodobieństwo ciągłe

5.1 Definicje

Definicja 5.1.1. Zmienna losowa X jest ciągła, jeśli istnieje funkcja $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$ taka, że

$$\forall_{B \in \mathcal{B}(\mathbb{R})} \; P(X \in B) = \int_B f(x)\,\mathrm{d}x.$$

Taką funkcję nazywamy funkcją gęstości lub gęstością zmiennej X.

Własności funkcji gęstości:

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

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

Definicja 5.1.2. Dystrybuanta zmiennej losowej X to funkcja $F(x) = P(X \leq x)$. Dla zmiennej ciągłej jest $F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$, a więc $f(x) = F'(x)$.

Definicja 5.1.3. Wartość oczekiwana ciągłej zmiennej X to

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

Lemat 5.1.1 (Lemat 8.1 P&C)**.** Jeśli zmienna losowa X przyjmuje wartości nieujemne to

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

Dowód. Niech f będzie gęstością X. Mamy

$$\int_0^\infty P(X \geq x)\,\mathrm{d}x = \int_{x=0}^\infty \int_{y=x}^\infty f(y)\,\mathrm{d}y\,\mathrm{d}x = \int_{y=0}^\infty \int_{x=0}^y f(y)\,\mathrm{d}x\,\mathrm{d}y$$
$$= \int_{y=0}^\infty f(y) \int_{x=0}^y \mathrm{d}x\,\mathrm{d}y = \int_{y=0}^\infty yf(y)\,\mathrm{d}y = \int_{-\infty}^\infty yf(y)\,\mathrm{d}y,$$

gdzie ostatnie przejście wynika z nieujemności X. □

Definicja 5.1.4. Wspólna dystrybuanta zmiennych losowych losowych X, Y to

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

Definicja 5.1.5. Wspólna gęstość ciągłych zmiennych losowych X, Y to funkcja f taka, że

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

a więc

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

Definicja 5.1.6. Dla dwóch zmiennych X, Y o zadanej wspólnej dystrybuancie $F(x,y)$ brzegowa dystrybuanta zmiennej X to funkcja

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

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

Definicja 5.1.7. Zmienne X, Y są niezależne, jeśli

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

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 5.1.8. Prawdopodobieństwo warunkowe definiujemy jako całkę

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

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

Definicja 5.1.9. Warunkowa wartość oczekiwana to całka

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

5.2 Rozkład jednostajny

5.2.1 Definicja

Definicja 5.2.1. Mówimy, że zmienna losowa X ma **rozkład jednostajny** na przedziale $[a, b]$ jeśli dystrybuanta tej zmiennej zadana jest przez funkcję

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

Łatwo zauważyć, że gęstość takiej zmiennej wynosi $\frac{1}{b-a}$ w $[a, b]$ oraz 0 wszędzie indziej.

Twierdzenie 5.2.1. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy

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

Twierdzenie 5.2.2. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy

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

Twierdzenie 5.2.3. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy

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

Twierdzenie 5.2.4. Niech X ma rozkład jednostajny na przedziale $[a, b]$. Wtedy dla dowolnych $a \leq c \leq d \leq b$

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

Dowód.

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

□

Twierdzenie 5.2.5. Niech $X_1, \ldots, X_n$ będą niezależne i wszystkie mają rozkład jednostajny na $[0, 1]$. Ponadto, niech $Y_1, \ldots, Y_n$ 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 $P_0, \ldots, P_n$. 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 $\frac{k}{n+1}$. □

5.2.2 Ćwiczenia

Ćwiczenie 5.2.1. Łamiemy kij długości jednego metra w dwóch miejscach wybranych niezależnie i jednostajnie.

Podaj oczekiwaną długość najkrótszego, średniego, i najdłuższego kawałka.

Dowód. Niech X i Y oznaczają odległości punktów podziału odpowiednio od lewego i prawego końca.

Jeśli $X + Y \geq 1$ to znaczy, że punkty X, Y są zamienione kolejnością tj. Y jest bliżej lewego końca a X prawego.

Wygodniej nam będzie jednak gdy $X + Y \leq 1$ więc zanim przejdziemy do głównej części pokażmy mini-lemat

Niech A będzie dowolną zmienną losową.

Ponieważ $P(X \leq Y) = P(X + Y \leq 1) = \frac{1}{2}$ to

$$\mathbb{E}[A] = \mathbb{E}[A \mid X \leq Y] \cdot P(X + Y \leq 1) + \mathbb{E}[A \mid X \geq Y] \cdot P(X + Y \geq 1)$$

Zatem

$$\mathbb{E}[A] = \frac{1}{2} \cdot (\mathbb{E}[A \mid X + Y \leq 1] + \mathbb{E}[A \mid X + Y \geq 1])$$

Jeśli wartości przyjmowane przez A są symetryczne (a w naszym przypadku będą) to $\mathbb{E}[A] = \mathbb{E}[A \mid X + Y \leq 1]$.

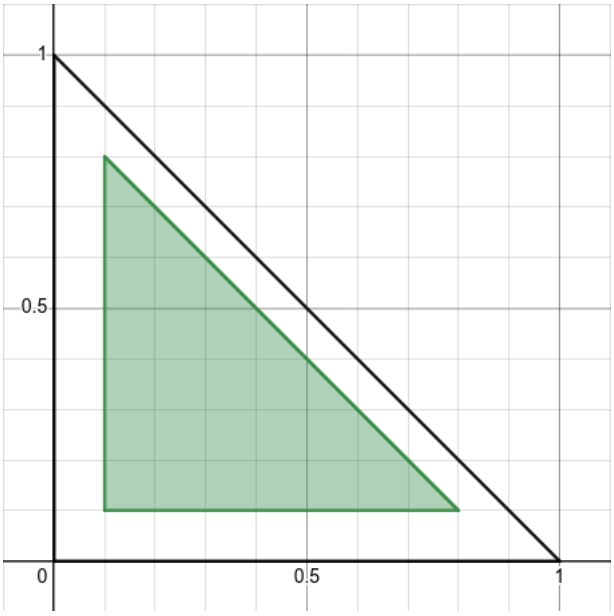
Widzimy, że długości kawałków są niezależne od kolejności punktów a jedynie od ich pozycji, zatem możemy założyć, że $X + Y \leq 1$ Skoro $X + Y \leq 1$ to punkty te dzielą kij na kawałki o długościach $X, Y, 1 - X - Y$

Niech $A = \min(X, Y, 1 - X - Y), B = \text{mid}(X, Y, 1 - X - Y), C = \max(X, Y, 1 - X - Y)$ Z liniowości wartości oczekiwanej mamy $\mathbb{E}[A] + \mathbb{E}[B] + \mathbb{E}[C] = \mathbb{E}[A + B + C] = 1$, wystarczy zatem policzyć $\mathbb{E}[A]$ oraz $\mathbb{E}[C]$.

Aby policzyć $\mathbb{E}[A]$będziemy chcieli skorzystać z lematu 5.1.1.

Zacznijmy więc od policzenia $P(A \geq a)$ dla dowolnego $a \in \mathbb{R}$

Widzimy, że $P(A \geq a) = P(X \geq a \wedge Y \geq a \wedge 1 - X - Y \geq a)$ Możemy zobrazować te zależności geometrycznie:



Rysunek 5.1: Obszar w którym zachodzą nierówności dla $a = 0.1$

Możemy policzyć pole tego trójkąta – wynosi ono dokładnie $\frac{(1-3a)^2}{2}$ Pole dużego trójkąta wynosi $\frac{1}{2}$ zatem $P(A \geq a) = (1-3a)^2 = 9a^2 - 6a + 1$

Widzimy tutaj, że sens ma jedynie rozważanie $a \leq \frac{1}{3}$ zatem finalnie

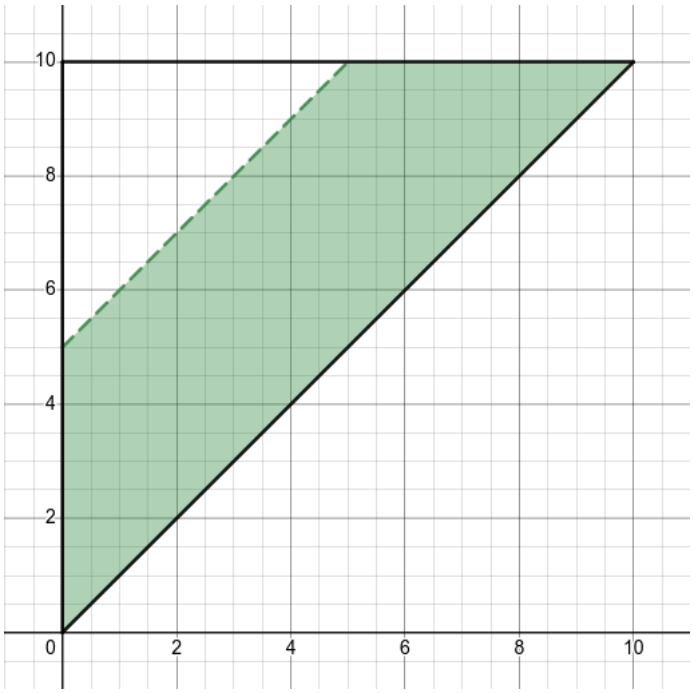
$$\mathbb{E}[A] = \int_0^{\frac{1}{3}} P(A \geq a)\,\mathrm{d}a = \int_0^{\frac{1}{3}} (9a^2 - 6a + 1)\,\mathrm{d}a = \frac{1}{9}$$

Policzenie $\mathbb{E}[C]$ jest bardzo podobne – zamiast brać koniunkcji trzech warunków to bierzemy ich alternatywę i liczymy pola otrzymanych figur. Pozostawiamy to jako ćwiczenie dla Czytelnika

Hint: Rozważ przedziały $a \in \left(0, \frac{1}{3}\right), a \in \left(\frac{1}{3}, \frac{1}{2}\right), a \in \left(\frac{1}{2}, 1\right)$. *Poprawny wynik to* $\frac{11}{18}$ □

Ćwiczenie 5.2.2. Losujemy X oraz Y niezależnie i jednostajnie z przedziału $[0, 10]$. Jakie jest prawdopodobieństwo, że $|X - Y| < 5$?

Dowód. Podobnie jak w ćwiczeniu 5.2.1 będziemy się warunkować pod założeniem, że $X < Y$ argumentując tym, że wartość bezwzględna jest symetryczna.



Rysunek 5.2: Obszar w którym $|X - Y| < 5$

Widać, że dopełnienie tego zdarzenia jest trójkątem o boku 5, a całość jest trójkątem o boku 10, zatem $P(|X - Y| \geq 5) = \frac{1}{4}$, czyli $P(|X - Y| < 5) = \frac{3}{4}$ □

5.3 Rozkład wykładniczy

Definicja 5.3.1. Rozkładem wykładniczym z parametrem λ nazywamy rozkład zadany dystrybuantą

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

Gęstość rozkładu wykładniczego $f = F'$ wynosi $f(x) = \lambda e^{-\lambda x}$

Momenty wynoszą

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x}\,\mathrm{d}x = \frac{1}{\lambda}$$
$$\mathbb{E}\big[X^2\big] = \int_0^\infty x^2\lambda e^{-\lambda x}\,\mathrm{d}x = \frac{2}{\lambda^2}$$
$$\mathrm{Var}[X] = \mathbb{E}\big[X^2\big] - \mathbb{E}[X]^2 = \frac{1}{\lambda^2}$$

Twierdzenie 5.3.1 (Lemat 8.4 P&C)**.** Rozkład wykładniczy jest **bez pamięci** tzn.

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

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

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

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

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

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

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

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

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

Twierdzenie 5.3.3 (MGF)**.** Niech X ma rozkład wykładniczy z parametrem λ. Wtedy dla $t < \lambda$

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

Dowód.

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

□

Twierdzenie 5.3.4 (Lemat 8.5 P&C)**.** Jeśli X_1, X_2 są **niezależnymi** zmiennymi losowymi o rozkładzie wykładniczym o parametrach (odpowiednio) λ_1 i λ_2 to zmienna losowa Y będąca ich minimum jest zmienną o rozkładzie wykładniczym o parametrze $\lambda_1 + \lambda_2$.

Dodatkowo, prawdopodobieństwo że $X_1 = Y$ (w sensie: że to X_1 będzie mniejsze) wynosi $\frac{\lambda_1}{\lambda_1+\lambda_2}$. Analogicznie, dla X_2 prawdopodobieństwo to wyniesie $\frac{\lambda_2}{\lambda_1+\lambda_2}$.

