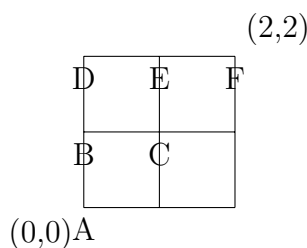


Zadanie 1 (Mateusz Wojaczek)

Wiewiórka chodzi po skończonej kracie jak na rysunku poniżej w sposób losowy. Rozpoczyna swoją trasę w wierzchołku $(0,0)$ i w każdym kroku wybiera w sposób losowy, jednostajny i niezależny jeden z sąsiednich wierzchołków aktualnego wierzchołka. W wierzchołku $(1,1)$ znajduje się żółędź. Należy wyznaczyć oczekiwaną liczbę kroków potrzebnych, aby wiewiórka dotarła do żółędzia.

Rozwiązanie

Ze względu na symetrię wystarczy rozważyć połowę kwadratu:



Rozwiązanie

Oznaczmy oczekiwane liczby kroków z poszczególnych wierzchołków odpowiednio:

$$A = E(0,0), \quad B = E(1,0), \quad C = E(1,1), \quad D = E(2,0), \quad E = E(2,1), \quad F = E(2,2)$$

Ponieważ wierzchołek $(2,2)$ zawiera żółędź, mamy

$$F = 0.$$

Dla pozostałych wierzchołków zapisujemy równania dla wartości oczekiwanej:

$$\begin{cases} A = 1 + B, \\ B = 1 + \frac{1}{3}A + \frac{1}{3}C + \frac{1}{3}D, \\ C = 1 + \frac{1}{2}B + \frac{1}{2}E, \\ D = 1 + \frac{1}{2}B + \frac{1}{2}C, \\ E = 1 + \frac{1}{3}C + \frac{1}{3}D + \frac{1}{3}F. \end{cases}$$

Ponieważ $D = C$, układ upraszcza się do:

$$\begin{cases} A = 1 + B, \\ B = 1 + \frac{1}{3}A + \frac{2}{3}C, \\ C = 1 + \frac{1}{2}B + \frac{1}{2}E, \\ E = 1 + \frac{2}{3}C. \end{cases}$$

Z ostatniego równania otrzymujemy:

$$E = 1 + \frac{2}{3}C.$$

Podstawiając do trzeciego równania:

$$C = 1 + \frac{1}{2}B + \frac{1}{2}\left(1 + \frac{2}{3}C\right),$$

czyli

$$C = \frac{3}{2} + \frac{1}{2}B + \frac{1}{3}C.$$

Stąd

$$\frac{2}{3}C = \frac{3}{2} + \frac{1}{2}B \Rightarrow C = \frac{9}{4} + \frac{3}{4}B.$$

Podstawiając do równania dla B :

$$B = 1 + \frac{1}{3}A + \frac{2}{3}C.$$

Z $A = 1 + B$ mamy:

$$B = 1 + \frac{1}{3}(1 + B) + \frac{2}{3}C,$$

czyli

$$\frac{2}{3}B = \frac{4}{3} + \frac{2}{3}C \Rightarrow B = 2 + C.$$

Podstawiamy $C = \frac{9}{4} + \frac{3}{4}B$:

$$B = 2 + \frac{9}{4} + \frac{3}{4}B,$$

czyli

$$\frac{1}{4}B = \frac{17}{4} \Rightarrow B = 17.$$

Stąd:

$$C = \frac{9}{4} + \frac{3}{4} \cdot 17 = \frac{9 + 51}{4} = 15, \quad A = 1 + B = 18.$$

Oczekiwana liczba kroków, jakie wiewiórka wykona zanim znajdzie żołędzia, wynosi **18**.

Odpowiedź

18

Zadanie 2 (Hubert Jastrzębski)

Rzucamy kostką, aż nie zanotujemy dwóch szóstek pod rząd. Jaka jest oczekiwana liczba rzutów?

Wskazówka: Odpowiedzią nie jest 36.

Rozwiązanie

Zasada chyba (bo jeszcze nikt go nie spisał) podobna do tej z poprzedniego zadania, ale nawet trochę prostsza.

Generalnie będziemy chcieli stworzyć układ równań, który pozwoli nam wyliczyć oczekiwaną liczbę rzutów – rozpisałem się trochę (+ w vs code ciężko czasem nie kliknąć tego tab), ale samo zadanie nie jest skomplikowane.

Zdefiniujmy sobie trzy stany:

- Stan N (neutralny) – stan w którym w ostatnim rzucie nie wypadła szóstka (albo jest to pierwszy rzut).
- Stan S (jedna szóstka) – stan w którym w ostatnim rzucie wypadła szóstka, ale nie mamy jeszcze dwóch szóstek pod rząd.
- Stan K (dwie szóstki, koniec) – stan w którym mamy już dwie szóstki pod rząd.

Niech X będzie zmienną losową, oznaczającą liczbę rzutów potrzebnych do zakończenia gry.

Niech:

- $E_N = E(X | \text{jesteśmy w stanie } N)$ – oczekiwana liczba rzutów do zakończenia gry, jeśli zaczynamy w stanie N ,
- $E_S = E(X | \text{jesteśmy w stanie } S)$ – oczekiwana liczba rzutów do zakończenia gry, jeśli zaczynamy w stanie S ,
- $E_K = E(X | \text{jesteśmy w stanie } K)$ – oczekiwana liczba rzutów do zakończenia gry, jeśli zaczynamy w stanie K .

Oczywiście $E_K = 0$, bo gra się już skończyła.

Rozważmy teraz stan N . W tym stanie rzucamy kostką:

- Z prawdopodobieństwem $\frac{1}{6}$ wypadła szóstka i przechodzimy do stanu S .
- Z prawdopodobieństwem $\frac{5}{6}$ nie wypadła szóstka i pozostajemy w stanie N .

