

Spis treści

Wykład 1 (Kelthur, Kacper Orszulak)	2
Wykład 2 (Adam, Mikołaj)	8
Wykład 3 (Autor/ka)	20
Wykład 4 (Autor/ka)	20
Wykład 5 (Autor/ka)	20
Wykład 6 (Autor/ka)	20
Wykład 7 (Autor/ka)	20
Wykład 8 (Autor/ka)	20
Wykład 9 (Autor/ka)	20
Wykład 10 (Autor/ka)	20
Wykład 11 (Autor/ka)	20
Wykład 12 (Autor/ka)	20
Wykład 13 (Autor/ka)	20

Wykład 1 (Kelthur, Kacper Orszulak)

Przestrzeń Probabilistyczna

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

- Ω to zbiór zdarzeń elementarnych.
- $\mathcal{F} \subseteq 2^\Omega$ to rodzina zdarzeń mierzalnych.
W przypadku dyskretnym $\mathcal{F} = 2^\Omega$.
W przypadku ogólnym, $\mathcal{F}$ musi być σ -ciałem (również σ -algebra)¹, czyli:
 - $\emptyset \in \mathcal{F}$
 - $X \in \mathcal{F} \Rightarrow \Omega - X \in \mathcal{F}$
 - $\forall_{i \in I} X_i \in \mathcal{F} \Rightarrow \cup_{i \in I} X_i \in \mathcal{F}$ (Gdzie I jest przeliczalne)
 - $\forall_{i \in I} X_i \in \mathcal{F} \Rightarrow \cap_{i \in I} X_i \in \mathcal{F}$
- $P : \mathcal{F} \rightarrow \mathbb{R}$ t. że:
 - $\forall_{E \in \mathcal{F}} 0 \leq P(E) \leq 1$
 - $P(\Omega) = 1$
 - $\forall_{\{E_i\}_{E_i \in \mathcal{F}}}$ przeliczalna rodzina $P(\cup_i E_i) = \sum_i P(E_i)$
 E_i parami rozłączne

Lemat:

Dla dowolnych $E_1, E_2 \in \mathcal{F}$:

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

Lemat (prawd. sumy / union bound):

Dla dowolnej skończonej lub przeliczalnej rodziny zbiorów $E_1, E_2, \dots \forall_i E_i \in \mathcal{F}$
 $P(\cup E_i) \leq \sum P(E_i)$

Lemat (zasada włączeń i wyłączeń):

Dla dowolnych $E_1, \dots, E_n \in \mathcal{F}$

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

Def. (zdarzenia niezależne):

zbiory $E_1, E_2 \in \mathcal{F}$ są niezależne, jeśli $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2)$

Def (zdarzenia niezależne w ogólności):

zbiory $E_1, \dots, E_k \in \mathcal{F}$ są niezależne, jeśli

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

¹Całkiem przystępna definicja na polskiej Wikipedii.

Def. (*Prawdopodobieństwo warunkowe*):
zdarzenie E pod warunkiem zdarzenia F :

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

Jeżeli E i F są niezależne:

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

Przykłady algorytmów randomizowanych

Weryfikacja równości wielomianów

Problem - mamy algorytm sprawdzający równość wielomianów. Np. czy:

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

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

Nasz algorytm:

- wybierz r losowo (każdy z równym prawdopodobieństwem - *rozkład jednostajny - uniform distribution*) ze zbioru $\{1, \dots, 100d\}$
- oblicz $F(r)$ i $G(r)$ (gdzie F i G to nasze dwa wielomiany na wejściu)
- jeśli $F(r) \neq G(r)$, to zwróć **false**
- jeśli $F(r) = G(r)$, to zwróć **true**

Jeżeli algorytm zwróci **false**, to na pewno F i G są różne. Jeżeli zwróci **true**, to pomyli się tylko wtedy, gdy r jest pierwiastkiem $F(x) - G(x)$, a tych jest co najwyżej d . Możemy się więc spodziewać, że prawdopodobieństwo, że algorytm się pomyli jest równe: $p = \frac{d}{100d} = \frac{1}{100}$

Bardziej formalnie:

$$\Omega = \{1, \dots, 100d\}$$

E - zdarzenie takie, że algorytm się pomylił

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

Mamy pierwszą nierówność, bo nie wszystkie pierwiastki muszą być w zbiorze $\{1, \dots, 100d\}$.
Mamy drugą, bo może być mniej niż d pierwiastków.