Dowód. Policzmy sobie dystrybuantę zmiennej losowej Y (w nieco śmieszny sposób):

$$1 - F_Y(x) = \mathrm{P}(x \leq Y) = \mathrm{P}(x \leq X_1 \wedge x \leq X_2) = \mathrm{P}(x \leq X_1) \cdot \mathrm{P}(x \leq X_2)$$

$$P(x \leq X_1) \cdot P(x \leq X_2) = (1 - F_{X_1}(x)) \cdot (1 - F_{X_2}(x))$$

Jako, że dla zmiennych X rozkładu wykładniczego $F_X(x) = 1 - e^{-\lambda x}$ to $1 - F_X(x) = e^{-\lambda x}$. Zatem:

$$(1 - F_{X_1}(x)) \cdot (1 - F_{X_2}(x)) = e^{-\lambda_1 x} \cdot e^{-\lambda_2 x} = e^{-(\lambda_1+\lambda_2)x} = 1 - F_Y(x)$$

czyli

$$F_Y(x) = 1 - e^{-(\lambda_1+\lambda_2)x}$$

skąd Y jest zmienną rozkładu wykładniczego z parametrem $\lambda_1 + \lambda_2$. Teraz pozostaje pokazać, że:

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

Zatem liczymy:

$$
\begin{aligned}
\mathrm{P}(X_1 \leq X_2) = \int_{x_2=-\infty}^{\infty} \int_{x_1=-\infty}^{x_2} f_{X_1X_2}(x_1, x_2)\, dx_1 dx_2 = \\
\int_{x_2=-\infty}^{\infty} f_{X_2}(x_2) \int_{x_1=-\infty}^{x_2} f_{X_1}(x_1)\, dx_1 dx_2 = \\
\int_{x_2=0}^{\infty} \lambda_2 e^{-\lambda_2 x_2} \int_{x_1=0}^{x_2} \lambda_1 e^{-\lambda_1 x_1}\, dx_1 dx_2 = \\
\lambda_1\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \int_{x_1=0}^{x_2} e^{-\lambda_1 x_1}\, dx_1 dx_2 = \\
\lambda_1\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \left(\Big|_0^{x_2} \frac{-1}{\lambda_1} e^{-\lambda_1 x_1} \right) dx_2 = \\
\lambda_1\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \left(\frac{-1}{\lambda_1} e^{-\lambda_1 x_2} - \frac{-1}{\lambda_1} e^{0} \right) dx_2 = \\
\lambda_1\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \frac{-1}{\lambda_1} \left(e^{-\lambda_1 x_2} - 1 \right) dx_2 = \\
-\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2} \left(e^{-\lambda_1 x_2} - 1 \right) dx_2 = \\
-\lambda_2 \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2 - \lambda_1 x_2} - e^{-\lambda_2 x_2}\, dx_2 = \\
-\lambda_2 \int_{x_2=0}^{\infty} e^{-x_2(\lambda_2+\lambda_1)} - e^{-\lambda_2 x_2}\, dx_2 = \\
-\lambda_2 \left(\int_{x_2=0}^{\infty} e^{-x_2(\lambda_2+\lambda_1)}\, dx_2 - \int_{x_2=0}^{\infty} e^{-\lambda_2 x_2}\, dx_2 \right) = \\
-\lambda_2 \left(\left(\Big|_{x_2=0}^{\infty} \frac{-1}{\lambda_1+\lambda_2} e^{-x_2(\lambda_1+\lambda_2)} \right) - \left(\Big|_{x_2=0}^{\infty} \frac{-1}{\lambda_2} e^{-\lambda_2 x_2} \right) \right) = \\
-\lambda_2 \left(\frac{1}{\lambda_1+\lambda_2} - \frac{1}{\lambda_2} \right) = \\
-\lambda_2 \left(\frac{\lambda_2}{\lambda_2 \cdot (\lambda_1+\lambda_2)} - \frac{\lambda_1+\lambda_2}{\lambda_2 \cdot (\lambda_1+\lambda_2)} \right) = \\
(-\lambda_2) \cdot \frac{-\lambda_1}{\lambda_2 \cdot (\lambda_1+\lambda_2)} = \\
\frac{\lambda_1}{\lambda_1+\lambda_2}
\end{aligned}
$$

□

5.3.1 Kule i urny z feedbackiem

Jak zwykle, zanim zaczniemy to pokażemy pomocniczy lemat:

Lemat 5.3.1. 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

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

□

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 5.3.5. 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 5.3.1 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. □

5.3.2 Ćwiczenia

Ćwiczenie 5.3.1. Rozważ zmienną o rozkładzie wykładniczym z parametrem λ. Jaki rozkład ma zmienna $\lceil X \rceil$? Czy ten rozkład coś Ci przypomina?

Dowód. Niech F będzie dystrybuantą X, $F(x) = 1 - e^{-\lambda x}$ dla $x \geq 0$.

Wtedy

$$\begin{aligned} P(\lceil X \rceil = n) &= F(n) - F(n-1) \\ &= e^{-\lambda(n-1)} - e^{-\lambda n} \\ &= \left(e^{-\lambda}\right)^{n-1} \cdot \left(1 - e^{-\lambda}\right) \end{aligned}$$

Jeśli przyjmiemy $p = 1 - e^{-\lambda} = F(1)$ to dostaniemy $P(\lceil X \rceil = n) = (1-p)^{n-1} \cdot p$

Rozkład geometryczny z parametrem $F(1)$. Wow.

□

Rozdział 6

Igła Buffona

6.1 Opis Problemu

W problemie Igły Buffona rzucamy „igłą" na podłogę złożoną z „desek". Pytamy się o prawdopodobieństwo przecięcia igły z krawędzią deski.

Bardziej formalnie, mamy płaszczyznę na której są zaznaczone proste, które są parami równoległe. Odstęp między kolejnymi prostymi jest ustalony i wynosi on d - szerokość naszej deski.

Następnie losujemy odcinek o długości l - długość igły, gdzieś na tej płaszczyźnie i pytamy się o prawdopodobieństwo, że ten odcinek przecina jedną z naszych prostych.

6.2 Krótka Igła

Definicja 6.2.1. Krótka igła jest to piękne zdarzenie które zachodzi gdy $l < d$ - długość igły jest mniejsza niż szerokość deski.

Twierdzenie 6.2.1. Prawdopodobieństwo że igła przetnie którąś z krawędzi wynosi $\frac{2l}{4\pi}$.

Dowód. Niech X oznacza odległość od środka igły do najbliższej krawędzi deski. Oznaczmy θ kąt ostry pomiędzy naszą igłą a jedną z wielu równoległych prostych.

Oczywiście obie z tych zmiennych mają rozkład jednostajny.

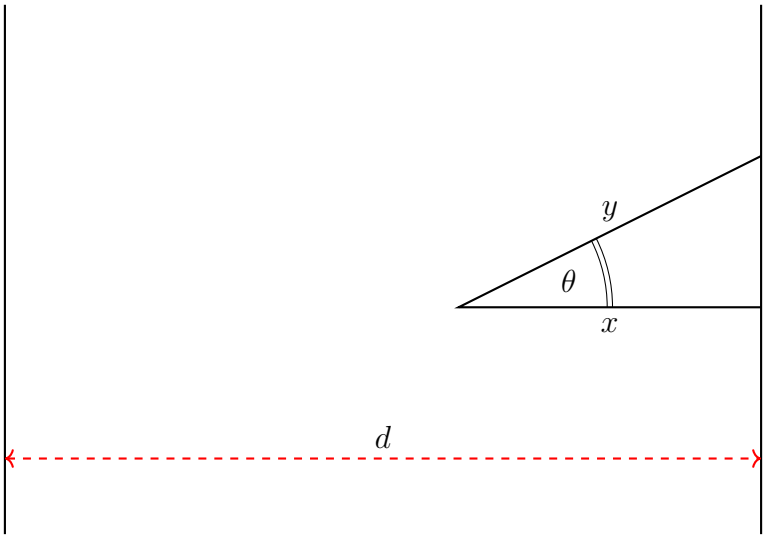
$$f_X(x) = \begin{cases} \frac{2}{d} : & 0 \leq x \leq \frac{d}{2} \\ 0 : & \text{wpw} \end{cases}$$

$$f_\theta(\theta) = \begin{cases} \frac{2}{\pi} : & 0 \leq \theta \leq \frac{\pi}{2} \\ 0 : & \text{wpw} \end{cases}$$

Powyżej korzystamy z faktu, że jest bijekcja między sytuacją gdy środek igły jest bliżej lewej a sytuacją gdy jest bliżej prawej krawędzi (bo rozumowanie jest to samo). Analogicznie zakładamy z kątem „odchylenia" igły. Tu też pojawia się miejsce w którym użyjemy założenia, że $l \leq d$: dzięki temu wiemy, że igła zawsze przetnie maksymalnie jedną krawędź[1].

Zmienne te są w oczywisty sposób niezależne, więc ich wspólny rozkład prawdopodobieństwa będzie ich iloczynem.

$$f_{X\theta}(x,\theta) = \begin{cases} \frac{4}{d\pi} : & 0 \leq x \leq \frac{d}{2}, 0 \leq \theta \leq \frac{\pi}{2} \\ 0 : & \text{wpw} \end{cases}$$



Igła będzie przecinać krawędź, gdy $\frac{l}{2} \geq y$. Jednocześnie:

$$\cos\theta = \frac{x}{y} \iff y = \frac{x}{\cos\theta}$$

Więc aby igła przecinała krawędź musimy mieć:

$$l \geq 2y = \frac{2x}{\cos\theta}$$

Czyli w takim razie musi być tak, że:

$$x \leq \frac{l\cos\theta}{2}$$

[1]Modulo sytuacja gdzie wypadnie dokładnie na środku i kąt będzie zerowy, ale takie zdarzenie ma miarę zero i nie wpłynie na wynik naszego prawdopodobieństwa; tym samym równie dobrze możemy założyć, że $l < d$.

Liczymy zatem nasze piękne prawdopodobieństwo:

$$\begin{aligned}\int_{\theta=0}^{\frac{\pi}{2}}\int_{x=0}^{\frac{l}{2}\cos\theta}\frac{4}{d\pi}\,dx\,d\theta = \frac{4}{d\pi}\int_{\theta=0}^{\frac{\pi}{2}}\int_{x=0}^{\frac{l}{2}\cos\theta}1\,dx\,d\theta = \\ \frac{4}{d\pi}\int_{\theta=0}^{\frac{\pi}{2}}\frac{l}{2}\cos\theta\,d\theta = \\ \frac{2l}{d\pi}\int_{\theta=0}^{\frac{\pi}{2}}\cos\theta\,d\theta = \\ \frac{2l}{d\pi}\Big|_{\theta=0}^{\frac{\pi}{2}}\sin\theta = \\ \frac{2l}{d\pi}\cdot\left(\sin\frac{\pi}{2}-\sin 0\right) = \\ \frac{2l}{d\pi}\end{aligned}$$

□

6.3 Długa Igła

Definicja 6.3.1. Długa igła jest to jeszcze piękniejsze zdarzenie, różni się tym że długa igła może przeciąć dwie krawędzie deski $l \geq d$

Twierdzenie 6.3.1. Prawdopodobieństwo na przecięcie długiej igły z krawędzią wynosi

$$\frac{2l}{d\pi}-\frac{2}{d\pi}\left(\sqrt{l^2-d^2}+d\arcsin\frac{d}{l}\right)+1$$

Dowód. Zobaczmy, że jeśli $l\sin\theta > d$ to wówczas prawdopodobieństwo, że dana igła przetnie krawędź wynosi 1. Natomiast w w przeciwnym przypadku zachowuje się tak samo jak krótka igła. Gdy $0 \leq \theta \leq \arcsin\left(\frac{d}{l}\right)$ to mamy krótką igłę. A zatem:

$$\int_{\theta=0}^{\arcsin\frac{d}{l}}\int_{x=0}^{\frac{l}{2}\sin\theta}\frac{4}{d\pi}\,dx\,d\theta+\int_{\arcsin\frac{d}{l}}^{\frac{\pi}{2}}\frac{2}{\pi}d\theta = \frac{2l}{d\pi}-\frac{2}{d\pi}\left(\sqrt{l^2-d^2}+d\arcsin\frac{d}{l}\right)+1$$

Po trywialnych przeliczeniach.

□

Rozdział 7

Proces Poissona

7.1 Definicja

Definicja 7.1.1. Procesem Poissona z parametrem λ nazywamy proces stochastyczny

$$\{N(t) \mid t \in \mathbb{R}, t \geq 0\}$$

(intuicyjnie: $N(t)$ mówi ile *jakichś* zdarzeń zaszło od momentu rozpoczęcia procesu do jakiejś chwili t.

taki, że:

1. $N(0) = 0$
2. Rozłączne przedziały są niezależne tj. zmienne
   $N(a) - N(b)$ i $N(c) - N(d)$ są niezależne dla $[b, a] \cap [d, c] = \varnothing$
3. Liczba zdarzeń na przedziałach jest stacjonarna tj.
   $N(t + s) - N(s)$ ma taki sam rozkład jak $N(t)$
4. Prawdopodobieństwo jednego zdarzenia w małym przedziale długości t zbiega do λ

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

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

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

Powyższa definicja nie jest jedyną możliwą definicją procesu Poissona. Okazuje się, że możemy skorzystać też z nieco wygodniejszej definicji bez tych dwóch limesów, ale za to korzystającej z rozkładu Poissona.

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

Twierdzenie 7.1.1 (Twierdzenie 8.7 P&C)**.** Niech $\{N(t) \mid t \geq 0\}$ będzie procesem Poissona z parametrem λ. Wtedy dla dowolnego $t \geq 0$ oraz $n \in \mathbb{N}$

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

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

Zauważamy, że z niezależności rozłącznych przedziałów mamy

$$P_0(t+h) = P_0(t) \cdot P_0(h)$$

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

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

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

Mamy zatem równanie różniczkowe

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

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

Całkujemy po t i dostajemy

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

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

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

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

Fakt 7.1.1.

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

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

Korzystając z wyżej wymienionego faktu, mamy:

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

Zrobiliśmy tu bardzo sprytną rzecz – mianowicie rozbiliśmy sumę na trzy części tak, aby przy liczeniu pochodnych wszystko nam się ładnie zwinęło.

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

Znowu dostajemy równanie różniczkowe

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

I z założenia indukcyjnego:

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

Całkujemy obustronnie:

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

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

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

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

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

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

□

7.2 Rozkład czasów między zdarzeniami

Twierdzenie 7.2.1 (Twierdzenie 8.11 P&C)**.** Niech $\{N(t) \mid t \geq 0\}$ będzie procesem stochastycznym takim, że

1. N(0) = 0
2. Czasy między kolejnymi zdarzeniami są niezależne i zadane rozkładem wykładniczym z parametrem λ

Wtedy proces ten jest procesem Poissona z parametrem λ

Dowód. Pokażemy kolejne własności z definicji 7.1.1

1. $N(0) = 0$ z definicji
2. Weźmy dowolne dwa przedziały $[b, a] \cap [d, c] = \varnothing$, przy czym $d > a$

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

   Rozkład X pod warunkiem, że $X > t$ jest wykładniczy z parametrem λ i jest niezależny od tego co się działo przed d, zatem wszystko co zarejestrujemy na przedziale $[d, c]$ jest niezależne od zdarzeń na przedziale $[b, a]$.
3. Aby pokazać, że rozkład $N(s + t) - N(s)$ jest taki sam jak rozkład $N(t)$ zrobimy tę samą sztuczkę co przed chwilą - w chwili s resetujemy ostatnią zmienną. Wszystko teraz dzieje się na przedziale długości t bez żadnych zależności od tego co było wcześniej, zatem rozkład liczby zdarzeń musi być taki sam jak rozkład $N(t)$
4. Niech X_1 będzie zmienną opisującą czas do pierwszego zdarzenia, a X_2 od pierwszego zdarzenia do drugiego. Widzimy, że $P(N(t) = 1) = P(X_1 < t \wedge X_1 + X_2 > t)$

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

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

   Nasze oszacowania będą wyglądały następująco:

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

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

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

Możemy zatem policzyć granice ograniczeń.