W każdym przypadku wykonaliśmy jeden rzut, więc możemy zapisać równanie:

$$E_N = \frac{1}{6}(E_S + 1) + \frac{5}{6}(E_N + 1)$$

Podobnie jest ze stanem S :

- Z prawdopodobieństwem $\frac{1}{6}$ wypadła szóstka i przechodzimy do stanu K .
- Z prawdopodobieństwem $\frac{5}{6}$ nie wypadła szóstka i wracamy do stanu N .

Znowu wykonaliśmy jeden rzut, więc mamy:

$$E_S = \frac{1}{6}(E_K + 1) + \frac{5}{6}(E_N + 1)$$

Teraz pozostaje tylko rozwiązać układ równań (bardzo ciężki):

$$\begin{aligned} E_N &= \frac{1}{6}(E_S + 1) + \frac{5}{6}(E_N + 1) \\ &= \frac{1}{6}\left(\frac{1}{6}(E_K + 1) + \frac{5}{6}(E_N + 1) + 1\right) + \frac{5}{6}(E_N + 1) \\ &= \frac{1}{6}\left(\frac{1}{6} + \frac{5}{6}E_N + \frac{5}{6} + 1\right) + \frac{5}{6}(E_N + 1) \\ &= \frac{2}{6} + \frac{5}{36}E_N + \frac{5}{6}E_N + \frac{5}{6} \\ &= \frac{7}{6} + \frac{35}{36}E_N \\ E_N &= \frac{7}{6} \cdot \frac{36}{1} = 42 \end{aligned}$$

Jako iż N to jest stan z którego rozpoczynamy grę, to odpowiedzią jest właśnie $E_N = 42$.

Odpowiedź

42

Zadanie 3 (Dominik)

Ze standardowej talii 52 kart wyciągamy 13 (ręka brydżowa). Niech X to liczba wyciągniętych asów, a Y to liczba wyciągniętych trefli.

- (i) Czy X i Y są niezależne?
- (ii) Czy X i Y są nieskorelowane?

Przypomnienie

Zmienne losowe X i Y są **niezależne**, gdy dla każdego $x, y \in \mathbb{R}$:

$$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$$

Zmienne losowe X i Y są **nieskorelowane**, gdy:

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

Rozwiązanie podpunktu (i)

Nie są niezależne, pokażę kontrprzykład.

- $P(X = 4)$ to prawdopodobieństwo, że wyciągniemy 4 asy, oczywiście $P(X = 4) > 0$.
- $P(Y = 13)$ to prawdopodobieństwo, że wyciągniemy tylko trefle, podobnie $P(Y = 13) > 0$.
- $P(X = 4 \wedge Y = 13)$ to prawdopodobieństwo, że wyciągniemy tylko trefle, w tym cztery asy. Takie zdarzenie jest niemożliwe, więc $P(X = 4 \wedge Y = 13) = 0$.

Czyli nie jest prawdą, że dla każdego $x, y \in \mathbb{R}$ zachodzi

$$P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y),$$

więc zmienne losowe X i Y nie są niezależne.

Rozwiązanie podpunktu (ii)

Niech X_i to indykator zdarzenia, że i -ta wylosowana karta to as:

$$E[X_i = 1] = \frac{4}{52} = \frac{1}{13} \quad (\text{dla } i \in [13])$$

Z liniowości wartości oczekiwanej:

$$E[X] = E[X_1 + \dots + X_{13}] = E[X_1] + \dots + E[X_{13}] = 13 \cdot \frac{1}{13} = 1$$

Analogicznie niech Y_i to indykator zdarzenia, że i -ta karta to trefl:

$$E[Y] = \dots = 13 \cdot \frac{1}{4} = \frac{13}{4}$$

$X_i Y_i$ to indykator zdarzenia, że i -ta wylosowana karta to as trefl: $E[X_i Y_i] = \frac{1}{52}$.

Dla $i \neq j$ $X_i Y_j$ to indykator zdarzenia, że i -ta wylosowana karta jest asem, a j -ta to trefl. Te zdarzenia są niezależne, więc:

$$E[X_i Y_j] = P(X_i = 1 \wedge Y_j = 1) = P(X_i = 1) \cdot P(Y_j = 1) = \frac{1}{4} \cdot \frac{1}{13} = \frac{1}{52}$$

Czyli dla każdej pary $(i, j) \in [13] \times [13]$ zachodzi $E[X_i Y_i] = \frac{1}{52}$.

Z liniowości wartości oczekiwanej:

$$E[XY] = E\left[\sum_{(i,j) \in [13] \times [13]} X_i Y_j\right] = \sum_{(i,j) \in [13] \times [13]} E[X_i Y_j] = 13^2 \cdot \frac{1}{52} = \frac{13}{4},$$

Wiemy też, że:

$$E[X] \cdot E[Y] = 1 \cdot \frac{13}{4} = \frac{13}{4},$$

więc warunek na **nieskorelowanie** zmiennych losowych X i Y zachodzi.

Odpowiedź

X i Y nie są niezależne, ale **są nieskorelowane**

Zadanie 4 (Michał)

Każdy z n graczy rzuca niezależnie od pozostałych kostką.

- (i) Za dowolną parę graczy którzy wyrzucili ten sam wynik grupa dostaje 1 punkt. Znaleźć wariancję liczby uzyskanych punktów.
- (ii) Za dowolną parę graczy którzy wyrzucili ten sam wynik grupa dostaje tyle punktów ile wyrzucili. Znaleźć wariancję liczby uzyskanych punktów.

Wstęp

Niech X_i będzie wynikiem rzutu i -tego gracza, dla $i = 1, \dots, n$. Zmienne X_i są niezależne i mają jednakowy rozkład: $P(X_i = k) = \frac{1}{6}$ dla $k \in \{1, 2, 3, 4, 5, 6\}$.