Jeżeli puścimy ten algorytm kilka razy:

$$E_1, \dots, E_k$$

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

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

Weryfikacja wyniku mnożenia macierzy

Mamy trzy macierze $n \times n$, A, B, C , nad zbiorem $\{0, 1\}$ (arytmetyka mod 2).

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

Normalne mnożenie działa w $O(n^3)$

Algorytm sprawdzający:

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

Jeżeli $AB \neq C$ i $\bar{r}$ jest wybrane losowo i jednostajnie, to $P(A(B\bar{r}) = C\bar{r}) \leq \frac{1}{2}$

Dowód:

Niech $D = AB - C \neq 0$. Zakładamy $d_{11} \neq 0$

$$D\bar{r} = 0 \Rightarrow \sum_{j=1}^n d_{1j}r_j = 0 \Rightarrow r_1 = \frac{-\sum_{j=2}^n d_{1j}r_j}{d_{11}}$$

Zmienne losowe

Definicja. *Zmienna losowa*

X jest zmienną losową przestrzeni $(\Omega, \mathcal{F}, \mathcal{P})$ jeśli $X : \Omega \rightarrow \mathbb{R}$ taka że $\forall_{F \in \mathcal{F}} X(F) \in \mathcal{B}(\mathbb{R})$.²

Definicja. *Dyskretna zmienna losowa*

Zmienna losowa, która przyjmuje (w obrazie) przeliczalnie wiele wartości.

Przykład. Rzut dwa razy symetryczną kostką.

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

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

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

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

Definicja. Zmienne losowe X i Y są *niezależne* jeśli

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

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

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

²Zbiór Borelowski.

³Wystarczy iterować się po obrazie X i Y .

Definicja. Wartość oczekiwana zmiennej X , oznaczana $E(X)$, to ⁴

$$E(X) = \sum_{i \text{ po obrazie } X} i \cdot \mathcal{P}(X = i)$$

Dla co najwyżej przeliczalnych obrazów wartość oczekiwana jest określona wtw. szereg $\sum |i| \mathcal{P}(X = i)$ jest zbieżny.

Twierdzenie. Liniowość wartości oczekiwanej (variant z sumą)

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

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

Dowód. ($n = 2$)

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

Twierdzenie. Liniowość wartości oczekiwanej (variant ze współczynnikiem)

$$\forall_{\substack{c \in \mathbb{R} \\ X \text{ zmienna losowa}}} E(cX) = c \cdot E(X).$$

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

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

Przykład. X zmienna losowa reprezentująca sumę oczek na dwóch kostkach.

Z definicji bardzo trudno jest policzyć

$$E(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \dots + 12 \cdot \frac{1}{36} = 7.$$

⁴W tej części kursu zakładamy, że obraz jest przeliczalny. Ogólnie sumowanie continuum obiektów jest problematyczne.

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

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7.$$

Przykład.

Urna 4 kulki. Każde innego koloru. 4 razy wyjmujemy kule, obserwujemy kolor i wrzucamy kulę z powrotem. Jaka jest oczekiwana liczba zaobserwowanych kolorów?

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

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

Jeżeli Y - indykator, to $E(Y) = 0 \cdot \mathcal{P}(Y = 0) + 1 \cdot \mathcal{P}(Y = 1) = \mathcal{P}(Y = 1)$.

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

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

Przykład. Mrówki chodzące po kiju.

Zacznijmy od zagadki:

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

Można zauważyć, że mrówki są nierozróżnialne, a więc gdy dwie mrówki się spotykają, możemy przyjąć, iż się nie odbijają, a po prostu przenikają się wzajemnie i idą dalej. Zatem odpowiedź to 100s.

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

Niech X = liczba odbić = liczba przeniknięć. Ponadto niech $X_{i,j}$ = liczba przeniknięć i -tej i j -tej mrówki $\in \{0, 1\}$. Zakładamy, że mrówki są posortowane i dla $i < j$ mrówka i -ta jest przed j -tą. Zauważmy, że $E(X_{i,j}) = 0 \cdot \mathcal{P}(X_{i,j} = 0) + 1 \cdot \mathcal{P}(X_{i,j} = 1) = \mathcal{P}(X_{i,j} = 1) = \frac{1}{4}$.

Wtedy

$$\begin{aligned} X &= \sum_{1 \leq i < j \leq 25} X_{i,j} \\ E(X) &= E\left(\sum X_{i,j}\right) = \sum E(X_{i,j}) \\ &= \sum \mathcal{P}(X_{i,j} = 1) = \sum \frac{1}{4} = \binom{25}{2} \frac{1}{4}. \end{aligned}$$

Wykład 2 (Adam, Mikołaj)

Wartość oczekiwana - ciąg dalszy

$E(X^2)$ vs $E(X)^2$

Dowodzimy, że $E(X^2) \geq E(X)^2$.