Ograniczenie dolne:

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

Ograniczenie górne:

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

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

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

   Pokazanie
   $$\lim_{t\to 0}\frac{P(X_1<t\wedge X_2<t)}{t}=0$$
   pozostawiamy jako ćwiczenie.

□

7.3 Scalanie i rozdzielanie

7.3.1 Scalanie

Ta prostsza część.

Twierdzenie 7.3.1 (Twierdzenie 8.12 P&C)**.** Niech N_1, N_2 będą niezależnymi procesami Poissona z parametrami λ_1, λ_2. Wtedy $N(t) = N_1(t) + N_2(t)$ jest procesem Poissona z parametrem $\lambda_1 + \lambda_2$ a ponadto każde zdarzenie procesu N przyszło z procesu N_1 z prawdopodobieństwem $\frac{\lambda_1}{\lambda_1+\lambda_2}$

Dowód. Pierwsze trzy warunki mamy za darmo.

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

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

7.3.2 Rozdzielanie

Ta smutniejsza część.

Twierdzenie 7.3.2 (Twierdzenie 8.13 P&C)**.** Niech N będzie procesem Poissona z parametrem λ. Każde zdarzenie jest niezależnie typu 1 z prawdopodobieństwem p oraz typu 2 z prawdopodobieństwem $1 - p$.

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

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

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

Dostaliśmy rozkład Poissona z parametrem λpt, czyli N_1 jest procesem Poissona z parametrem λpt. Tak samo pokazujemy N_2.

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

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

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

Zauważamy bardzo odkrywczą rzecz, mianowicie $N_2(u) = N_2(t) + (N_2(u) - N_2(t))$ Pokazaliśmy już, że $N_1(t)$ oraz $N_2(t)$ są niezależne, więc wystarczy pokazać, że $N_1(t)$ oraz $N_2(u) - N_2(t)$ też są niezależne. A tak jest, dlatego, że oryginalny N był procesem Poissona i rozdzielanie robiliśmy niezależnie, więc to ile zdarzeń z przedziału (t, u) wpadło do N_2 jest niezależne od $N_1(t)$. □

7.4 Warunkowe czasy zdarzeń

Lemat 7.4.1. Niech X_1 będzie pierwszym międzyczasem procesu Poissona N z parametrem λ. Zmienna $X_1 \mid N(t) = 1$ ma rozkład jednostajny na $[0, t]$.

Dowód.

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

□

Twierdzenie 7.4.1. Niech $\{N(t) : t \geq 0\}$ będzie procesem Poissona z parametrem λ. Niech T_i będzie czasem przyjścia i-tego zdarzenia. Przy warunku $N(t) = n$ rozkład $(T_1, \ldots, T_n)$ jest taki sam jak $\operatorname{sort}(X_1, \ldots, X_n)$, gdzie zmienne $X_1, \ldots, X_n$ mają rozkład jednostajny na $[0, t]$ i są niezależne.

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

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

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

$$P((Y_1, \ldots, Y_n) \leq (s_1, \ldots, s_n)) = \sum_{(i_1,\ldots,i_n)\in S_n} P(X_{i_1} \leq \ldots \leq X_{i_n} \cap X_{i_1} \leq s_1 \cap \ldots \cap X_{i_n} \leq s_n)$$
$$= n!P(X_1 \leq \ldots \leq X_n \cap (X_1, \ldots, X_n) \leq (s_1, \ldots, s_n)) = n! \int_{u_1=0}^{s_1} \ldots \int_{u_n=u_{n-1}}^{s_n} \left(\frac{1}{t}\right)^n \mathrm{d}u_n \ldots \mathrm{d}u_1$$
$$= \frac{n!}{t^n} \int_{u_1=0}^{s_1} \ldots \int_{u_n=u_{n-1}}^{s_n} \mathrm{d}u_n \ldots \mathrm{d}u_1.$$

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

dzyczas. Mamy

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

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

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

Rozdział 8

Proces Markowa

8.1 Definicje

Poznaliśmy już łańcuchy Markowa, które opisywały zachowanie jakiegoś procesu mierzonego w dyskretnych odstępach czasowych. W tym rozdziale opiszemy procesy Markowa, które są ich ciągłym odpowiednikiem.

Definicja 8.1.1. Procesem Markowa nazywamy proces stochastyczny

$$\{X_t \mid t \in \mathbb{R}, t \geq 0\}$$

w którym

$$\forall s,t : P(X_{s+t} = x \mid X_u = a(u), 0 \leq u \leq t) = P(X_{s+t} = x \mid X_t = a(t))$$

gdzie a jest dowolną funkcją mapującą czas na zdarzenia.

Idea jest ta sama co w łańcuchach Markowa – prawdopodobieństwo na zdarzenie w danym momencie pod warunkiem że znamy jakąś jego historię, ma być takie samo co prawdopodobieństwo tego zdarzenia gdy znamy tylko ostatnie zdarzenie z tej historii. Tak jak w łańcuchach Markowa, będziemy zakładać że czas jest homogeniczny, tj. jest jedynie indeksem i nie wpływa w żaden sposób na prawdopodobieństwa zdarzeń.

Będziemy zajmować się jedynie dyskretnymi procesami tj. takimi, które przyjmują przeliczalnie wiele stanów.

Dzięki temu możemy myśleć o procesie Markowa jako połączenie dwóch procesów:

1. łańcucha Markowa zadanego macierzą przejścia $\mathbf{P}$
2. parametrów $\lambda_1, \lambda_2, \ldots$ opisujących rozkład trwania stanu i.

   Ściślej mówiąc – czas spędzony w stanie i ma rozkład wykładniczy z parametrem λ_i

8.1.1 Rozkład stacjonarny

Tutaj ponownie mamy analogię do łańcuchów Markowa (ciekawa sprawa, nie?) – rozkład stacjonarny procesu to będzie taki wektor $\bar{\pi} = [\pi_1, \pi_2, \ldots]$, który opisuje jakie mamy prawdopodobieństwo, że w losowym momencie t daleko w przyszłości będziemy akurat w stanie i.

Oczywiście, jako że proces ten jest ciągły to wartość π_i musi jakoś uwzględniać zarówno prawdopodobieństwa przejścia z macierzy $\mathbf{P}$ jak i czas przez który siedzimy w tym stanie jak już do niego trafimy zadany parametrem λ_i

Twierdzenie 8.1.1 (strona 228 P&C)**.** Rozkład stacjonarny procesu Markowa zadany jest przez równania:

$$\forall i : \pi_i \lambda_i = \sum_k \pi_k \lambda_k P_{k,i}$$

$$\sum_i \pi_i = 1$$

W szczególności jeśli $\forall_i \lambda_i = \lambda$ to $\pi_i = \sum_k \pi_k P_{k,i}$ czyli rozkład procesu Markowa jest taki sam jak jego łańcucha. Ma to sens – skoro w każdym czasie spędzamy tyle samo czasu, to asymptotycznie wszystko powinno zależeć jedynie od tego jak często wchodzimy do tych stanów.

Warto zwrócić uwagę na indeksy – sumujemy się po stanach k z których **wchodzimy** do stanu i. Patrząc na macierz odpowiada to wzięciu i-tej kolumny.

Ćwiczenie 8.1.1. Maszyna pracuje nieprzerwanie przez liczbę godzin będącą zmienną losową X, która ma rozkład wykładniczy z parametrem $\frac{1}{10}$. Po tym czasie ulega awarii, i jest naprawiana przez kolejne Y godzin, gdzie Y jest niezależne od X i ma rozkład wykładniczy z parametrem $\frac{9}{10}$. Po naprawie maszyna zaczyna od razu pracować i cały cykl się powtarza.

1. Pokaż, że możemy modelować tę maszynę procesem Markowa
2. Wyznacz macierz przejścia stowarzyszonego łańcucha Markowa
3. Wyznacz rozkład stacjonarny procesu oraz rozkład stacjonarny łańcucha

Dowód.

1. W naturalny sposób będziemy mieli dwa stany – nazwijmy je $P = 1$ (praca) oraz $A = 2$ (awaria). Aby pokazać, że jest to proces Markowa, to musimy udowodnić, że jest bez pamięci.

   Jeśli jesteśmy w momencie t to po pierwsze – to w jakim stanie była maszyna przed t nie ma już wpływu na kolejne wydarzenia, bo wszystkie zmienne są niezależne.

   Ponadto w czasie t możemy zresetować obecną zmienną odliczającą czas do zmiany stanu (tzn. pracować pod warunkiem, że jest ona większa niż czas do poprzedniego zdarzenia)

i tym samym zniwelować jakąkolwiek zależność, która wynikałaby z tego, że wiemy jak długo czekamy na zmianę stanu.

2. Zauważmy, że przejścia zawsze są $P \to A$ oraz $A \to P$

   W takim razie macierz ma postać:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

3. Niech $\bar{\pi}$ będzie rozkładem stacjonarnym procesu.

   Mamy dane równania:

$$\pi_P \cdot \frac{1}{10} = \pi_P \cdot \frac{1}{10} \cdot 0 + \pi_A \cdot \frac{9}{10} \cdot 1$$
$$\pi_A \cdot \frac{9}{10} = \pi_P \cdot \frac{1}{10} \cdot 1 + \pi_A \cdot \frac{9}{10} \cdot 0$$
$$\pi_P + \pi_A = 1$$

   Rozwiązując ten układ równań otrzymamy $\pi_P = \frac{9}{10}, \pi_A = \frac{1}{10}$ Wyniki nawet pokrywają się z oczekiwaniami – parametr czasu pracy jest dość mały w porównaniu do parametru awarii, co oznacza, że przez większość czasu maszyna powinna pracować.

   Aby dostać rozkład stacjonarny łańcucha $\bar{\pi}'$ rozwiązujemy podobny układ równań

$$\pi'_P = \pi'_P \cdot 0 + \pi'_A \cdot 1$$

$$\pi'_A = \pi'_P \cdot 1 + \pi'_A \cdot 0$$
$$\pi'_P + \pi'_A = 1$$

   Wychodzi nam $\pi'_P = \pi'_A = \frac{1}{2}$ – maszyna jest raz w jednym stanie a raz w drugim, więc asymptotycznie będzie w każdym tyle samo razy.

□

8.2 Kolejka Markowa

8.2.1 Definicja

W różnych sytuacjach, zwłaszcza w informatyce, mamy do czynienia z kolejkami do których przychodzą zdarzenia w pewnym sensie losowo, i które są potem obsługiwane również względnie losowo. Kolejka Markowa służy właśnie do modelowania takich procesów.

Na opisanie modelu kolejki będziemy stosować notację $Y/Z/n$, gdzie:

1. Y opisuje rozkład z jakim przychodzą zdarzenia (klienci)
2. Z opisuje rozkład czasu jaki zajmuje obsłużenie pojedynczego klienta
3. n mówi ile mamy wątków, które jednocześnie będą obsługiwać zapytania z kolejki

8.2.2 Kolejka $M/M/1$

M w naszej definicji oznacza *Memoryless*. Innymi słowy, nowi klienci przychodzą zgodnie z procesem Poissona z parametrem λ a czas obsługi (przez jedyny wątek) ma rozkład wykładniczy z parametrem μ.

Liczbę klientów oczekujących w kolejce możemy modelować procesem Markowa $\{M_t \mid t \geq 0\}$

Niech $P_k(t) = P(M_t = k)$ oznacza prawdopodobieństwo na to, że obecnie w kolejce mamy k klientów.

Twierdzenie 8.2.1 (strona 231 P&C)**.** Niech $\bar{\pi}$ będzie rozkładem stacjonarnym naszego procesu. Wtedy

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^k$$

Oczekiwana liczba klientów w kolejce wynosi

$$\begin{aligned} C &= \sum_{k=0}^{\infty} k\pi_k \\ &= \frac{\lambda}{\mu}\sum_{k=1}^{\infty} k \cdot \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{k-1} \\ &= \frac{\lambda}{\mu} \cdot \frac{1}{1 - \lambda/\mu} \\ &= \frac{\lambda}{\mu - \lambda} \end{aligned}$$

W drugim przejściu zauważamy, że wygląda to jak rozkład geometryczny z parametrem $p = 1 - \frac{\lambda}{\mu}$, którego wartość oczekiwana wynosi $\frac{1}{p}$

8.2.3 Kolejka M/M/1/K

Ograniczamy rozmiar naszej kolejki przez K – każdy klient ponad nadmiar jest odsyłany z kwitkiem. Jaki teraz mamy rozkład stacjonarny?

Twierdzenie 8.2.2 (strona 233 P&C)**.** Rozkład stacjonarny kolejki ograniczonej przez K ma postać:

$$\pi_0 = \frac{1}{\sum_{k=0}^{K}(\lambda/\mu)^k}$$

$$\pi_k = \begin{cases} \pi_0 \cdot \left(\frac{\lambda}{\mu}\right)^k & \text{gdy } k \leq K \\ 0 & \text{gdy } k > K \end{cases}$$

8.2.4 Ćwiczenia

Ćwiczenie 8.2.1. Do małego kantoru przychodzą klienci zgodnie z procesem Poissona w tempie 0.2 klienta na minutę. Średni czas obsługi klienta zajmuje 2 minuty. Ponieważ pomieszczenie jest niewielkie to kolejka mieści maksymalnie dwóch klientów (nie licząc obecnie obsługiwanego), jeśli kolejka jest pełna to nowy klient udaje się gdzie indziej.

1. Zamodeluj ten proces jako proces Markowa – wyznacz graf stanów i macierz przejścia.
2. Wyznacz rozkład stacjonarny procesu oraz jego łańcucha
3. Asymptotycznie jaką część czasu kolejka jest pełna?
4. Asymptotycznie jaka część klientów uda się gdzie indziej z uwagi na pełną kolejkę?
5. W tym momencie w kolejce czeka dokładnie jeden klient. Jakie jest prawdopodobieństwo, że zostanie obsłużony zanim przyjdzie kolejna osoba?