Będziemy korzystać ze wzoru na wariancję sumy zmiennych losowych:

$$\text{Var}\left(\sum_k A_k\right) = \sum_k \text{Var}(A_k) + \sum_{k \neq l} \text{Cov}(A_k, A_l)$$

Rozwiązanie podpunktu (i)

Definiujemy zmienną losową Y jako całkowitą liczbę punktów. Liczba punktów jest sumą punktów przyznanych każdej parze graczy. Jest $\binom{n}{2}$ takich par.

Dla każdej pary graczy (i, j) z $1 \leq i < j \leq n$, zdefiniujemy indykator Y_{ij} :

$$Y_{ij} = \begin{cases} 1 & \text{gdy } X_i = X_j \\ 0 & \text{gdy } X_i \neq X_j \end{cases}$$

Całkowita liczba punktów Y to oczywiście suma tych zmiennych:

$$Y = \sum_{1 \leq i < j \leq n} Y_{ij}$$

Aby znaleźć $\text{Var}(Y)$, musimy obliczyć $\text{Var}(Y_{ij})$ oraz $\text{Cov}(Y_{ij}, Y_{kl})$ dla $(i, j) \neq (k, l)$.

Wariancja

Wartość oczekiwana: $E[Y_{ij}] = p = \frac{1}{6}$.

Wariancja: $\text{Var}(Y_{ij}) = p(1-p) = \frac{1}{6} \left(1 - \frac{1}{6}\right) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$.

Kowariancja

Dla par rozłącznych Kowariancja jest równa 0 zastanówmy co dzieje się dla par mających jeden wspólny wskaźnik. BSO weźmy pary (i, j) i (i, k) , wtedy

$$P(Y_{ij} = 1 \text{ i } Y_{ik} = 1) = P(\text{wszyscy trzej gracze mają ten sam wynik}) = \frac{1}{36}$$

stąd

$$\text{Cov}(Y_{ij}, Y_{ik}) = \frac{1}{36} - \left(\frac{1}{6}\right)^2 = 0$$

zatem wszystkie kowariancje pomiędzy dwoma różnymi Y_{ij} są równe 0

A zatem

wariancja sumy jest po prostu sumą wariancji:

$$\text{Var}(Y) = \sum_{1 \leq i < j \leq n} \text{Var}(Y_{ij})$$

Liczba par (i, j) wynosi $\binom{n}{2} = \frac{n(n-1)}{2}$.

$$\text{Var}(Y) = \binom{n}{2} \cdot \text{Var}(Y_{ij}) = \frac{n(n-1)}{2} \cdot \frac{5}{36} = \frac{5n(n-1)}{72}$$

Rozwiązanie podpunktu (ii)

Definiujemy zmienną losową Y jako całkowitą liczbę punktów. Dla każdej pary graczy (i, j) z $1 \leq i < j \leq n$, zdefiniujemy zmienną Y_{ij} :

$$Y_{ij} = \begin{cases} X_i & \text{gdy } X_i = X_j \\ 0 & \text{gdy } X_i \neq X_j \end{cases}$$

Moment pierwszego i drugiego rzędu

$$[Y_{ij}] = \sum_{k=1}^6 k \cdot \frac{1}{36} = \frac{21}{36} = \frac{7}{12},$$

$$[Y_{ij}^2] = \sum_{k=1}^6 k^2 \cdot \frac{1}{36} = \frac{91}{36},$$

$$\text{Var}(Y_{ij}) = \frac{91}{36} - \left(\frac{7}{12}\right)^2 = \frac{35}{16}.$$

Kowariancja dwóch zmiennych

Jeśli pary są rozłączne, kowariancja = 0. Jeśli pary mają wspólnego gracza (Y_{ij} i Y_{ik}):

$$[Y_{ij}Y_{ik}] = \sum_{k=1}^6 k^2 \cdot \frac{1}{216} = \frac{91}{216},$$

$$\text{Cov}(Y_{ij}, Y_{ik}) = \frac{91}{216} - \left(\frac{7}{12}\right)^2 = \frac{35}{432}.$$

Liczba par wskaźników ze wspólnym graczem

Liczba uporządkowanych par o wspólnym wskaźniku (takich potrzebujemy do sumy) to:

$$2 \cdot 3 \cdot \binom{n}{3} = n(n-1)(n-2).$$

Wariancja całkowita

Wiemy, że

$$Y = \sum_{1 \leq i < j \leq n} Y_{ij}.$$

Wtedy

$$\begin{aligned} \text{Var}(Y) &= \sum_{\{i,j\}} \text{Var}(Y_{ij}) + \sum_{(i<j) \neq (k<l)} \text{Cov}(Y_{ij}, Y_{ik}) \\ &= \binom{n}{2} \cdot \frac{35}{16} + n(n-1)(n-2) \cdot \frac{35}{432} \\ &= \frac{35 n(n-1)}{32} + \frac{35 n(n-1)(n-2)}{432} \\ &= 35 \cdot \frac{n(n-1)(2n+23)}{864}. \end{aligned}$$

Odpowiedź

(i) $\binom{n}{2} \cdot \frac{5}{36} = \frac{5n(n-1)}{72}$ (ii) $\frac{35 n(n-1)(2n+23)}{864}$
--

Zadanie 5 (Filip Manijak)

W urnie znajduje się b białych kul i c czarnych kul. Wyciągamy kolejno wszystkie kule.

Wyciągnięta kula daje *zmianę*, jeżeli jej kolor jest różny od koloru kuli wyciągniętej bezpośrednio przed nią.

Wyciągnięta biała kula daje *dobrą zmianę*, jeżeli bezpośrednio przed nią została wyciągnięta czarna kula.

Oblicz:

- oczekiwaną liczbę i wariancję *zmian*,
- oczekiwaną liczbę i wariancję *dobrych zmian*.

Rozwiązanie

Oczekiwana liczba zmian i wariancja

Niech w urnie jest b białych kul i c czarnych kul, $N = b + c$. Losujemy wszystkie kule kolejno bez zwracania.