Przykład.

niech X - zmienna losowa

$$P(X = n) = \frac{1}{99}, n \in \{1, \dots, 99\}$$

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

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

Dowód.

Niech Y - zmienna losowa

$$Y = (X - E(X))^2 \geq 0$$

Wtedy:

$$\begin{aligned} E(Y) &= E((X - E(X))^2) = E(X^2 - 2 \cdot E(X) \cdot X + E(X)^2) \\ &= E(X^2) - 2 \cdot E(X) \cdot E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \\ &\geq 0 \end{aligned}$$

Fakt. *Wartość oczekiwana funkcji*

$$E(f(x)) = \sum_{z \text{ po obrazie } f(X)} P(f(X) = z) \cdot z = \sum_{z \text{ po obrazie } X} P(X = z) \cdot f(z)$$

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

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

Nierówność Jensena.

Jeśli f wypukła i $E(X) < \infty$ to $E(f(x)) \geq f(E(X))$

Dowód. Zakładamy, że f ma rozwinięcie Taylora.

Niech $\mu = E(X)$

Z Taylora wiemy, że:

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

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

Definicja. *Indykator*

X jest indykatorem zdarzenia A , jeśli:

$$X = I_A = \begin{cases} 1, & \text{jeśli zdarzenie } A \text{ zaszło} \\ 0 & \text{wpp.} \end{cases}$$

Wtedy zachodzi:

$$E(X) = P(X = 0) \cdot 0 + P(X = 1) \cdot 1 = P(X = 1) = P(A)$$

Definicja. *Rozkład dwumianowy*

Zmienna losowa X ma rozkład dwumianowy typu (n, p) ($n \geq 1, p \in (0, 1)$), jeśli:

$$P(X = j) = \binom{n}{j} \cdot p^j \cdot (1 - p)^{n-j}, j \in [n]$$

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

Fakt. Funkcja f może być rozkładem, gdy jest nieujemna i sumuje się do 1 na całej dziedzinie. Możemy sprawdzić, czy rozkład dwumianowy jest poprawny:

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

Tak więc ten rozkład jest poprawny.

Wartość oczekiwana zmiennej losowej o rozkładzie dwumianowym:

$$E(X) = np$$

Dowód.

Z definicji:

$$\begin{aligned}
 E(X) &= \sum_{j=0}^n j \cdot P(X = j) \\
 &= \sum_{j=0}^n j \cdot \frac{n!}{j! \cdot (n-j)!} \cdot p^j \cdot (1-p)^{n-j} \\
 &= np \cdot \sum_{j=1}^n \frac{(n-1)!}{(j-1)! \cdot (n-1-(j-1))!} \cdot p^{j-1} \cdot (1-p)^{n-1-(j-1)} \\
 &= np \cdot (p + (1-p))^{n-1} \\
 &= np
 \end{aligned}$$

Można to też zrobić prościej⁵:

$$\begin{aligned}
 X &= \sum_{i=1}^n X_i, \quad X_i = \begin{cases} 1, & \text{sukces w } i\text{-tej próbie} \\ 0 & \text{wpp} \end{cases} \\
 E(X) &= E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n P(X_i = 1) = \sum_{i=1}^n p = np
 \end{aligned}$$

Warunkowa wartość oczekiwana

Definicja.

$$E(Y|Z = z) \stackrel{\text{def}}{=} \sum_y y \cdot P(Y = y|Z = z)$$

Przykład.

X_1, X_2 - wyniki niezależnych rzutów kostką 6-ścienną

$$X = X_1 + X_2$$

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

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

⁵Podobno jest to pułapka na kolokwium, nie ma co liczyć z definicji

⁶Sumę po y rozumiemy jako sumę po obrazie Y

Lemat.

X, Y - zmienne losowe

$$E(X) = \sum_y P(Y = y) \cdot E(X|Y = y)$$

Dowód.

$$\begin{aligned} \sum_y P(Y = y) \cdot E(X|Y = y) &= \sum_y P(Y = y) \cdot \sum_x x \cdot P(X = x|Y = y) \\ &= \sum_y \sum_x x \cdot P(Y = y) \cdot \frac{P(X = x \cap Y = y)}{P(Y = y)} \\ &= \sum_x x \cdot \sum_y P(X = x \cap Y = y) \\ &= \sum_x x \cdot P(X = x) \\ &= E(X) \end{aligned}$$

Fakt.

$$E\left(\sum_i X_i | Y = y\right) = \sum_i E(X_i | Y = y)$$

Czyli liniowość zachodzi też dla warunkowej wartości oczekiwanej

Definicja.