Dowód.

1. Na początek warto przetłumaczyć opis słowny na model matematyczny. Mamy dany proces Poissona z parametrem $\lambda = 0.2$

   Zakładamy, że czas przetwarzania ma rozkład wykładniczy (bo w sumie nie przerabialiśmy żadnej innej opcji) i skoro oczekiwany czas wynosi 2 to musi mieć on parametr $\mu = \frac{1}{2}$

   Będziemy mieli trzy możliwe stany $0, 1, 2, 3$ na możliwe liczby osób w kantorze (wliczając w to osobę obsługiwaną) między którymi przejścia opisuje macierz

   $$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ p & 0 & 1-p & 0 \\ 0 & p & 0 & 1-p \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

   Wyjaśnijmy co tu się dzieje:

   - Jeśli w kantorze jest 0 klientów to jedyne co się może zdarzyć to przyjść nowy, i w ten sposób mamy jednego klienta z prawdopodobieństwem 1
   - Kiedy mamy jednego lub dwóch klientów to są dwie opcje – albo przychodzi nowy klient i kolejka rośnie o 1 albo obsłużony zostaje obecny klient i kolejka maleje o 1. p jest tutaj prawdopodobieństwem, że kolejka maleje, czyli że zanim przyjdzie kolejna osoba to kończymy obsługiwać obecną.

- Jeśli kantor jest pełny (czyli są w nim trzy osoby) to jedyne co może się stać, to może zostać obsłużony klient i zwolnić jedno miejsce.

Pasowałoby policzyć p - czyli prawdopodobieństwo, że obsłużymy klienta zanim przyjdzie kolejny. Mamy dwie zmienne wykładnicze X z parametrem λ oraz Y z parametrem μ. X liczy czas do przyjścia kolejnego klienta a Y czas do końca obsługi obecnego.

$p = P(\min(X, Y) = Y) = \frac{\mu}{\mu+\lambda} = \frac{0.5}{0.7} = \frac{5}{7}$

Potrzebujemy jeszcze wyznaczyć parametry $\lambda_0, \lambda_1, \lambda_2, \lambda_3$ dla czasów spędzonych w każdym ze stanów.

Łatwo pokazać, że

$$\begin{cases} \lambda_0 = \lambda = 0.2 \\ \lambda_1 = \lambda + \mu = 0.7 \\ \lambda_2 = \lambda + \mu = 0.7 \\ \lambda_3 = \mu = 0.5 \end{cases}$$

2. Dla procesu rozwiązujemy układ równań

$$\begin{cases} \pi_0\lambda_0 = \pi_1\lambda_1 \cdot p \\ \pi_1\lambda_1 = \pi_0\lambda_0 \cdot 1 + \pi_2\lambda_2 \cdot p \\ \pi_2\lambda_2 = \pi_1\lambda_1 \cdot (1-p) + \pi_3\lambda_3 \cdot 1 \\ \pi_3\lambda_3 = \pi_2\lambda_2 \cdot (1-p) \\ \pi_0 + \pi_1 + \pi_2 + \pi_3 = 1 \end{cases}$$

Dla łańcucha jest podobnie, tylko bez lambd

$$\begin{cases} \pi'_0 = \pi'_1 \cdot p \\ \pi'_1 = \pi'_0 \cdot 1 + \pi'_2 \cdot p \\ \pi'_2 = \pi'_1 \cdot (1-p) + \pi'_3 \cdot 1 \\ \pi'_3 = \pi'_2 \cdot (1-p) \\ \pi'_0 + \pi'_1 + \pi'_2 + \pi'_3 = 1 \end{cases}$$

3. Sprawdzamy π_3 wyliczone w poprzednim punkcie
4. Klienci widzą stan kolejki zgodny z rozkładem stacjonarnym, zatem asymptotycznie π_3 klientów zobaczy pełną kolejkę.
5. jest to dokładnie p

□

Rozdział 9

Rozkład normalny

9.1 Definicja

Definicja 9.1.1. Definiujemy rozkład normalny (uogólniony) jako rozkład o następującej funkcji gęstości prawdopodobieństwa:

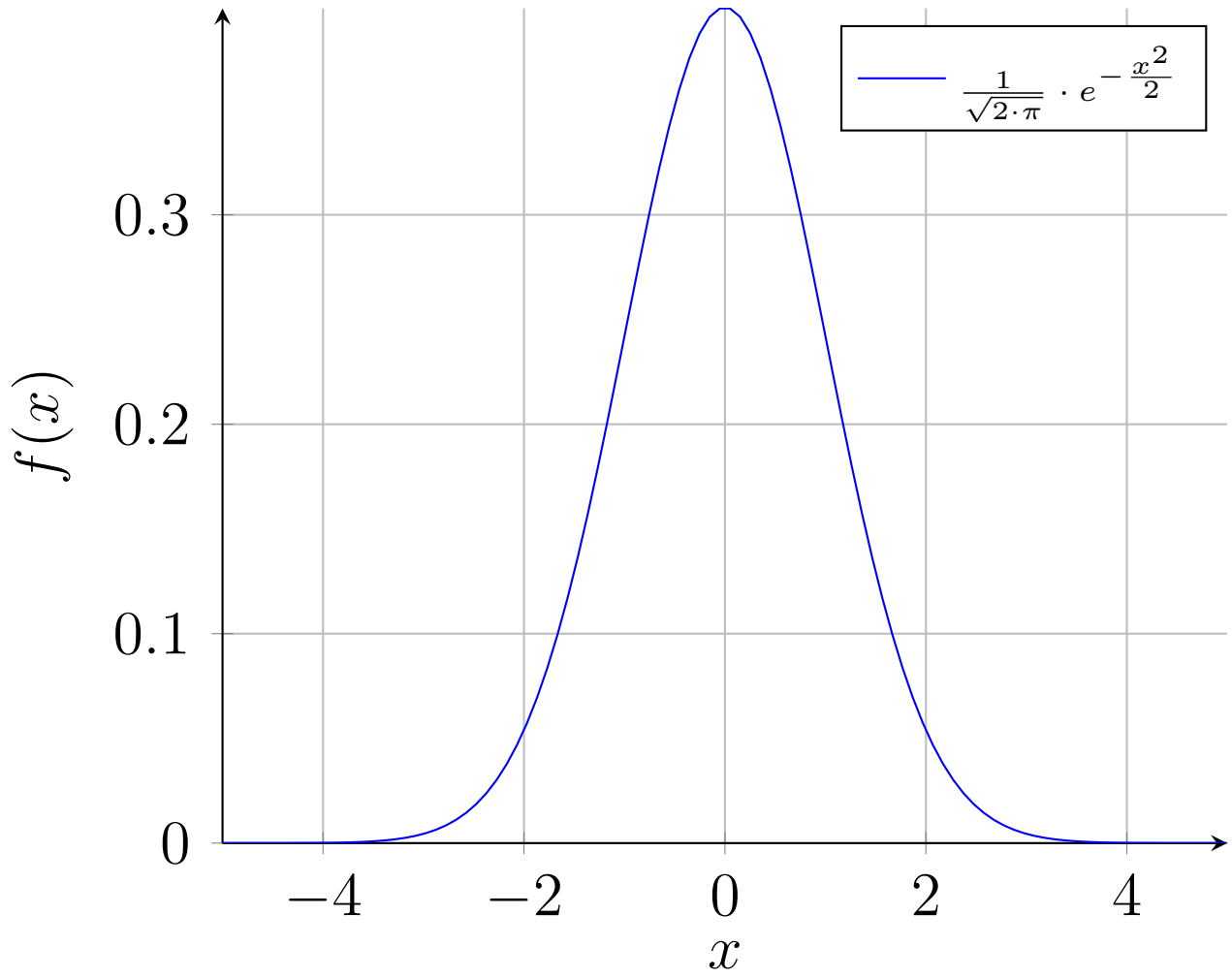
$$f_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-((z-\mu)/\sigma)^2/2}$$

Definicja 9.1.2. Rozkład normalny o parametrach μ i σ^2 oznaczamy jako $N(\mu, \sigma^2)$.

Definicja 9.1.3. Standardowy rozkład Normalny to rozkład normalny o parametrach $\mu = 0$ i $\sigma^2 = 1$; oznaczamy go (bez większego szoku) jako $N(0, 1)$.

Definicja 9.1.4. Dystrybuantę standardowego rozkładu normalnego oznaczamy jako Φ.

Funkcja gęstości prawdopodobieństwa standardowego rozkładu normalnego wygląda jak ~~dzban~~ dzwon.



9.2 Wartość oczekiwana i wariancja

Najpierw wyliczmy wartość oczekiwaną standardowego rozkładu normalnego.

Twierdzenie 9.2.1. Wartość oczekiwana standardowego rozkładu normalnego wynosi 0, wariancja wynosi 1.

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

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

ponieważ $\mathbb{E}[Z] = 0$

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

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

całkowanie przez części

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

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

Lemat 9.2.1. Zmienna losowa ma rozkład normalny wtedy i tylko wtedy gdy jest transformacją liniową zmiennej losowej o standardowym rozkładzie normalnym

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

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

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

□

czyli ten ~~dzban~~ dzwon ma efektywnie przesunięcie o μ.

9.3 Przykład

Przykład z książki i wykładu rozkładu normalnego: Detekcja Sygnału (p. 245)

Ćwiczenie 9.3.1.

Przyjmijmy że mamy transmiter, który wysyła bit zakodowany za pomocą $S\in -1,+1$. Ale nie żyjemy w idealnym świecie i dodawany jest szum Y, który jest zmienną losową o normalnym rozkładzie ze średnią 0 i odchyleniem standardowym σ.
Sygnał ten jest odbierany i dekodowany w zależności jaki znak danego bita odbierzemy.

$$R=\mathrm{sgn}(S+Y)$$

Chcemy znaleźć prawdopodobieństwo, że otrzymany bit będzie inny niż ten, który nadamy (szum zmieni znak) $P(R\neq S)$.

Prawdopodobieństwo, że błąd wyskoczy dla $S=1$ jest równe prawdopodobieństwu, że $Y\leq -1$

$$P(Y\leq -1)=P\left(\frac{Y-\mu}{\sigma}\leq\frac{-1-\mu}{\sigma}\right)=\Phi\left(-\frac{1}{\sigma}\right)$$

Dla $S=-1$ mamy podobnie (to jest symetryczne btw)

$$P(Y\geq 1)=1-P\left(\frac{Y-\mu}{\sigma}\leq\frac{1-\mu}{\sigma}\right)=1-\Phi\left(\frac{1}{\sigma}\right)$$

Ponieważ symetria btw $\Phi(-\frac{1}{\sigma}) = 1 - \Phi(\frac{1}{\sigma})$

Wynik wynosi zatem $2(1 - \Phi(\frac{1}{\sigma})$

Dalej musimy odczytać z tabeli (BOOOOOOOOOORING)

9.4 Centralne Twierdzenie Graniczne

Intuicyjnie: Centralne Twierdzenie Graniczne mówi, że jak mamy jakieś niezależne zmienne losowe (niekoniecznie o takim samym rozkładzie) to rozkład średnich tych wylosowanych wartości będzie zbiegać do rozkładu normalnego po wykonaniu pewnej, dużej liczby prób. Twierdzenie to uzasadnia występowanie w naturze rozkładu normalnego.

Definicja 9.4.1. Ciąg dystrybuant $F_1, F_2, ...$ zbiega w dystrybuancie do dystrybuanty F, co oznaczamy jako $F_n \to F$, jeśli dla każdego $a \in \mathbb{R}$ w którym F jest ciągła zachodzi:

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

Twierdzenie 9.4.1 (Centralne Twierdzenie Graniczne)**.** Niech $X_1, X_2, ...X_n$ będą niezależnymi zmiennymi losowymi o takim samym rozkładzie, średniej μ i wariancji σ^2. Niech $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Wówczas dla dowolnych a, b

$$\lim_{n\to\infty} \mathrm{P}\left(a \leq \frac{\overline{X_n} - \mu}{\sigma} \cdot \sqrt{n} \leq b\right) \to \Phi(b) - \Phi(a)$$

Dowód. Aby dowieść CTG, będziemy musieli przytoczyć *pomocne twierdzonko*, którego (mamy nadzieję) nikt nie będzie musiał dowodzić:

Twierdzenie 9.4.2 (Lévy-Cramér)**.** Niech $Y_1, Y_2, \ldots$ będzie sekwencją zmiennych losowych, gdzie Y_i ma dystrybuantę F_i i funkcję tworzącą momentów M_i. Niech Y będzie zmienną losową o dystrybuancie F i funkcji tworzącej momenty M. Jeżeli dla każdego t zachodzi:

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

to $F_n \to F$ dla wszystkich t w których $F(t)$ jest ciągła.

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

Przystępujemy teraz do dowodzenia CTG.

Definiujemy $Z_i = (X_i - \mu)/\sigma$. Wówczas Z_i są to niezależne zmienne losowe oraz:

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

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

(gdzie korzystamy z faktu, że dla stałej c $\mathrm{Var}[cX] = c^2 \mathrm{Var}[X]$).

Ponadto mamy, że:

$$\frac{\bar{X_n} - \mu}{\sigma} \cdot \sqrt{n} = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} = \frac{\sqrt{n}}{n} \sum_{i=1}^{n} Z_i = \frac{\sum_{i=1}^{n} Z_i}{\sqrt{n}}$$

Żeby zastosować teraz przywołane przez nas twierdzenie Levy'ego i tego drugiego musimy pokazać, że MGF[1] zmiennych losowych postaci

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

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

W takim razie, chcemy w tym celu pokazać coś takiego:

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

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

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

Ponieważ Z_i są niezależne i mają ten sam rozkład:

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

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

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

Dodatkowo, *również bez jakiejkolwiek przyczyny*, policzmy sobie pierwszą i drugą pochodną

[1] Moment Generating Function

$L(0)$.