Dla $i = 2, \dots, N$ definiujemy wskaźniki:

$$X_i = \begin{cases} 1, & \text{jeśli } i\text{-ta kula ma inny kolor niż } (i-1)\text{-ta,} \\ 0, & \text{w przeciwnym razie,} \end{cases}$$

$$Y_i = \begin{cases} 1, & \text{jeśli } i\text{-ta kula jest biała i poprzednia czarna (dobra zmiana),} \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

I) Zmiany koloru

Całkowita liczba zmian koloru:

$$C = \sum_{i=2}^N X_i.$$

Oczekiwana liczba zmian

Dla dowolnego i :

$$\Pr(X_i = 1) = \Pr(B \rightarrow W) + \Pr(W \rightarrow B) = \frac{bc}{N(N-1)} + \frac{bc}{N(N-1)} = \frac{2bc}{N(N-1)}.$$

Zatem:

$$\mathbb{E}[C] = (N-1) \cdot \frac{2bc}{N(N-1)} = \frac{2bc}{N}.$$

Wariancja — pełna

Każdy wskaźnik X_i to zmienna Bernoulliego:

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

Ogólna formuła wariancji sumy:

$$\text{Var}(C) = \sum_{i=2}^N \text{Var}(X_i) + 2 \sum_{2 \leq i < j \leq N} \text{Cov}(X_i, X_j).$$

1. Kowariancje sąsiadujące ($|i - j| = 1$) Niech X_i i X_{i+1} mają wspólną kulę w pozycji i . Wtedy:

$$E[X_i X_{i+1}] = \Pr(X_i = 1, X_{i+1} = 1) = \frac{bc(b-1) + bc(c-1)}{N(N-1)(N-2)} = \frac{bc}{N(N-1)},$$

więc

$$\text{Cov}(X_i, X_{i+1}) = \frac{bc}{N(N-1)} - p^2 = \frac{bc}{N(N-1)} - \left(\frac{2bc}{N(N-1)} \right)^2.$$

Liczba takich par: $(N-2)$.

2. Kowariancje niesąsiadujące ($|i - j| > 1$) Niech X_i i X_j nie sąsiadują. Wtedy mamy 4 różne kule w pozycjach $i-1, i, j-1, j$. Zatem:

$$E[X_i X_j] = \Pr(X_i = 1, X_j = 1) = \frac{2bc}{N(N-1)} \cdot \frac{2(b-1)(c-1)}{(N-2)(N-3)}.$$

Liczba takich par: $\binom{N-1}{2} - (N-2) = \frac{(N-1)(N-2)}{2} - (N-2) = \frac{(N-2)(N-3)}{2}$.

Zatem kowariancja dla każdej pary niesąsiadującej:

$$\text{Cov}(X_i, X_j) = \frac{4bc(b-1)(c-1)}{N(N-1)(N-2)(N-3)} - p^2.$$

3. Wariancja całkowita Sumując:

$$\text{Var}(C) = (N-1)p(1-p) + 2(N-2) \left(\frac{bc}{N(N-1)} - p^2 \right) + 2 \cdot \frac{(N-2)(N-3)}{2} \left(\frac{4bc(b-1)(c-1)}{N(N-1)(N-2)(N-3)} - p^2 \right).$$

Po uproszczeniu otrzymujemy:

$$\text{Var}(C) = \frac{2bc(2bc - N)}{N^2(N-1)}.$$

II) Dobre zmiany

Dla $i = 2, \dots, N$ definiujemy:

$$Y_i = \begin{cases} 1, & \text{jeśli } i\text{-ta kula jest biała i poprzednia czarna (dobra zmiana),} \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

Całkowita liczba dobrych zmian:

$$G = \sum_{i=2}^N Y_i.$$

Oczekiwana liczba

$$\Pr(Y_i = 1) = \frac{bc}{N(N-1)}, \quad \mathbb{E}[G] = (N-1) \frac{bc}{N(N-1)} = \frac{bc}{N}.$$

Wariancja

Każdy wskaźnik Y_i to zmienna Bernoulliego:

$$\text{Var}(Y_i) = p(1-p), \quad p = \frac{bc}{N(N-1)}.$$

Dla sąsiadujących wskaźników Y_i i Y_{i+1} :

$$\text{Cov}(Y_i, Y_{i+1}) = \mathbb{E}[Y_i Y_{i+1}] - \mathbb{E}[Y_i] \mathbb{E}[Y_{i+1}] = 0 - p^2 = -p^2,$$

ponieważ zdarzenia są rozłączne (kula i nie może jednocześnie być biała i czarna).

Dla niesąsiadujących wskaźników Y_i, Y_j ($|i-j| > 1$) kowariancja wynosi:

$$\text{Cov}(Y_i, Y_j) = \Pr(Y_i = 1 \text{ i } Y_j = 1) - p^2.$$

Wariancja całkowitej liczby dobrych zmian:

$$\text{Var}(G) = \sum_{i=2}^N \text{Var}(Y_i) + 2 \sum_{2 \leq i < j \leq N} \text{Cov}(Y_i, Y_j).$$

Po uwzględnieniu wszystkich kowariancji (sąsiadujących i niesąsiadujących) otrzymujemy:

$$\text{Var}(G) = \frac{b^2 c^2}{N^2(N-1)}.$$

Zadanie 6 (Jakub Stepaniuk)

Niech X będzie zmienną losową liczącą liczbę kopii kliki K_4 na czterech wierzchołkach w grafie wylosowanym z modelu $G_{n,p}$ (to znaczy, mając $n \geq 4$ wierzchołków dla każdej pary wierzchołków losujemy niezależnie z prawdopodobieństwem $p \in (0, 1)$, czy są połączone krawędzią, czy nie). Oblicz wartość oczekiwaną i wariancję X .