$E(Y|Z)$ to funkcja zmiennej Z (czyli zmienna losowa) taka, że dla zmiennych losowych Y, Z :

$$E(Y|Z)(\omega) = E(Y|Z = Z(\omega))$$

Przykład.

X_1, X_2 - wyniki niezależnych rzutów kostką 6-ścienną

$$X = X_1 + X_2$$

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

A więc $E(X|X_1) = X_1 + 3,5$

Lemat.

Dla dowolnych zdarzeń Y, Z :

$$E(Y) = E(E(Y|Z))$$

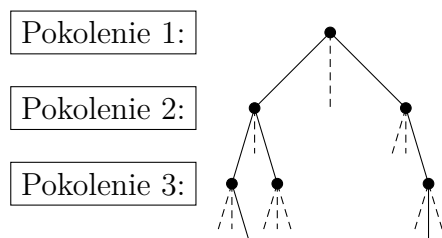
Dowód.

$$E(E(Y|Z)) = \sum_z P(Z = z) E(Y|Z = z) = E(Y)$$

Przykład. Proces gałązkowy.

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

Przykład dla $n = 3$



Jaka jest oczekiwana liczba wszystkich odpalonych programów?

Y_i - liczba programów w i -tym pokoleniu

$Y_0 = 0$, Y_1 ma rozkład dwumianowy z parametrami (n, p)

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

$$E(Y_i | Y_{i-1}) = Y_{i-1} np$$

Teraz, dowodzimy indukcyjnie, że $E(Y_i) = (np)^i$.

$$E(Y_1) = np$$

$$E(Y_i) = E(E(Y_i | Y_{i-1})) = E(Y_{i-1} np) = np E(Y_{i-1}) = np (np)^{i-1} = (np)^i$$

⁷Ponieważ $E(Y|Z)$ jest funkcją zmiennej Z , możemy rozbić wartość oczekiwaną względem obrazu Z

A następnie obliczamy wynik:

$$E\left(\sum_{i=0}^{\infty} Y_i\right) = {}^8 \sum_{i=0}^{\infty} E(Y_i) = \sum_{i=0}^{\infty} (np)^i$$

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

Rozkład geometryczny

Definicja.

X ma rozkład geometryczny z parametrem $p \in (0, 1)$ gdy:

$$\forall_{n \in \mathbb{N}_+} P(X = n) = (1 - p)^{n-1} p$$

Rozkład geometryczny jest **bez pamięci (memoryless)** - dokonane do danej chwili próby nie mają wpływu na dalszy rozkład X . Formalnie:

$$P(X = n + k | X > k) = P(X = n)$$

Dowód:

$$P(X = n + k | X > k) = \frac{P(X = n + k \cap X > k)}{P(X > k)} = \frac{(1 - p)^{n+k-1} p}{(1 - p)^k} = (1 - p)^{n-1} p = P(X = n)$$

Lemat.

X - zmienna losowa przyjmująca wartości całkowite, nieujemne

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

Dowód.

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

Wartość oczekiwana zmiennej z rozkładem geometrycznym.

Jeśli X ma rozkład geometryczny z parametrem p , to $E(X) = \frac{1}{p}$

Dowód.

$$E(X) = \sum_{i=1}^{\infty} P(X \geq i) = \sum_{i=1}^{\infty} (1 - p)^{i-1} = \frac{1}{p}$$

⁸Według Micka można to zrobić, o ile prawa strona równości jest zbieżna

Alternatywna metoda:

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

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

$$pE(X) = 1 \implies E(X) = \frac{1}{p}$$

Przykład. Problem kolekcjonera kuponów

Kupujemy paczki, w każdej z których jest jeden z n kuponów. Prawdopodobieństwo tego, że i -ty kupon jest w danej paczce wynosi $\frac{1}{n}$. Jaka jest oczekiwana liczba paczek które musimy kupić przed uzyskaniem wszystkich kuponów?