Zacznijmy od trywialnych obserwacji:

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

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

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

W ostatnim przejściu korzystamy z faktu, że $\mathbb{E}[Z_i^2] = 1$. Wynika to z faktu, że $\mathbb{E}[Z_i^2] - \mathbb{E}[Z_i]^2 = \sigma^2 = 1$.

Przypomnijmy, że chcieliśmy pokazać, że:

$$\lim_{n\to\infty} M_{Y_n}(t) \to e^{t^2/2}$$

lub równoważnie:

$$\lim_{n\to\infty} \left(M_{Z_i}\left(\frac{t}{\sqrt{n}}\right)\right)^n \to e^{t^2/2}$$

po zlogarytmowaniu stronami:

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

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

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

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

No i lecimy z pochodnymi!

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

I w sumie to mieliśmy dowieść. Ale fajnie. □

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

Twierdzenie 9.4.3. Niech $X_1, \ldots, X_n$ będzie ciągiem niezależnych zmiennych losowych z $\mathbb{E}[X_i] = \mu_i$ i $\mathrm{Var}[X_i] = \sigma_i^2$. Niech zachodzi

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

Wówczas

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

Twierdzenie 9.4.4 (Berry-Esséen)**.** Istnieje taka stała C, że zachodzi następujące: niech $X_1, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie ze skończoną wartością oczekiwaną μ i wariancją σ^2. Dalej niech $\rho = \mathbb{E}\left[|X_i - \mu|^3\right] < \infty$ i $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Mamy

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

Rozdział 10

Entropia

10.1 Definicja

Definicja 10.1.1. Entropię dyskretnej zmiennej losowej X definiujemy jako

$$H(X) = \sum_{x \in \operatorname{im} X} -P(X = x) \cdot \lg P(X = x) = \mathbb{E}\left[\lg \frac{1}{P(X)}\right]$$

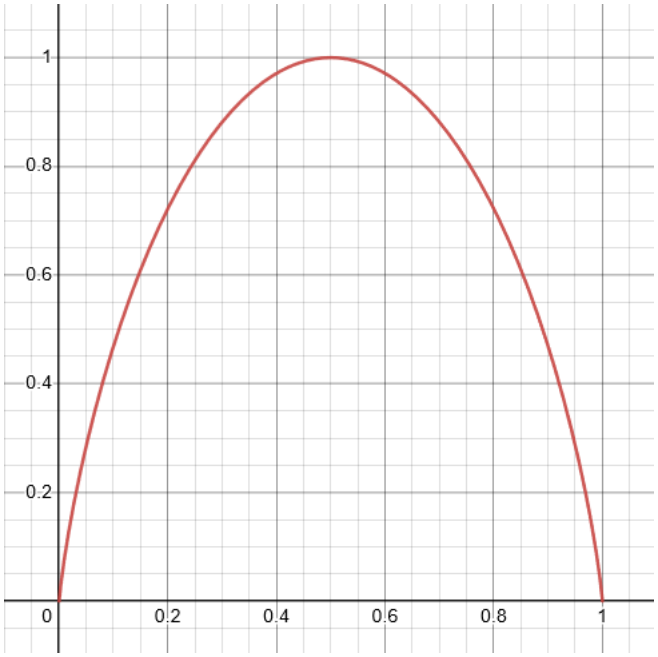
Zaznaczmy, że w rozdziale o entropii będziemy posługiwać się logarytmem dwójkowym (lg), a nienaturalnym (ln), tak jak zazwyczaj.

Zdefiniujemy też entropię indykatora (skrzywionego rzutu monetą)

Definicja 10.1.2. Niech $p \in (0, 1)$. Definiujemy:

$$H(p) = -p \lg p - (1 - p) \lg(1 - p)$$

Dla wygody przyjmujemy $H(0) = H(1) = 0$.



Rysunek 10.1: Wykres funkcji H

Widzimy, że entropia jest największa dla $p = \frac{1}{2}$ – intuicyjnie odpowiada to faktowi, że sprawiedliwa moneta generuje najwięcej losowości (cały jeden bit informacji), a bardzo skrzywiona moneta bardzo mało (prawie zero bitów informacji).

Nieco dziwnie jest myśleć o ułamkowych bitach informacji; koncept ten ma więcej sensu jeśli rzucamy taką monetą wiele razy – jeśli pojedynczy rzut niesie nam pół bitu to 100 rzutów daje nam 50 bitów „czystej" informacji.

Bardziej namacalne konsekwencje tej intuicji zobaczymy w późniejszych sekcjach, w których pokażemy jak można mierzyć losowość i kompresję za pomocą entropii.

Tym czasem pokażmy inny ciekawy fakt – jeśli wykonujemy dwa niezależne eksperymenty to łączna entropia takiej zabawy jest sumą entropii każdego z osobna.

Twierdzenie 10.1.1 (Lemat 10.1 P&C)**.** Jeśli X_1, X_2 są niezależne, oraz $Y = (X_1, X_2)$ to

$$H(Y) = H(X_1) + H(X_2)$$

Dowód. Stety lub nie, jest to jedna wielka pała.

$$H(Y) = -\sum_{x_1,x_2} P((X_1, X_2) = (x_1, x_2)) \cdot \lg P((X_1, X_2) = (x_1, x_2))$$

Z niezależności rozbijamy na iloczyn prawdopodobieństw

$$H(Y) = -\sum_{x_1,x_2} P(X_1 = x_1)P(X_2 = x_2) \cdot (\lg P(X_1 = x_1) + \lg P(X_2 = x_2))$$

Rozbijamy teraz sumę na tej sumie logarytmów:

$$\begin{aligned} H(Y) = &- \sum_{x_1,x_2} P(X_1 = x_1)P(X_2 = x_2) \cdot \lg P(X_1 = x_1) \\ &- \sum_{x_1,x_2} P(X_1 = x_1)P(X_2 = x_2) \cdot \lg P(X_2 = x_2) \end{aligned}$$

Zauważamy, że możemy wyciągnąć odpowiednio $\sum P(X_2 = x_2)$ oraz $\sum P(X_1 = x_1)$

$$\begin{aligned} H(Y) = &- \sum_{x_2} P(X_2 = x_2) \sum_{x_1} P(X_1 = x_1) \cdot \lg P(X_1 = x_1) \\ &- \sum_{x_1} P(X_1 = x_1) \sum_{x_2} P(X_2 = x_2) \cdot \lg P(X_2 = x_2) \end{aligned}$$

Lewe czynniki sumują się do 1, a prawe do odpowiednich entropii, zatem

$$H(Y) = H(X_1) + H(X_2)$$

□

10.2 Szacowanie współczynników dwumianowych

Twierdzenie 10.2.1 (Lemat 10.2 P&C)**.** Niech $n \in \mathbb{N}, q \in [0,1]$ oraz $nq \in \mathbb{N}$. Wtedy

$$\frac{2^{nH(q)}}{n+1} \leq \binom{n}{nq} \leq 2^{nH(q)}$$

Dowód. Dla $q = 0$ i $q = 1$ dowód jest oczywisty. Niech zatem $q \in (0,1)$

1. Pokażmy najpierw ograniczenie górne.

   Ponieważ $nq \in \mathbb{N}$ wyrażenie $\binom{n}{nq}$ występuje we wzorze

   $$1 = (q + (1-q))^n = \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} = \binom{n}{nq} q^{qn} (1-q)^{(1-q)n} + \text{coś}$$

   Zatem
   $$\binom{n}{nq} \leq q^{-qn}(1-q)^{-(1-q)n} = 2^{-nq\lg q - n(q-1)\lg(1-q)} = 2^{nH(q)}$$

2. Teraz pokażemy ograniczenie dolne.

   Będziemy chcieli pokazać, że $\binom{n}{q} q^{qn}(1-q)^{(1-q)n}$ jest największym składnikiem sumy – w ten sposób dostaniemy
   $$\binom{n}{nq} q^{qn}(1-q)^{(1-q)n} \geq \frac{1}{n+1}$$
   co przekształcamy do postulowanej nierówności tak jak zrobiliśmy to przed chwilą.

   Różnica dwóch kolejnych składników sumy wynosi

   $$\binom{n}{k} q^k (1-q)^{n-k} - \binom{n}{k+1} q^{q+1}(1-q)^{n-k-1}$$

   $$= \binom{n}{k} q^k (1-q)^{n-k} \cdot \left(1 - \frac{n-k}{k+1} \cdot \frac{q}{1-q}\right)$$

   Jak wyliczymy nawias do sensownej postaci to nam wyjdzie, że jest on dodatni tylko gdy

   $$k \geq qn - 1 + q$$

   W takim razie pierwszym takim k jest $k = qn$, czyli qn-ty składnik jest tym największym.

□

Twierdzenie 10.2.2 (Wniosek 10.3 P&C)**.**
Dla $0 \leq q \leq \frac{1}{2}$

$$\frac{2^{nH(q)}}{n+1} \leq \binom{n}{\lfloor nq \rfloor} \leq 2^{nH(q)}$$

Dla $\frac{1}{2} \leq q \leq 1$

$$\frac{2^{nH(q)}}{n+1} \leq \binom{n}{\lceil nq \rceil} \leq 2^{nH(q)}$$

Dowód. 1. Pokażmy najpierw ograniczenia górne.

Skorzystamy tutaj z prostej obserwacji:

$$\lfloor nq \rfloor \leq nq \leq \lceil nq \rceil$$

Zobaczmy jak to wygląda dla pierwszego przypadku. Skoro $\lfloor nq \rfloor \leq nq$ oraz $q \leq \frac{1}{2}$ to $q^{qn} \leq q^{\lfloor qn \rfloor}$ oraz $(1-q)^{n-qn} \leq (1-q)^{n-\lfloor qn \rfloor}$, a zatem

$$\binom{n}{\lfloor nq \rfloor} q^{qn}(1-q)^{(1-q)n} \leq \binom{n}{\lfloor nq \rfloor} q^{\lfloor qn \rfloor}(1-q)^{n-\lfloor qn \rfloor} \leq 1$$

Stąd stosujemy rozumowanie pokazane w dowodzie twierdzenia 10.2.1 aby dostać ograniczenie górne.

Analogicznie pokazujemy przypadek drugi.

2. Przejdźmy teraz do ograniczeń dolnych i ponownie ograniczmy się do pierwszego przypadku.

Najlepiej byłoby pokazać

$$\binom{n}{\lfloor nq \rfloor} q^{qn}(1-q)^{(1-q)n} \geq \frac{1}{n+1}$$

ale niestety w książce tego nie ma (jest za to blef) a na wykładzie tego nie omawialiśmy.

□

10.3 Entropia jako miara losowości

Definicja 10.3.1. Funkcją ekstrakcji nazywamy taką funkcję Ext, która zwraca bity w oparciu o wartości zwracane przez zmienną X w taki sposób, że dla dowolnego ciągu binarnego y zachodzi

$$P(|\mathrm{Ext}(X)| = k) \neq 0 \implies P(\mathrm{Ext}(X) = y \mid |y| = k) = 2^{-k}$$

Innymi słowy – jeśli możemy dostać *jakiś* ciąg k-bitowy to możemy dostać *każdy* ciąg k-bitowy, wszystkie z równym prawdopodobieństwem. Dalece nie jest oczywiste, że taka funkcja ekstrakcji w ogóle istnieje (być może nasza zmienna zachowuje się w brzydki sposób).

Czasem jednak jest dobrze.

Twierdzenie 10.3.1 (Twierdzenie 10.4 P&C)**.** Niech X ma rozkład jednostajny na zbiorze $\{0, \ldots, m-1\}$. Wtedy istnieje funkcja ekstrakcji Ext, która średnio zwraca co najmniej $\lfloor H(X) \rfloor - 1 = \lfloor \lg m \rfloor - 1$ bitów.

Dowód. Naszą funkcję konstruujemy rekurencyjnie. Niech Ext_m będzie funkcją ekstrakcji dla ustalonego m.

Jeśli $m = 1$ to zwracamy pusty ciąg $\mathrm{Ext}_1(0) = \varnothing$. Rozważmy zatem $m \geq 2$.

Niech $\alpha = \lfloor \lg m \rfloor$. Definiujemy

$$\mathrm{Ext}_m(x) = \begin{cases} x & \text{dla } x < 2^\alpha \\ \mathrm{Ext}_{m-2^\alpha}(x - 2^\alpha) & \text{dla } x \geq 2^\alpha \end{cases}$$

Intuicyjnie działa to tak, że wybieramy największą potęgę dwójki, która mieści się w m i dla wszystkich liczb mniejszych od niej zwracamy po prostu ich reprezentację bitową. (wraz z zerami wiodącymi, aby miała ona długość α) Pozostałe wartości wypełniamy rekurencyjnie krótszymi ciągami.

Pozostaje policzyć oczekiwaną liczbę zwróconych bitów – nazwijmy ją $\mathbb{E}[X]$. Ponadto niech $\beta = \lfloor \lg(m - 2^\alpha) \rfloor$

$$\begin{aligned} \mathbb{E}[Y] &\geq \frac{2^\alpha}{m} \cdot \alpha + \frac{m - 2^\alpha}{m} \cdot (\lfloor \lg(m - 2^\alpha) \rfloor - 1) \\ &= \frac{m + 2^\alpha - m}{m} \cdot \alpha + \frac{m - 2^\alpha}{m} \cdot (\beta - 1) \\ &= \alpha + \frac{m - 2^\alpha}{m} \cdot (-\alpha + \beta - 1) \end{aligned}$$

Mamy ponadto oszacowanie

$$\frac{m - 2^\alpha}{m} \leq \frac{2^{\beta+1} - 1}{2^\alpha + 2^{\beta+1} - 1}$$

Zatem

$$\begin{aligned}\mathbb{E}[Y] &\geq \alpha - \frac{2^{\beta+1}-1}{2^{\alpha}+2^{\beta+1}-1}\cdot(\alpha-\beta+1)\\ &\geq \alpha - \frac{2^{\beta+1}-1}{(2^{\beta+1}-1)\cdot(2^{\alpha-\beta-1}+1)}\cdot(\alpha-\beta+1)\\ &\geq \alpha - \frac{\alpha-\beta+1}{2^{\alpha-\beta-1}+1}\\ &\geq \alpha - 1\end{aligned}$$

□

Twierdzenie 10.3.2 (Twierdzenie 10.5 P&C)**.** Niech $p \in \left(\frac{1}{2}, 1\right)$ będzie prawdopodobieństwem sukcesu pojedynczej próby oraz $\delta > 0$

Wtedy dla odpowiednio dużego n oraz ciągu n niezależnych prób:

1. Istnieje funkcja ekstrakcji, która średnio zwraca co najmniej $(1-\delta)\cdot n \cdot H(p)$ niezależnych bitów.
2. Średnia liczba bitów zwracana przez dowolną funkcję ekstrakcji to co najwyżej $n \cdot H(p)$

Dowód.