Rozwiązanie

Wartość oczekiwana jest prosta do wyliczenia. Wystarczy zdefiniować indykatory $\{X_I\}_{I \in \binom{V}{4}}$ dla każdego 4-elementowego podzbioru wierzchołków, które są równe 1 wpp. gdy ten zbiór indukuje kopię K_4 . Otrzymujemy wartość oczekiwaną:

$$E[X] = E\left[\sum_{I \in \binom{V}{4}} X_I\right] = \sum_{I \in \binom{V}{4}} E[X_I] = \sum_{I \in \binom{V}{4}} p^6 = \binom{n}{4} p^6$$

Z wariancją jest nieco więcej zabawy, ale na szczęście możemy zauważyć, że indykatory zbiorów o 0 lub 1 wspólnych wierzchołkach byłyby rozłączne krawędziowo, więc byłyby niezależne (a więc $Cov(X_I, X_J) = 0$). Możemy więc wyliczyć wariancję pierwsze poznając wartości kowariancji dla przecinających się krawędziowo zbiorów.

2 wspólne wierzchołki

Niech $X, Y \in \binom{V}{4}, |X \cap Y| = 2$.

$$\begin{aligned} Cov(X_I, X_J) &= E[X_I X_J] - E[X_I] E[X_J] = \\ &= p^{11} - (p^6)^2 = (1 - p)p^{11} \end{aligned}$$

3 wspólne wierzchołki

Niech $X, Y \in \binom{V}{4}, |X \cap Y| = 3$

$$\begin{aligned} Cov(X_I, X_J) &= E[X_I X_J] - E[X_I] E[X_J] = \\ &= p^9 - (p^6)^2 = (1 - p^3)p^9 \end{aligned}$$

Wariancje

Pozostaje jeszcze tylko wyliczyć wariancję pojedynczego indykatora:

$$Var(X_I) = E[X_I^2] - E[X_I]^2 = p^6 - p^{12} = (1 - p^6)p^6$$

I ostateczny wynik:

$$\begin{aligned}
 \text{Var}(X) &= \text{Var}\left(\sum_{I \in \binom{V}{4}} X_I\right) = \sum_{I \in \binom{V}{4}} \text{Var}(X_I) + \sum_{I, J \in \binom{V}{4}, I \neq J} \text{Cov}(X_I, X_J) \\
 &= \sum_{I \in \binom{V}{4}} \text{Var}(X_I) + \sum_{I, J \in \binom{V}{4}, I \cap J = 2} \text{Cov}(X_I, X_J) + \sum_{I, J \in \binom{V}{4}, I \cap J = 3} \text{Cov}(X_I, X_J) \\
 &= \sum_{I \in \binom{V}{4}} \text{Var}(X_I) + \sum_{I, J \in \binom{V}{4}, I \cap J = 2} \text{Cov}(X_I, X_J) + \sum_{I, J \in \binom{V}{4}, I \cap J = 3} \text{Cov}(X_I, X_J) \\
 &= \binom{n}{4} (1 - p^6) p^6 + \binom{n}{6} \binom{6}{2} \binom{4}{2} (1 - p) p^{11} + \binom{n}{5} \binom{5}{3} \binom{2}{1} (1 - p^3) p^9 \\
 &= \binom{n}{4} (1 - p^6) p^6 + 90 \binom{n}{6} (1 - p) p^{11} + 20 \binom{n}{5} (1 - p^3) p^9
 \end{aligned}$$

Odpowiedź

$$E[X] = \binom{n}{4} p^6, \text{Var}[X] \text{ powyżej}$$

Zadanie 7 (Wiktor Marć)

Powiedzmy, że w urnie znajduje się N kul z czego K jest zielonych a reszta czerwone. Losujemy n kul bez zwracania. Niech X będzie liczbą zielonych kul które wylosowaliśmy. Jaki rozkład ma X ? Innymi słowy, ile wynosi $P(X = k)$?

Rozwiązanie

k kul zielonych spośród wszystkich kul zielonych można wybrać na $\binom{K}{k}$ sposobów, analogicznie $n - k$ kul czerwonych na $\binom{N-K}{n-k}$ sposobów. Wszystkich sposobów na wybranie n kul ze zbioru N kul jest $\binom{N}{n}$. Wynik wynosi więc:

$$\frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

Odpowiedź

$$\frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}, \text{ jest to rozkład hipergeometryczny}$$

Zadanie 8 (Krzysztof Peszko)

Jaka jest wartość oczekiwana i wariancja zmiennej losowej X z poprzedniego zadania.

Rozwiązanie

Robimy zdarzenie X_i , czyli, że i -ta kula jest zielona. Z liniowości wartości oczekiwanej, wiemy, że $E(X) = \sum_{i=0}^n E(X_i)$.

$E(X_i) = P(X_i) = \frac{K}{N}$, więc $E(X) = n * \frac{K}{N}$.

Teraz obliczmy $E(X^2)$, możemy to rozpisać, jako $E(X(X-1) + X) = E(X * (X-1)) + E(X)$.

Teraz powiedzmy, sobie czym jest $E(X * (X-1))$. Będzie to liczba uporządkowanych par, takich, że obie kule są zielone.