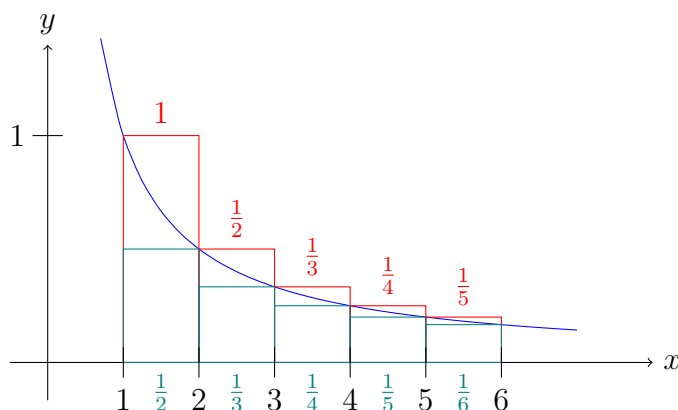
X - liczba kupionych paczek do uzyskania n kuponów

X_i - liczba kupionych paczek posiadając $i-1$ kuponów (paczki kupione próbując zdobyć i -ty kupon)

$X_1 = 1$, X_i ma rozkład geometryczny z parametrem równym $\frac{n-(i-1)}{n}$

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \frac{n}{n-(i-1)} = n \sum_{i=1}^n \frac{1}{n} = nH_n \approx n \cdot \ln(n)$$

Dowód. $H_n \approx \ln(n)$



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

Przykład. Analiza algorytmu Quicksort w wersji randomizowanej

Jaka jest oczekiwana liczba porównań algorytmu Quicksort, w którym każda wartość ma równe prawdopodobieństwo na bycie wybranym jako pivot?

Niech $y_1, \dots, y_n$ to będą posortowane wartości na wejściu Quicksorta.

X - liczba wykonanych porównań

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

Quicksort każdą parę porównuje co najwyżej raz, a więc:

$$X = \sum_{1 \leq i < j \leq n} X_{i,j} \implies E(X) = E\left(\sum_{1 \leq i < j \leq n} X_{i,j}\right) = \sum_{1 \leq i < j \leq n} E(X_{i,j}) = \sum_{1 \leq i < j \leq n} P(X_{i,j} = 1)$$

Musimy teraz tylko obliczyć $P(X_{i,j} = 1)$



Jeśli pivot zostanie wybrany w $[y_1, y_i)$ albo $(y_j, y_n]$, to y_i i y_j dalej będą w tej samej instancji Quicksorta. Za to jeśli jako pivot zostanie wybrane coś w (y_i, y_j) , to y_i nigdy nie zostanie porównane z y_j . Jediną możliwością na ich porównanie jest to, że kiedy po raz pierwszy jako pivot zostanie wybrana wartość w $[y_i, y_j]$, będzie to y_i albo y_j .

Niech Q to w takim razie pierwsza instancja Quicksorta w której zostanie wybrany element w $[y_i, y_j]$. Dodatkowo niech Z to przedział liczb, które Q dostaje na wejściu. Interesuje nas wartość:

$$P(Q \text{ wybrał jako pivot } y_i \text{ lub } y_j | Q \text{ wybrał pivota z } [y_i, y_j]) = \frac{\frac{2}{|Z|}}{\frac{j-i+1}{|Z|}} = \frac{2}{j-i+1}$$

A więc:

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

Zostaje teraz tylko wyliczenie $E(X)$:

$$\begin{aligned} E(X) &= \sum_{1 \leq i < j \leq n} P(X_{i,j} = 1) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{2}{j-i+1} \\ &= \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} = \sum_{k=2}^n \sum_{i=1}^{n-k+1} \frac{2}{k} = \sum_{k=2}^n (n+1-k) \frac{2}{k} \\ &= \sum_{k=2}^n \frac{2(n+1)}{k} - 2(n-1) = 2(n+1)(H_n - 1) - 2(n-1) \\ &= 2n \ln(n) + \theta(n) \end{aligned}$$

Nierówność Markowa

Jest to pierwsza *nierówność koncentrująca* omawiana na kursie.

X - zmienna losowa przyjmująca wartości nieujemne

$$\forall a > 0 \quad P(X \geq a) \leq \frac{E(X)}{a}$$

Dowód

ustalmy $a > 0$ $X \geq 0$

ustalmy indykator $I = \begin{cases} 1, & X \geq a \\ 0 & \text{wpp} \end{cases}$

$$X \geq 0 \implies I \leq \frac{X}{a}$$

Jeżeli $I = 1$, to $X \geq a \implies \frac{X}{a} \geq 1$, natomiast dla $I = 0$, $\frac{X}{a}$ jest dodatnie, ponieważ zarówno X jak i a są nieujemne.