1. Zaczynamy od pierwszego podpunktu

   Niech j będzie liczbą sukcesów. Mamy $\binom{n}{j}$ możliwych ciągów, każdy wypada z takim samym prawdopodobieństwem.

   Na podstawie poprzedniego twierdzenia możemy zdefiniować na takich ciągach funkcję ekstrakcji.

   Niech Z będzie liczbą sukcesów które wypadły, a B liczbą bitów, którą zwracamy w opisany sposób. Wtedy

   $$\mathbb{E}[B] = \sum_{k=0}^{n} P(Z=k)\cdot\mathbb{E}[B \mid Z=k]$$

   Z poprzedniego twierdzenia mamy

   $$\mathbb{E}[B \mid Z=k] \geq \left\lfloor \lg \binom{n}{k} \right\rfloor - 1$$

   Będziemy chcieli skorzystać z oszacowania współczynników dwumianowych przez entropię.

   Dobieramy zatem jakiegoś małego $0 < \varepsilon < \min\left(p-\frac{1}{2}, 1-p\right)$ i będziemy się zajmować $k \in [n(p-\varepsilon), n(p+\varepsilon)]$

Teraz mówimy, że

$$\binom{n}{k} \geq \binom{n}{\lfloor n(p+\varepsilon) \rfloor} \geq \frac{2^{nH(p+\varepsilon)}}{n+1}$$

Teraz szacujemy

$$\begin{aligned}
\mathbb{E}[B] &\geq \sum_{k=\lfloor n(p-\varepsilon) \rfloor}^{\lceil n(p+\varepsilon) \rceil} P(Z=k) \cdot \mathbb{E}[B \mid Z=k] \\
&\geq \sum_{k=\lfloor n(p-\varepsilon) \rfloor}^{\lceil n(p+\varepsilon) \rceil} P(Z=k) \cdot \left(\left\lfloor \lg \binom{n}{k} \right\rfloor - 1 \right) \\
&\geq \sum_{k=\lfloor n(p-\varepsilon) \rfloor}^{\lceil n(p+\varepsilon) \rceil} P(Z=k) \cdot (nH(p+\varepsilon) - \lg(n+1) - 2) \\
&\geq (nH(p+\varepsilon) - \lg(n+1) - 2) \cdot P(|Z - np| \leq \varepsilon n)
\end{aligned}$$

Drugi czynnik możemy ograniczyć nierównością Chernoffa na niezależne próby Poissona. Mamy:

$$P(|Z - np| > \varepsilon n) \leq 2 \cdot \exp\left(\frac{-n\varepsilon^2}{3p} \right)$$

Zatem

$$\begin{aligned}
\mathbb{E}[B] &\geq (nH(p+\varepsilon) - \lg(n+1) - 2) \cdot \left(1 - \exp\left(\frac{-n\varepsilon^2}{3p} \right) \right) \\
&\geq (1-\delta) nH(p)
\end{aligned}$$

2. Teraz drugi podpunkt, na szczęście prostszy.

   Jeśli mamy jakąś funkcję ekstrakcji Ext i wrzucimy do niej jakiś ciąg x taki, że $P(X = x) = q$ to Ext może zwrócić co najwyżej $\lg \frac{1}{q}$ bitów.

   Dzieje się tak, ponieważ jeśli wyplujemy jakieś k bitów z prawdopodobieństwem q to każde k bitów musimy zwracać z takim prawdopodobieństwem.

   Wychodzi nam z tego, że $2^{|\mathrm{Ext}(x)|} \cdot q \leq 1$.

   Zatem

$$\begin{aligned}
\mathbb{E}[B] &= \sum P(X=x) \cdot |\mathrm{Ext}(x)| \\
&\leq \sum P(X=x) \cdot \lg \frac{1}{P(X=x)} \\
&= H(X) \\
&= n \cdot H(p)
\end{aligned}$$

□

10.4 Entropia a kompresja

Entropia pozwala nam mierzyć losowość, ale jest to też w pewnym sensie miara informacji. A informacja jest dość mocno powiązana z kompresowaniem tej informacji do najbardziej upakowanej postaci. O tym jest ten rozdział.

Definicja 10.4.1. Funkcją kompresji nazywamy taką iniekcję Com, która dla dowolnego ciągu n rzutów monetą zwraca ciąg unikatowy ciągi binarny.

Pokażemy teraz twierdzenie, które wygląda bardzo podobnie do twierdzenie z poprzedniego rozdziału, które mówiło o ekstrakcji bitów ze zmiennej.

Analogia wynika z tego, że kompresowanie jest bardzo podobnym procesem do wytwarzania losowych bitów, z tym, że tutaj wymagamy dodatkowo, aby zwracane przez nas ciągi były zawsze różne, co może wymagać dodatkowych bitów.

Twierdzenie 10.4.1. Niech $p > \frac{1}{2}$ będzie prawdopodobieństwem sukcesu pojedynczej próby. Wtedy dla dowolnego $\delta > 0$ oraz wystarczająco dużego n:

1. Istnieje funkcja kompresji Com, która średnio zwraca co najwyżej $(1+\delta)H(p)$ bitów
2. Średnia liczba bitów zwracana przez dowolną funkcję kompresji wynosi co najmniej $(1-\delta)H(p)$

Dowód.

1. Zaczynamy od pierwszej części – Skonstruujemy naszą funkcję kompresji explicite.

   Ustalmy sobie takie $\varepsilon > 0$, żeby $p - \varepsilon > \frac{1}{2}$

   Pierwszy bit będzie flagą, która jest równa 0 gdy w wejściowym ciągu występuje co najmniej $n(p-\varepsilon)$ jedynek, i 1 w przeciwnym wypadku.

   Dla tych co mają 1 przepisujemy na pałę bity z wejścia. Robimy tak, ponieważ szansa na to, że wypadło mniej niż $n(p-\varepsilon)$ jedynek jest mała ,a dokładniej zbiega do zera dla dowolnego (stałego) ε. Dowodzimy ten fakt nierównością Chernoffa.

   Szacujemy od góry liczbę pozostałych ciągów. Korzystamy tutaj z faktu, że $p - \varepsilon > \frac{1}{2}$ więc współczynniki dwumianowe od $\lceil n(p-\varepsilon) \rceil$ maleją.

$$\sum_{j=\lceil n(p-\varepsilon)\rceil}^{n} \binom{n}{j} \leq \sum_{j=\lceil n(p-\varepsilon)\rceil}^{n} \binom{n}{\lceil n(p-\varepsilon)\rceil} \leq \frac{n}{2} \cdot 2^{nH(p-\varepsilon)}$$

   Tyle jest tych ciągów – każdemu przydzielamy jakikolwiek ciąg bitów, wystarczy użyć logarytmicznie wiele bitów, czyli co najwyżej.

$$nH(p-\varepsilon) + \lg n + 1$$

Aby policzyć oczekiwaną bierzemy oszacowanie z punktu pierwszego. Mitzenmacher bierze tu dopełnienie, ale przecież

$$P(\text{liczba jedynek} > \lfloor n(p-\varepsilon) \rfloor) \geq 1 - \exp\left(\frac{-n\varepsilon^2}{2p}\right)$$

więc trochę nie działa.

Dlatego my wstawiamy 1 jako górne oszacowanie tego prawdopodobieństwa.

Wychodzi nam

$$\exp\left(\frac{-n\varepsilon^2}{2p}\right) \cdot (n+1) + 1 \cdot (nH(p-\varepsilon) + \lg n + 1)$$

Dla dowolnie ustalonego ε pierwszy składnik zbiega do zera, zatem zostaje nam zatem pokazać, że dla dowolnie ustalonej δ możemy dobrać takie ε aby dla dużych n zachodziło:

$$nH(p-\varepsilon) + \lg n + 1 \leq (1+\delta)nH(p)$$

Przekształcając otrzymujemy

$$\frac{\lg n + 1}{n} \leq (1+\delta)H(p) - H(p-\varepsilon)$$

Teraz zauważamy, że im mniejsze ε tym bliżej siebie są obie entropie po prawej stronie, zatem dla odpowiednio małego ε zachodzi

$$(1+\delta)H(p) - H(p-\varepsilon) > 0$$

Z drugiej strony $\frac{\lg n+1}{n}$ zbiega do zera, więc jak dobierzemy małe epsilon, to potem dobieramy duże n tak, aby te dwa oszacowania były rozdzielone jakąś małą stałą.

No i fajnie.

2. Teraz część druga tezy

□

10.5 Twierdzenie Shannona

Podobnie jak kompresja wprowadza kodowanie, które zmniejsza liczbę bitów potrzebnych na zapisanie informacji, tak możemy wprowadzić kodowanie, które zwiększa liczbę bitów aby uodpornić dane na błędy wynikające z niedokładności przekazu.

10.5.1 Definicje

Definicja 10.5.1. Kanałem z parametrem p nazywamy „funkcję" (bo matematycznie to nie jest funkcja), która dla zadanego ciągu $x_1, \ldots, x_k$ zwraca losowo ciąg $y_1, \ldots y_k$ taki, że

$$\forall_{1 \leq i \leq k} : P(x_i \neq y_i) = p$$

Kanał dla ustalonego będziemy oznaczać przez Channel dla p wynikającego z kontekstu.

Zauważmy, że jeśli $p = \frac{1}{2}$ to efektywnie dostajemy losowy ciąg bitów, dlatego dalej będziemy się zajmować optymistyczną sytuacją, w której $p < \frac{1}{2}$.

Definicja 10.5.2. Dla ustalonych $k, n; k \leq n$ **$(\mathbf{k}, \mathbf{n})$-enkoderem** będziemy nazywać dowolną funkcję $\{0,1\}^k \to \{0,1\}^n$, natomiast **$(\mathbf{k}, \mathbf{n})$-dekoderem** będziemy nazywać dowolną funkcję $\{0,1\}^n \to \{0,1\}^k$.

Będziemy oznaczać Enc oraz Dec dla k, n wynikających z kontekstu.

Definicja 10.5.3. Odległość edycyjną, zapisywaną jako $\mathbf{\Delta}$, ciągów $a, b \in \{0,1\}^k$ definiujemy jako liczbę pozycji na których ciągi a, b się różnią, tj. $\Delta(a,b) = |\{i \mid a_i \neq b_i\}|$

Wyposażeni w definicje szukamy jakiegoś oszacowania, jak wiele informacji jesteśmy w stanie upakować w n bitach, tak aby przesyłając dowolną wiadomość $m \in \{0,1\}^n$ przez kanał o parametrze p prawdopodobieństwo, że odczytaliśmy niepoprawną wiadomość było mniejsze niż γ.
Bardziej formalnie dla ustalonych p, n, γ szukamy $\max k$ takiego, że

$$\exists_{\text{Enc,Dec}} : \forall_{m \in \{0,1\}^n} : P(\text{Dec}(\text{Channel}(\text{Enc}(m))) = m) \geq 1 - \gamma$$

Widzimy, że jeśli $p = 0$, tj. kanał jest idealny, to $\max k = n$ a Enc = Dec = $x \mapsto x$

Dla $p > 0$ to intuicyjnie $k \approx n \cdot (1 - H(p))$, gdzie $H(p)$ to entropia skrzywionego rzutu monetą.

10.5.2 Twierdzenie Shannona

Mitzenmacher podaje zblefiony dowód. (Blef jest przy liczeniu $\mathbb{E}[W_0]$ – zbiory T_1, T_2 są zmiennymi losowymi, więc nie można ich wyciągnąć przed operator oczekiwanej).

My przedstawimy bardzo podobny, lecz poprawny dowód poniższego twierdzenia.

Twierdzenie 10.5.1 (Shannon, tylko że lepiej niż Mitzenmacher)**.**

$$\forall_{p\in\left(0,\frac{1}{2}\right)}\forall_{\delta,\gamma>0}\exists_{n_0}\forall_{n>n_0}:$$

1. $\forall_k : k \leq n\cdot(1-H(p)-\delta):$

$$\exists_{\text{Enc,Dec}}\forall_{m\in\{0,1\}^k} : P(\text{Dec}(\text{Channel}(\text{Enc}(m)))\neq m)\leq\gamma$$

2. Niech M będzie zmienną losową, która przyjmuje losową wiadomość jednostajnie. Wtedy:
$$\forall k \geq n\cdot(1-H(p)+\delta):$$

$$\forall_{\text{Enc,Dec}} P(\text{Dec}(\text{Channel}(\text{Enc}(M)))=M)\leq\gamma$$

Dowód. Pokażemy tylko pierwszą nierówność, drugą zostawiamy dla dociekliwego Czytelnika. Dowód zaczynamy tak jak robi to książka.

Na początku pokażemy coś słabszego – mianowicie, że jeśli wysyłamy losowe wiadomości z jednostajnym prawdopodobieństem to oczekiwany błąd jest mniejszy niż γ. Potem pokażemy jak wykazać nierówność postawioną w tezie.

Niech M będzie zmienną losową, która każdą wiadomość przyjmuje jednostajnie. Skonstruujemy takie funkcje Enc i Dec aby

$$P(\text{Dec}(\text{Channel}(\text{Enc}(M)))\neq M)\leq\gamma$$

Zaczynamy od wylosowania 2^k kodów n-bitowych - będą to kody wiadomości. Mamy więc zmienne losowe na każdą wiadomość – $\{X_m\}_{m\in\{0,1\}^k}$. Ponadto definiujemy $\text{Enc}(m)=X_m$.

Dec dostaje teraz wiadomość, która w oczekiwaniu ma pn przekłamanych bitów.