Wiemy, że takich par będzie $2 * \binom{K}{2}$, a prawdopodobieństwo, że konkretna para się znajdzie to $\frac{\binom{N-2}{n-2}}{\binom{N}{n}}$. Także korzystając z liniowości, wychodzi nam, że:

$$E(X * (X-1)) = 2 * \binom{K}{2} * \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{K * (K-1) * (n) * (n-1)}{(N) * (N-1)}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{K * (K-1) * n * (n-1)}{N * (N-1)} + n * \frac{K}{N} - (n * \frac{K}{N})^2$$

$$Var(X) = \frac{n * K}{N} * (\frac{(K-1) * (n-1)}{(N-1)} + 1 - (n * \frac{K}{N}))$$

$$Var(X) = \frac{n * K}{N} * (\frac{(K-1) * (n-1)}{(N-1)} + \frac{N-1}{N-1} - (n * \frac{K}{N}))$$

$$Var(X) = \frac{n * K}{N} * (\frac{(K-1) * (n-1) + (N-1)}{(N-1)} - (n * \frac{K}{N}))$$

$$Var(X) = \frac{n * K}{N} * (\frac{K * (n-1) - n + N}{(N-1)} - (n * \frac{K}{N}))$$

$$Var(X) = \frac{n * K}{N} * \frac{N * K * (n-1) - n * N + N^2 - n * K * (N-1)}{N * (N-1)}$$

$$Var(X) = \frac{n * K}{N} * \frac{K * (N * (n-1) - n * (N-1)) - n * N + N^2}{N * (N-1)}$$

$$Var(X) = \frac{n * K}{N} * \frac{K * (-N + n) - n * N + N^2}{N * (N-1)}$$

$$Var(X) = \frac{n * K * (N - n)}{N} * \frac{-K + N}{N * (N-1)}$$

$$Var(X) = \frac{n * K * (N - n)}{N * (N-1)} * (1 - \frac{K}{N})$$

Odpowiedź

$$E(X) = n * \frac{K}{N}, \text{Var}(X) = \frac{n * K * (N - n)}{N * (N - 1)} * (1 - \frac{K}{N})$$

Zadanie 9 (Pavlo Tsikalysyn)

Pokaż, że nierówność Markowa jest optymalna. To znaczy, dla danego $a > 1$ znajdź zmienną losową X , że:

$$P(X \geq aE(X)) = \frac{1}{a}$$

Rozwiązanie

Niech dla naszej zmiennej losowej $P(X = 0) = \frac{a-1}{a}$ oraz $P(X = a) = \frac{1}{a}$.

Wtedy $E(X) = P(X = 0) \cdot 0 + P(X = a) \cdot a = \frac{a-1}{a} \cdot 0 + \frac{1}{a} \cdot a = 1$,

stąd mamy, że $P(X \geq aE(X)) = P(X \geq a) = P(X = a) = \frac{1}{a}$.

Odpowiedź

$$P(X = 0) = \frac{a-1}{a} \text{ oraz } P(X = a) = \frac{1}{a}$$

Zadanie 10 (Kacper Orszulak)

Algorytm bąbelkowy sortowania przegląda podaną na wejściu permutację i odwraca sąsiadujące elementy będące w inwersji. Przejścia wzdłuż permutacji są powtarzane tak długo, aż permutacja jest posortowana. Załóż, że podajemy temu algorytmowi losowo wybraną permutację zbioru $[n]$. Jaka jest oczekiwana liczba inwersji, którą odwróci algorytm podczas swojego działania. Jaka jest wariancja? Za pomocą nierówności Czebyszewa pokaż, że algorytm sortowania bąbelkowego wykonuje prawie zawsze liczbę kroków rzędu $\Omega(n^2)$.

Rozwiązanie

Wstęp

Zamiana miejscami dwóch sąsiednich elementów będących w inwersji zmniejsza ogólną liczbę inwersji dokładnie o jeden. Z tego powodu liczba zamian w algorytmie jest równa liczbie inwersji w wejściowej permutacji. W zadaniu będziemy zatem analizować własności zmiennej losowej X będącą liczbą inwersji wylosowanej permutacji.

Niech X_{ij} to indykator tego, czy dla $i < j$ zachodzi $\sigma(i) > \sigma(j)$. Oczywiście $X = \sum_{1 \leq i < j \leq n} X_{ij}$.

Zachodzi $\mathcal{P}(X_{ij} = 1) = \frac{n(n-1)/2}{n(n-1)} = \frac{1}{2}$.

Ponadto dla indykatorów zachodzi $\mathbb{E}(X_{ij}^2) = \mathbb{E}(X_{ij}) = \mathcal{P}(X_{ij} = 1)$. Zatem $\text{Var}(X_{ij}) = \frac{1}{4}$.

Wartość oczekiwana liczby zamian

$$\mathbb{E}(X) = \sum_{1 \leq i < j \leq n} \mathbb{E}(X_{ij}) = \sum_{1 \leq i < j \leq n} \mathcal{P}(X_{ij} = 1) = \binom{n}{2} \cdot \frac{1}{2} = \frac{n(n-1)}{4}$$

Wariancja liczby zamian

Lemat. *Wariancja sumy n zmiennych losowych.*

$$\text{Var}\left(\sum_{i \in [n]} X_i\right) = \sum_{i \in [n]} \text{Var}(X_i) + 2 \sum_{\{i,j\} \in \binom{[n]}{2}} \text{Cov}(X_i, X_j)$$

Dowód. *Pozostawiam jako ćwiczenie dla czytelnika ;).*

Wracając do zadania:

$$\text{Var}(X) = \text{Var}\left(\sum_{\{i,j\} \in \binom{[n]}{2}} X_{ij}\right) = \sum_{\{i,j\} \in \binom{[n]}{2}} \text{Var}(X_{ij}) + 2 \sum_{\{\{i,j\}, \{k,l\}\} \in \binom{\binom{[n]}{2}}{2}} \text{Cov}(X_{ij}, X_{kl})$$

Zastanówmy się teraz jak obliczyć sumę kowariancji.

Jeżeli rozważymy składnik taki, że i, j, k, l są parami różne, to dostaniemy zmienne X_{ij}, X_{kl} , które na pewno są niezależne, ponieważ nierówność $\sigma(i) \lesssim \sigma(j)$ nie wpływa na nierówność $\sigma(k) \lesssim \sigma(l)$. Oznacza to, że $Cov(X_{ij}, X_{kl}) = 0$.

Nie może się zdarzyć, że $i = j = k = l$, ponieważ $\{i, i\} = \{i\} \notin \binom{[n]}{2}$.