Stąd mamy:

$$P(X \geq a) = P(I = 1) = E(I) \leq E\left(\frac{X}{a}\right) = \frac{E(X)}{a}$$

Czym w ogóle jest koncentracja prawopodobieństwa? Pytamy się, jak prawdopodobne jest, że zmienna przyjmie wartość bliską wartości oczekiwanej. Koncentrujemy zmienną wokół jej wartości oczekiwanej, jeżeli jesteśmy w stanie stwierdzić, że to prawdopodobieństwo jest duże, bądź prawdopodobieństwo odchylenia jest małe.

Przykład.

Rzucamy n razy monetą - zliczamy liczbę orłów.

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

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

$$E(X) = E\left(\sum_i X_i\right) = \sum_i E(X_i) = \frac{n}{2}$$

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

Z Markowa:

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

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

Wariancja i momenty zmiennych losowych

Definicja. *Moment zmiennej losowej*

k -ty moment ($k \geq 1$) zmiennej losowej X to:

$$E(X^k)$$

W szczególności wartość oczekiwana jest pierwszym momentem zmiennej losowej.

Definicja. *Wariancja zmiennej losowej*

Wariancja zmiennej X to:

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

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

Operator wariancji nie jest liniowy.

Definicja. *Odchylenie standardowe zmiennej losowej*

Odchylenie standardowe zmiennej X to:

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

Przykład.

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

Wartość oczekiwana X :

$$E(X) = c \cdot P(X = c) = c \cdot 1 = c$$

Wariancja:

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

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

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

Wartość oczekiwana Y :

$$E(Y) = 0 \cdot \frac{k-1}{k} + k \cdot c \cdot \frac{1}{k} = c$$

Wariancja Y :

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

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

Definicja. Kowariancja

Kowariancja zmiennych losowych X i Y to:

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

Lemat.

Dla dowolnych zmiennych losowych X, Y zachodzi:

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

Dowód.

Rozpisujemy z definicji wariancję sumy:

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

Lemat.

Dla dowolnych *niezależnych* zmiennych losowych X i Y :

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

Dowód.

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

Wniosek.

Dla dowolnych *niezależnych* zmiennych losowych X i Y :

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

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Dowód.

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

Korzystamy tutaj z wyżej dowiedzonego faktu o wartości oczekiwanej iloczynu.

Przykład. *Wariancja zmiennej o rozkładzie dwumianowym*

X - zmienna o rozkładzie dwumianowym z parametrami (n, p)

$$E(X) = np$$

$$\begin{aligned} E(X^2) &= \sum \binom{n}{j} \cdot p^j (1-p)^{n-j} \cdot j^2 \\ &= \sum \frac{n!}{j! \cdot (n-j)!} \cdot p^j (1-p)^{n-j} \cdot j \cdot (j-1) + \sum \frac{n!}{j! \cdot (n-j)!} \cdot p^j (1-p)^{n-j} \cdot j \\ &= n(n-1)p^2 \sum \frac{(n-2)!}{(j-2)! \cdot (n-j)!} p^{j-2} (1-p)^{n-j} + np \sum \frac{(n-1)!}{(j-1)! \cdot (n-j)!} p^{j-1} (1-p)^{n-j} \\ &= n(n-1)p^2 \cdot (p + (1-p))^{n-2} + np \cdot (p + (1-p))^{n-1} \\ &= n(n-1)p^2 + np \end{aligned}$$

W drugim przekształceniu chcemy otrzymać $j \cdot (j-1)$ zamiast j^2 . Żeby zachować równość dodajemy naszą sumę z czynnikiem j zamiast j^2 .

Teraz możemy policzyć wariancję:

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

To było liczenie z definicji. Ponieważ kolejne próby są niezależne, można obliczyć to prościej:

$$\text{Var}(X) = \text{Var}\left(\sum_i X_i\right) = \sum_i \text{Var}(X_i)$$

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

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

Sumując po i dostajemy:

$$\sum_i \text{Var}(X_i) = \sum_i (1-p)p = np(1-p)$$

Wykład 3 (Autor/ka)

Wykład 4 (Autor/ka)

Wykład 5 (Autor/ka)

Wykład 6 (Autor/ka)

Wykład 7 (Autor/ka)

Wykład 8 (Autor/ka)

Wykład 9 (Autor/ka)

Wykład 10 (Autor/ka)

Wykład 11 (Autor/ka)

Wykład 12 (Autor/ka)

Wykład 13 (Autor/ka)