Niech s będzie wejściem dla dekodera, a ε dobrane wystarczająco małe.

Jeśli

$$\exists!_{m\in\{0,1\}^k} : \Delta(s,X_m)\in[(p-\varepsilon)n,(p+\varepsilon)n]$$

to dekoder jednoznacznie zwraca m. Warunek ten można ilustrować jako pierścień o szerokości 2ε wokół odpowiedniego punktu oraz pytanie czy w tym pierścieniu jest dokładnie jeden kod jakiejś wiadomości.

W przeciwnym razie albo nie mamy żadnej albo mamy wręcz za dużo opcji i zwracamy cokolwiek, ale będziemy zakładać że mamy pecha i zawsze zwracamy źle. Zastanówmy się kiedy tak się dzieje.

Okazuje się że istnieją dwie opcje na sprawienie aby warunek na poprawną odpowiedź był fałszywy.

1. Kanał wykonał za dużo przekłamań
2. Odebrana wiadomość jest pomiędzy dwoma możliwymi źródłami.

Możemy policzyć prawdopodobieństwo że jeśli do kanału wejdzie s_1 to kanał zwróci s_2. Wynosi ono dokładnie

$$w(s_1, s_2) = p^{\Delta(s_1,s_2)}(1-p)^{n-\Delta(s_1,s_2)}$$

Dla każdej pary $m \in \{0,1\}^k, s \in \{0,1\}^n$ definiujemy indykator

$$I_{m,s} = \begin{cases} 1 & \text{gdy } \operatorname{Dec}(s) \neq m \\ 0 & \text{gdy } \operatorname{Dec}(s) = m \end{cases}$$

który nam mówi, czy został popełniony błąd.

Dla dowolnej wiadomości $m \in \{0,1\}^k$ definiujemy prawdopodobieństwo W_m, że wysyłając m nie udało nam się odzyskać wiadomości:

$$W_m = \sum_{s \in \{0,1\}^n} w(X_m, s) \cdot I_{m,s}$$

Podkreślmy, że W_m jest zmienną losową, która jest zależna od X_m.

Będziemy chcieli policzyć $\mathbb{E}[W_m]$ po każdym możliwym wyborze kodów $X_1, \ldots, X_{2^k}$

Zdefiniujmy sobie najpierw dwa rodzaje zbiorów, które nam kategoryzują odbierane kody:

$$T_1(x) = \{s \in \{0,1\}^n \mid |\Delta(x,s) - pn| > \varepsilon n\}$$

$$T_2(x) = \{s \in \{0,1\}^n \mid |\Delta(x,s) - pn| \leq \varepsilon n\}$$

Teraz rozpisujemy sobie $\mathbb{E}[W_m]$ na osobne przypadki:

$$\begin{aligned} \mathbb{E}[W_m] &= \mathbb{E}\left[\sum_{s \in \{0,1\}^n} w(X_m, s) \cdot I_{m,s}\right] \\ &= \mathbb{E}\left[\sum_{s \in T_1(X_m)} w(X_m, s) \cdot I_{m,s}\right] + \mathbb{E}\left[\sum_{s \in T_2(X_m)} w(X_m, s) \cdot I_{m,s}\right] \end{aligned}$$

Będziemy chcieli oszacować każdy składnik osobno.

Zrobimy tutaj jednak coś innego niż książka, która wchodzi operatorem wartości oczekiwanej pod sumę, ignorując zupełnie fakt, że sumujemy się po czymś co jest zależne od zmiennych $X_1, \ldots, X_{2^k}$.

1. Pierwszy składnik.

Zaczniemy od oszacowania sumy pod wartością oczekiwaną przy ustalonych $X_1, \ldots, X_{2^k}$

Skoro $s \in T_1(X_m)$ to $I_{m,s} = 1$ bo z definicji T_1 mamy za dużo przekłamań aby odzyskać m, upraszczamy zatem to wyrażenie do

$$\sum_{s \in T_1(X_m)} w(X_m, s) \cdot I_{m,s} = \sum_{s \in T_1(X_m)} w(X_m, s)$$

Zauważamy, że możemy zapisać, trochę bardziej intuicyjnie

$$w(X_m, s) = P(\text{Channel}(X_m) = s)$$

przy czym X_m jest ustalone; prawdopodobieństwo liczymy jedynie względem zachowania kanału. Przepisujemy zatem wyrażenie i korzystamy z definicji $T_1(X_m)$

$$\begin{aligned}\sum_{s \in T_1(X_m)} w(X_m, s) &= \sum_{s \in T_1(X_m)} P(\text{Channel}(X_m) = s) \\ &= P(|\Delta(X_m, \text{Channel}(X_m)) - pn| > \varepsilon n)\end{aligned}$$

Zauważamy teraz fajną rzecz, mianowicie kanał wykonuje przekłamania niezależnie od tego co przez niego przechodzi. Niech D oznacza liczbę przekłamań wykonywanych przez kanał (na dowolnym wejściu).

Ograniczamy nierównością Chernoffa, gdzie $\mu = pn, \delta = \frac{\varepsilon}{p}$:

$$\begin{aligned}P(|\Delta(X_m, \text{Channel}(X_m)) - pn| > \varepsilon n) &= P(|D - pn| > \varepsilon n) \\ &\leq 2\exp\left(\frac{-n\varepsilon^2}{3p}\right)\end{aligned}$$

Dostaliśmy oszacowanie każdej konfiguracji przez stałą, możemy zatem oszacować wartość oczekiwaną od góry, a następnie wziąć n wystarczająco duże aby uzyskane oszacowanie było mniejsze niż $\frac{\gamma}{2}$.

$$\mathbb{E}\left[\sum_{s \in T_1(X_m)} w(X_m, s) \cdot I_{m,s}\right] \leq 2\exp\left(\frac{-n\varepsilon^2}{3p}\right) \leq \frac{\gamma}{2}$$

2. Drugi składnik.

W tym miejscu książka ponownie robi dziwne rzeczy z operatorem wartości oczekiwanej. Przeprowadzimy dość podobne rozumowanie, ale wykorzystamy do tego warunkową wartość oczekiwaną.

Rozpisujemy zatem

$$\begin{aligned}
\mathbb{E}\left[\sum_{s\in T_2(X_m)} w(X_m,s)\cdot I_{m,s}\right] &= \sum_{y_m\in\{0,1\}^n} \mathbb{E}\left[\sum_{s\in T_2(X_m)} w(X_m,s)\cdot I_{m,s} \mid X_m = y_m\right]\cdot P(X_m = y_m)\\
&= \sum_{y_m\in\{0,1\}^n} \mathbb{E}\left[\sum_{s\in T_2(y_m)} w(y_m,s)\cdot I_{m,s} \mid X_m = y_m\right]\cdot 2^{-n}\\
&= 2^{-n}\cdot \sum_{y_m\in\{0,1\}^n}\sum_{s\in T_2(y_m)} w(y_m,s)\cdot \mathbb{E}[I_{m,s}\mid X_m = y_m]\\
&= 2^{-n}\cdot \sum_{y_m\in\{0,1\}^n}\sum_{s\in T_2(y_m)} w(y_m,s)\cdot P(I_{m,s}=1\mid X_m = y_m)
\end{aligned}$$

Oszacujemy teraz

$$P(I_{m,s}=1\mid X_m = y_m)$$

Prawdopodobieństwo, że ustalony X_i, $i\neq m$ zapala nam indykator wynosi dokładnie

$$P(X_i\in T_2(y_m)) = \sum_{j=\lceil n(p-\varepsilon)\rceil}^{\lfloor n(p-\varepsilon)\rfloor} 2^{-n}\cdot\binom{n}{j}$$

Szacujemy od góry za pomocą entropii

$$\sum_{j=\lceil n(p-\varepsilon)\rceil}^{\lfloor n(p-\varepsilon)\rfloor}\binom{n}{j} \leq n\cdot 2^{nH(p+\varepsilon)}$$

Przez union bound szacujemy, że prawdopodobieństwo, że dowolny z innych kodów spowoduje nam kolizję, wynosi co najwyżej

$$(2^k-1)\cdot 2^{-n}\cdot n\cdot 2^{nH(p+\varepsilon)} \leq n\cdot \exp_2(k+n(H(p+\varepsilon)-1))$$

Dostaliśmy ładne oszacowanie, możemy je włożyć do wyjściowego wyrażenia aby otrzymać

$$\begin{aligned}
&2^{-n}\cdot \sum_{y_m\in\{0,1\}^n}\sum_{s\in T_2(y_m)} w(y_m,s)\cdot P(I_{m,s}=1\mid X_m=y_m)\\
&\leq 2^{-n}\cdot \sum_{y_m\in\{0,1\}^n}\sum_{s\in T_2(y_m)} w(y_m,s)\cdot n\cdot \exp_2(k+n(H(p+\varepsilon)-1))\\
&= 2^{-n}\cdot n\cdot \exp_2(k+n(H(p+\varepsilon)-1))\cdot \sum_{y_m\in\{0,1\}^n}\sum_{s\in T_2(y_m)} w(y_m,s)\\
&\leq 2^{-n}\cdot n\cdot \exp_2(k+n(H(p+\varepsilon)-1))\cdot 2^n\\
&= n\cdot \exp_2(k+n(H(p+\varepsilon)-1))
\end{aligned}$$

Teraz możemy skorzystać z faktu, że $k \leq (1 - H(p) - \delta)n$ i dostać oszacowanie od góry przez

$$n \cdot \exp_2 n(H(p+\varepsilon) - H(p) - \delta))$$

Dopieramy ε na tyle małe, aby $H(p+\varepsilon) - H(p) < \delta$. Wtedy wykładnik jest ujemny, zatem dla odpowiednio dużego n otrzymane wyrażenie jest mniejsze niż $\frac{\gamma}{2}$

Pokazaliśmy zatem, że jeśli dobierzemy odpowiednio małe ε oraz wystarczająco duże n to oba składniki są mniejsze od $\frac{\gamma}{2}$ a zatem

$$\mathbb{E}[W_m] \leq \gamma$$

Sumując te oczekiwania dla wszystkich możliwych wiadomości dostajemy

$$\mathbb{E}\left[\sum_{m=0}^{2^k} W_m\right] = \sum_{m=0}^{2^k} \mathbb{E}[W_m] \leq \gamma \cdot 2^k$$

Skoro oczekiwana sumy jest ograniczona przez $\gamma \cdot 2^k$ to musi istnieć konkretny zbiór kodów $x_1, \ldots, x_{2^k} \in \{0,1\}^n$ dla których

$$\sum_{m=0}^{2^k} W_m \leq \gamma \cdot 2^k$$

Jeśli wiadomość wysyłamy losowo z prawdopodobieństwem $\frac{1}{2^k}$ to oczekiwane prawdopodobieństwo błędu wynosi co najwyżej

$$\frac{1}{2^k} \sum_{m=0}^{2^k} W_m \leq \frac{1}{2^k} \gamma \cdot 2^k = \gamma$$

co zamyka dowód słabszej tezy.

Pokażemy teraz silniejszą wersję tezy (1) to znaczy, że istnieje takie kodowanie, że **dla każdej** wiadomości prawdopodobieństwo porażki wynosi co najwyżej γ.

Ustalmy $\gamma' = \frac{\gamma}{2}, \delta' = \delta + \varepsilon$, gdzie ε jest wybrane dowolnie małe.

Pokazaliśmy, że dla γ', δ' jeśli $k \leq n \cdot (1 - H(p) - \delta')$ to istnieje taki wybór kodów $x_1, \ldots x_{2^k}$, że

$$\sum_{m=0}^{2^k} W_m \leq \gamma' \cdot 2^k$$

Bez straty ogólności możemy posortować x_i rosnąco po W_i.

Zauważmy teraz, że dla $i \leq 2^{k-1}$ musi zachodzić

$$W_i \leq 2\gamma'$$

Istotnie – w przeciwnym razie $W_{2^{k-1}} > 2\gamma'$ a co za tym idzie

$$\sum_{m=2^{k-1}+1}^{2^k} W_m > 2^{k-1} \cdot 2\gamma' = 2^k \gamma'$$

co jest sprzeczne z własnością wybranego kodowania.

W takim razie możemy wykorzystać $x_1, \ldots, x_{2^{k-1}}$ jako kodowanie dla wiadomości $k-1$ bitowych, i tym samym dostajemy

$$\forall_k : k-1 \leq n(1-H(p)-\delta') : \exists_{\text{Enc,Dec}} \forall_{m \in \{0,1\}^k} : P(\text{Dec}(\text{Channel}(\text{Enc}(m))) \neq m) \leq 2\gamma'$$

Odwińmy teraz oznaczenia:

$$\forall_k : k \leq n\left(1-H(p)-\delta-\varepsilon+\frac{1}{n}\right) : \exists_{\text{Enc,Dec}} \forall_{m \in \{0,1\}^k} : P(\text{Dec}(\text{Channel}(\text{Enc}(m))) \neq m) \leq \gamma$$

Dla dostatecznie dużych n mamy $\frac{1}{n} < \varepsilon$ a zatem

$$k \leq n\left(1-H(p)-\delta-\varepsilon+\frac{1}{n}\right) \leq n(1-H(p)-\delta)$$

Co pozwala nam ostatecznie stwierdzić, że dla γ, δ

$$\exists_{n_0} \forall_{n>n_0} \forall_k : k \leq n(1-H(p)-\delta) : \exists_{\text{Enc,Dec}} \forall_{m \in \{0,1\}^k} : P(\text{Dec}(\text{Channel}(\text{Enc}(m))) \neq m) \leq \gamma$$

co należało pokazać. □

Rozdział 11

Sprzęgania

11.1 Sprzęganie rozkładów

Definicja 11.1.1. 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 11.1.1. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S. Niech $B = \{x \in S : \mu(x) \geq \nu(x)\}$. Zachodzi

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

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

Lemat 11.1.2. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S. Zachodzi

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

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

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

□

Lemat 11.1.3. Niech μ, ν będą rozkładami prawdopodobieństwa nad skończonym zbiorem S. Zachodzi

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

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

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

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

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

która daje

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

□

Definicja 11.1.2. 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 11.1.4. 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 żądaną nierówność.