Analogicznie nie może zajść, że $i = j \vee k = l$. Nie może też zajść $i = k \wedge j = l$.

Pozostaje nam zatem rozważyć tylko przypadek, że $|\{i, j, k, l\}| = 3$.

Niech zatem $\{p, q, r\} \in \binom{[n]}{3}$ takie że $p < q < r$. BSO założmy, że $i < j \wedge k < l$ oraz $(i, j) \overset{\text{leksykograficznie}}{<} (k, l)$.

Oto składniki, których jeszcze nie rozważyliśmy:

$$Cov(X_{pq}, X_{pr}) + Cov(X_{pr}, X_{qr}) + Cov(X_{pq}, X_{qr})$$

Wiemy, że $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ ¹. Jedyne czego jeszcze nie wiemy zatem, to wartości oczekiwane iloczynów zmiennych losowych. Każdy iloczyn indyktorów, to też indyktor, a więc wystarczy policzyć tylko prawdopodobieństwa postaci $\mathcal{P}(X_{ij}X_{kl} = 1) = \mathcal{P}(X_{ij} = 1 \wedge X_{kl} = 1)$.

$\mathcal{P}(X_{pq} = 1 \wedge X_{pr} = 1) = P\left(\left\{\sigma_p > \sigma_q\right\}_{\sigma_p > \sigma_r}\right) = \frac{1}{3}$, ponieważ jest $\frac{2!}{3!}$ permutacji, w których σ_p jest największe. Zatem $Cov(X_{pq}, X_{qr}) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

$\mathcal{P}(X_{pr} = 1 \wedge X_{qr} = 1) = P\left(\left\{\sigma_p > \sigma_r\right\}_{\sigma_q > \sigma_r}\right) = \frac{1}{3}$, ponieważ jest $\frac{2!}{3!}$ permutacji, w których σ_r jest najmniejsze. Zatem $Cov(X_{pr}, X_{qr}) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

$\mathcal{P}(X_{pq} = 1 \wedge X_{qr} = 1) = P(\sigma_p > \sigma_q > \sigma_r) = \frac{1}{6} = \frac{1}{3!}$, ponieważ nie mamy żadnej swobody w permutacji trzejelementowej. Zatem $Cov(X_{pq}, X_{qr}) = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}$.

Dokończymy obliczenia:

$$Var(X) = \binom{n}{2} \cdot \frac{1}{4} + 2 \cdot \binom{n}{3} \cdot \left(\frac{1}{12} + \frac{1}{12} - \frac{1}{12}\right) = \frac{n(n-1)(2n+5)}{72}$$

Prawdopodobieństwo odchyłu

Twierdzenie. *Nierówność Czebyszewa* Dla dowolnego $a > 0$

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

Niech $\mu := \mathbb{E}[X]$, $\alpha \in (0; 1)$ oraz $a := (1 - \alpha)\mu$.

¹Na wykładzie i w P&C chyba nie ma tego wzoru, ale jest powszechnie znany.

Policzmy prawdopodobieństwo, że liczba inwersji jest co najwyżej $O(n^2)$.

$$\begin{aligned}\mathcal{P}(X = O(n^2)) &= \mathcal{P}(X \leq \alpha\mu) = \mathcal{P}(\mu - X \geq (1 - \alpha)\mu) = \\ &= \mathcal{P}(\mu - X \geq a) \leq \mathcal{P}(|X - \mu| \geq a) \\ &\leq \frac{\text{Var}[X]}{(1 - \alpha)^2\mu^2} = \frac{8}{9} \frac{2n + 5}{n(n - 1)} \frac{1}{(1 - \alpha)^2} \\ &\xrightarrow{n \rightarrow \infty} 0\end{aligned}$$

Zatem

$$\mathcal{P}(X = \Omega(n^2)) = 1 - \mathcal{P}(X = O(n^2)) \xrightarrow{n \rightarrow \infty} 1 - 0 = 1$$

Dostajemy więc szukaną tożsamość. $\square$

Odpowiedź

$$\mathbb{E}(X) = \frac{n(n-1)}{4}, \text{Var}[X] = \frac{n(n-1)(2n+5)}{72}$$

Zadanie 11 (Kuba Sułkowski)

Losujemy słowo binarne długości n . Rozważmy wszystkie $\binom{n}{2}$ par indeksów. Dla każdej takiej pary zmienna X_{ij} jest wartością funkcji xor na i -tym oraz j -tym bicie. Niech $X = \sum_{1 \leq i < j \leq n} X_{ij}$.

- (i) wykaż, że zmienne nie są niezależne jako rodzina zmiennych
- (ii) wykaż, że $\mathbb{E}(X_{ab}X_{cd}) = \mathbb{E}(X_{ab})\mathbb{E}(X_{cd})$ dla $\{a, b\} \neq \{c, d\}$
- (iii) wyznacz $\text{Var}(X)$ korzystając z poprzedniego podpunktu
- (iv) przy pomocy nierówności Czebyszewa ogranicz $P(|X - E(X)| \geq n)$

Rozwiązanie

(i)

Zakładamy $n \geq 3$ (dla mniejszego n jest ≤ 1 zmiennych losowych, więc na pewno są niezależne)

Oczywiście $P(X_{ij} = 0) = P(X_{ij} = 1) = \frac{1}{2} \quad \forall 1 \leq i < j \leq n$

Weźmy trzy bity A_1, A_2 i A_3 .

Zdarzenie $(X_{12} = 1 \wedge X_{13} = 1 \wedge X_{23} = 1)$ wymaga $A_1 \neq A_2, A_1 \neq A_3$ i $A_2 \neq A_3$, co nie jest możliwe jednocześnie.