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

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

11.2 Sprzęganie łańcuchów

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

Definicja 11.2.1. Definiujemy

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

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

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

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

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

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

Lemat 11.2.1. Niech $((X_t, Y_t))_{t\in\mathbb{N}}$ będzie sprzęganiem łańcuchów (skończonych, nieprzywiedlnych, nieokresowych) z macierzą przejścia P i zbiorem stanów S. Niech $T \in \mathbb{N}$ i $\varepsilon > 0$ będą

takie, że dla każdego $x, y \in S$ zachodzi

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

Wtedy czas mieszania łańcucha z macierzą P jest ograniczony:

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

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

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

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

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

Mamy zatem

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

a z tego wynika

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

□

Lemat 11.2.2 (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

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

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

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

□

Twierdzenie 11.2.1 (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

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

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 $((X_t, Y_t))_{t\in\mathbb{N}}$ będzie sprzęganiem łańcuchów o macierzy przejścia P^r zadanym 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 $\left(\frac{1}{p_3}\min(u(s), \nu(s)) : s \in S\right)$. 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 sprzęganie, 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 $j \in \{0, \ldots, r-1\}$ możemy zapisać

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

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

11.3 Przykłady sprzęgań

Twierdzenie 11.3.1. Rozważmy graf G o n wierzchołkach i stopniu maksymalnym Δ. Definiujemy następujący łańcuch Markowa: niech M_0 będzie ustalonym k-elementowym zbiorem niezależnym w G. W kroku łańcucha losujemy wierzchołek v jednostajnie z M_t i w jednostajnie

z $V(G)$. Ustalamy $M_{t+1} = M_t \setminus \{v\} \cup \{w\}$, jeśli taki zbiór jest niezależny i ma k wierzchołków. Inaczej $M_{t+1} = M_t$.

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

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

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

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

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

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

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

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

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

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

bo w pierwszym przypadku musimy wybrać wspólne v i takie w, że dokładnie jeden z łańcuchów zmieni stan – szacujemy z góry przez sumę sąsiedztw niewspólnych wierzchołków z obu łańcuchów (może być tak, że te sąsiedztwa się pokrywają i oba nie zmienią stanu, ale to szacowanie z góry więc jest dobrze). W drugim przypadku musimy wybrać różne v, v' i takie w, że oba

łańcuchy zmienią stan – odejmujemy największe możliwe sąsiedztwa wszystkich wierzchołków w obu łańcuchach poza v i v'. Teraz zakładając $d_t > 0$ możemy przeliczyć

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

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

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

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

Z nierówności Markowa

$$P\left(d_t \geq 1\right) \leq \mathbb{E}\left[d_t\right] \leq k e^{-tC} \implies P\left(X_t \neq Y_t\right) \leq k e^{-tC}.$$

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

Twierdzenie 11.3.2. Rozważmy graf G o n wierzchołkach i stopniu maksymalnym Δ. Definiujemy następujący łańcuch Markowa: niech M_0 będzie ustalonym poprawnym c-kolorowaniem wierzchołkowym G. W kroku łańcucha losujemy wierzchołek v jednostajnie z $V(G)$ i kolor ℓ jednostajnie z $[c]$. Jeśli $N(v)$ nie zawiera wierzchołków koloru ℓ, to zmieniamy kolor v na ℓ. Inaczej łańcuch stoi w miejscu.

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

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

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

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

Weźmy dowolną bijekcję $f : [c] \to [c]$ taką, że

$$\begin{array}{ll} f(S_X(v)) \subseteq S_Y(v) & \text{jeśli } |S_X(v)| \leq |S_Y(v)| \\ S_X(v) \supseteq f^{-1}(S_Y(v)) & \text{jeśli } |S_X(v)| > |S_Y(v)|. \end{array}$$

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

Oznaczmy $d_t = |D_t|$ i $d'(v) = \begin{cases} |N(v) \cap A_t|, & v \in D_t \\ |N(v) \cap D_t|, & v \in A_t \end{cases}$. Zauważmy, że mamy $\sum_{v\in A_t} d'(v) = m' = \sum_{v\in D_t} d'(v)$, gdzie m' jest liczbą krawędzi między wierzchołkami D_t a A_t. Zachodzi

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

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

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

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

Zatem mamy

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

czyli

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

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

Rozdział 12

Metoda Monte Carlo

12.1 Definicje

Definicja 12.1.1. Zmienna losowa X (randomizowany algorytm) daje (ε, δ)-aproksymację wartości V, jeśli

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

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

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

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

Definicja 12.1.2. Mając dany problem obliczeniowy $x \to R\left(x\right)$ mówimy, że mamy w pełni wielomianowy randomizowany schemat aproksymacji (FPRAS – fully polynomial randomized approximation scheme), jeśli dla wejścia x i parametrów ε, δ umiemy generować (ε, δ)-aproksymację $R\left(x\right)$ w czasie wielomianowym od $x, \frac{1}{\varepsilon}, \ln\left(\frac{1}{\delta}\right)$.

Mamy problem $x \to R\left(x\right)$ i chcemy znaleźć jego FPRAS. W tym celu projektujemy zmienną losową X o $\mathbb{E}\left[X\right] = R\left(x\right)$, powtarzamy ją tak dużo razy, aż ograniczenie Czernowa da nam odpowiednią aproksymację.

Zakładając, że X jest indykatorem, chcemy iterować zmienną $m \geq \frac{3\ln\left(\frac{2}{\delta}\right)}{R(x)\varepsilon^2}$ razy. Ta liczba jest wielomianowa od $\frac{1}{\varepsilon}$ i $\ln\left(\frac{1}{\delta}\right)$, ale $\frac{1}{R(x)}$ niekoniecznie jest wielomianowe od x. Jeśli jednak jest, to znaleźliśmy dobrą aproksymację.

Definicja 12.1.3. Zmienna losowa X (randomizowany algorytm) na skończonej przestrzeni Ω

daje ε-jednostajną próbkę Ω, jeśli dla każdego $S \subseteq \Omega$ jest

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

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

Definicja 12.1.4. Mając dany problem obliczeniowy $x \to \Omega\left(x\right)$ (gdzie $\Omega\left(x\right)$ to zbiór rozwiązań) mówimy, że mamy w pełni wielomianowy prawie jednostajny schemat próbkowania (FPAUS – fully polynomial almost uniform sampler), jeśli dla wejścia x i parametru ε umiemy generować ε-jednostajną próbkę $\Omega\left(x\right)$ w czasie wielomianowym od $x, \ln\left(\frac{1}{\varepsilon}\right)$.

12.2 Przykłady

12.2.1 DNF

Definicja 12.2.1. Dysjunktywna postać normalna (DNF – disjunctive normal form) formuły boolowskiej to postać, która jest alternatywą klauzul składających się z koniunkcji literałów, np.

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

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

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

Twierdzenie 12.2.1. Mając zadaną formułę w postaci DNF $F = C_1 \vee C_2 \vee \ldots \vee C_t$ o n zmiennych możemy wyznaczyć liczbę wartościowań ją spełniających $C\left(F\right)$ za pomocą następującego algorytmu: pozbywamy się klauzul zawierających wyrażenia postaci $x \wedge \overline{x}$ i definiujemy SC_i jako zbiór wartościowań spełniających C_i. Do tego $S = \{(i, a) : i \in [t], a \in SC_i\}$, $RC_i = SC_i \setminus \bigcup_{j=1}^{i-1} SC_j$, $R = \{(i, a) : i \in [t], a \in RC_i\}$.

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

Dowód. Zauważmy, że pozbycie się klauzul zawierających $x \wedge \overline{x}$ można wykonać liniowym przejściem po klauzulach. Jeśli C_i ma ℓ_i literałów (czyli ℓ_i zmiennych), to spełnia ją dokładnie $2^{n-\ell_i}$ wartościowań (ustalamy zmienne w klauzuli, reszta dowolnie). Mamy $|S| = \sum_{i=1}^{t} |SC_i| =$

$\sum_{i=1}^{t} 2^{n-\ell_i}$, a więc tę liczbę da się łatwo wyznaczyć. RC_i jest zbiorem wartościowań, które spełniają C_i, ale nie spełniają żadnej poprzedniej klauzuli. Jasne, że $C(F) = |R|$ (każde wartościowanie liczymy raz). Nasz algorytm właściwie losuje parę $(i, a) \in S$ i sprawdza, czy $(i, a) \in R$. Co więcej, robi to jednostajnie: najpierw wybieramy i z prawdopodobieństwem $\frac{|SC_i|}{|S|}$, a następnie a z prawdopodobieństwem $\frac{1}{|SC_i|}$ – razem $\frac{1}{|S|}$. Zauważmy, że wylosowanie a da się zrobić łatwo – losujemy jednostajnie wartości literałów nie ustalonych przez C_i. Dla $R(F) = \frac{C(F)}{|S|} = \frac{|R|}{|S|}$ mamy $\mathbb{E}[X] = R(F)$. Aby dostać (ε, δ)-aproksymację musimy powtórzyć losowanie $\frac{3\ln\left(\frac{2}{\delta}\right)}{R(F)\varepsilon^2}$ razy. Mamy $\frac{1}{R(F)} = \frac{|S|}{|R|} \leq t$, bo każde wartościowanie z R występuje w S co najwyżej t razy. Zatem $\frac{1}{R(F)}$ jest wielomianowe od F i faktycznie mamy FPRAS. □

12.2.2 FPRAS z FPAUS

Twierdzenie 12.2.2. Niech $\Omega(G)$ oznacza rodzinę zbiorów niezależnych w grafie G. Zakładając, że mamy dla niej FPAUS, istnieje FPRAS jej rozmiaru.

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

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

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

Teraz możemy przeliczyć

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

gdzie pierwsza nierówność jest konsekwencją ciągu nierówności

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

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

Z tego wynika, że jeśli znajdziemy FPRAS dla wartości $R = \frac{|\Omega(H)|}{|\Omega(H')|}$, gdzie $H = H' + e$ dla pewnej

krawędzi e, to uzyskamy naszą tezę. Wykorzystamy do tego FPAUS. Bierzemy $\frac{\varepsilon'}{3}$-jednostajnie zbiór $I \in \Omega(H')$ i ustalamy indykator $X = [I \in \Omega(H)]$. Powtarzamy taki eksperyment m razy: $Y = \frac{1}{m}\sum_{i=1}^{m} X_i$. Niech $\mu = \mathbb{E}[Y]$.

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

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

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

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

Rozpisując lewą stronę nierówności dostajemy

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

i podobnie dla prawej strony. Mamy zatem

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

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

12.2.3 FPAUS ze sprzęgania

Twierdzenie 12.2.3. Rozważmy graf G z $\Delta(G) \leq 4$, w którym do tego nie ma trójkątów (W książce *Probability and Computing* nie ma tego założenia. Jest ono jednak potrzebne w części dowodu, w której rozważamy przypadki.). Niech $\Omega(G)$ będzie rodziną zbiorów niezależnych w G. Istnieje FPAUS dla $\Omega(G)$.

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

$p \in \{1,2,3\}$. Teraz mamy przypadki

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

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

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

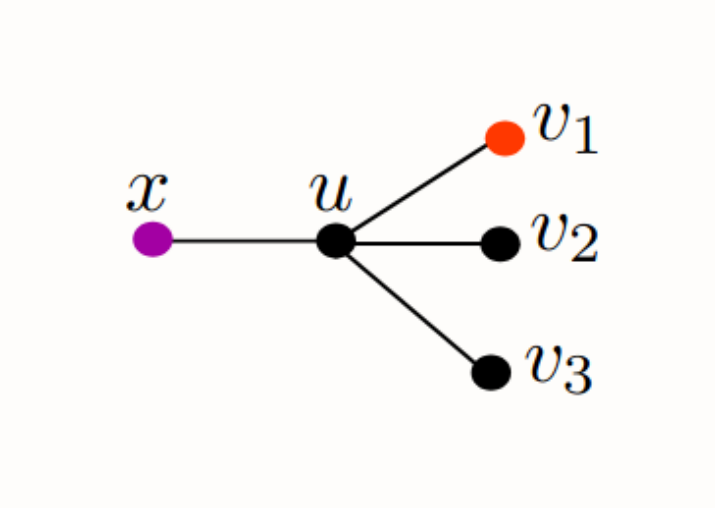
Jeśli $d_t = 0$, to $d_{t+1} = 0$. Rozważmy teraz przypadek $d_t = 1$ – mamy $I, J \in \Omega(G)$ takie, że $J = I \cup \{x\}$ i chcemy pokazać, że dla $D = |K(I) \Delta K(J)|$ zachodzi $\mathbb{E}[D] \leq 1$. Niech $e = uv$. Jeśli $u, v \notin N(x)$, to $D = 1$, bo oba zbiory zachowają się tak samo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

1. $|N(u) \cap I| \geq 2$. Tutaj zawsze $D \leq 1$, co sprawdzamy rozpatrując wszystkie przypadki.

2. $|N(u) \cap I| = 1$. Patrząc na rysunek: jeśli $v \in \{v_2, v_3\}$, to $D = 1$, bo w obu zbiorach zrobimy to samo (m.in. tu ważny jest brak trójkątów – v nie jest połączone z x). Jeśli $v = x$, to $p \in \{1,2\} \to D = 0$ (usuwamy albo dodajemy x do obu) a $p = 3 \to D = 1$, bo ruch się nie uda. Jeśli $v = v_1$, to $p \in \{1,2\} \to D = 1$, ale $p = 3 \to D = 3$, bo ruch nie uda się dla J, ale uda się dla I. Widzimy, że wartość oczekiwana D w tej sytuacji to 1.

3. $|N(u) \cap I| = 0$. Jeśli $v = x$, to $D = 0$. Jeśli $v = v_i$ dla $v_i \in N(u) \setminus \{x\}$, to $p \in \{1,2\} \to D = 1$, ale $p = 3 \to D = 2$, bo u zostanie dodane do I. Mamy

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



Rysunek 12.1: Przypadek, gdy I ma w sobie jednego sąsiada u.

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

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

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

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

a więc

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

To daje nam

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

Ostatecznie

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

□