Otrzymujemy $P(X_{12} = 1 \wedge X_{13} = 1 \wedge X_{23} = 1) = 0 \neq \frac{1}{8} = P(X_{12} = 1)P(X_{13} = 1)P(X_{23} = 1)$

(ii)

Jeśli $\{a, b\} \cap \{c, d\} = \emptyset$ to trywialne - X_{ab} i X_{cd} są niezależne.

Wpp. bez straty ogólności niech $a = c$, $b \neq d$. Wartości bitów A_a, A_b, A_d są niezależne i z równym prawdopodobieństwem równe 0 lub 1.

Oczywiście $E(X_{ab}) = P(X_{ab} = 1) = P(A_a \neq A_b) = \frac{1}{2} = E(X_{ad}) = E(X_{bd})$

$$E(X_{ab}X_{ad}) = P(X_{ab} = 1 \wedge X_{ad} = 1) = P(A_a \neq A_b \wedge A_a \neq A_d) \\ = P(A_b = A_d = 1 | A_a = 0)P(A_a = 0) + P(A_b = A_d = 0 | A_a = 1)P(A_a = 1)$$

Skoro A_a, A_b, A_d są niezależne, to można podstawić wszędzie $\frac{1}{2}$

$$E(X_{ab}X_{ad}) = \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^3 = \frac{1}{4} = E(X_{ab})E(X_{ad})$$

(iii)

Skoro każda para zmiennych jest niezależna, to wariancja sumy jest sumą wariancji.

$$\text{Var}\left(\sum_{1 \leq i < j \leq n} X_{ij}\right) = \sum_{1 \leq i < j \leq n} \text{Var}(X_{ij}) = \binom{n}{2} \cdot \frac{1}{4}$$

(iv)

$$P(|X - E(X)| \geq n) \leq \frac{\text{Var}(X)}{n^2} = \frac{\frac{1}{8}n(n-1)}{n^2} = \frac{n-1}{8n} < \frac{1}{8}$$

Odpowiedź (iv)

$$\boxed{\frac{n-1}{8n}}$$

Zadanie 12 (Kacper Poneta)

Niech $X_1, X_2, \dots, X_n$ będą niezależnymi zmiennymi losowymi z tego samego rozkładu o skończonej wartości oczekiwanej μ oraz skończonym odchyleniu standardowym σ . Na bazie nierówności Czebyszewa pokaż, że dla każdego $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| > \epsilon\right) = 0$$

Rozwiązanie

Zauważmy, że $\frac{X_1 + X_2 + \dots + X_n}{n}$ jest średnią arytmetyczną zmiennych losowych X_i . Zastosujmy nierówność Czebyszewa:

$$P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)}{\epsilon^2}$$

Obliczmy wariancję średniej:

$$\text{Var}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}$$

Podstawiając to do nierówności Czebyszewa, otrzymujemy:

$$P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Zauważmy, że gdy $n \rightarrow \infty$, to $\frac{\sigma^2}{n\epsilon^2} \rightarrow 0$. Zatem:

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| > \epsilon\right) = 0$$

Odpowiedź

Tak to prawda

Zadanie 13 (OSKIBOSKI123)

Wyznacz postać jawną funkcji tworzącej momenty dla rozkładu jednostajnego na zbiorze $[n]$.

Rozwiązanie

Zmienna X ma rozkład jednostajny dyskretny na zbiorze $\{1, 2, \dots, n\}$:

$$P(X = k) = \frac{1}{n} \quad \text{dla } k = 1, 2, \dots, n$$

Definicja funkcji tworzącej momentów (MGF)

$$M_X(t) = E(e^{tX}) = \sum_{k=1}^n e^{tk} \cdot P(X = k)$$

$$M_X(t) = \sum_{k=1}^n e^{tk} \cdot \frac{1}{n} = \frac{1}{n} \sum_{k=1}^n (e^t)^k$$

$$\sum_{k=1}^n (e^t)^k = e^t \cdot \frac{1 - (e^t)^n}{1 - e^t} = e^t \cdot \frac{1 - e^{nt}}{1 - e^t}$$

$$M_X(t) = \frac{1}{n} \cdot e^t \cdot \frac{1 - e^{nt}}{1 - e^t}$$

Dla $t = 0$ otrzymujemy nieoznaczoność, ale z definicji:

$$M_X(0) = E(e^{0 \cdot X}) = E(1) = 1$$

Odpowiedź

$$M_X(t) = \frac{e^t(1 - e^{nt})}{n(1 - e^t)} \text{ dla } t \neq 0, \text{ oraz } M_X(0) = 1.$$

Zadanie 14 (Stanisław Macura)

Wyznacz postać jawną funkcji tworzącej momenty dla rozkładu dwumianowego z parametrami n i p na dwa sposoby:

1. Oblicz $\mathbb{E}[e^{tX}]$ z definicji wartości oczekiwanej.
2. Skorzystaj z niezależności zmiennych Bernoulliego.

Jak na podstawie funkcji tworzącej dla rozkładu wyznaczyć wartość oczekiwaną i wariancję tego rozkładu?

Rozwiązanie

1. $\mathbb{E}[e^{tX}] = \sum_{j=0}^n \binom{n}{j} p^j (1-p)^{n-j} e^{tj} = \sum_{j=0}^n \binom{n}{j} (pe^t)^j (1-p)^{n-j} = (pe^t + (1-p))^n = (1-p + pe^t)^n$
2. Jako, iż X jest zmienną losową z rozkładem dwumianowym to $X = X_1 + \dots + X_n$, gdzie X_i jest indykatozem sukcesu w i -tej próbie, a zmienne X_i są niezależne. Wtedy $\mathbb{E}[e^{tX}] = \mathbb{E}[e^{t(X_1 + \dots + X_n)}] = \mathbb{E}[\prod_{i=1}^n e^{tX_i}] = \prod_{i=1}^n \mathbb{E}[e^{tX_i}] = \prod_{i=1}^n (pe^t + (1-p)) = (1-p + pe^t)^n$

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

$$\mathbb{E}(X) = M_X^{(1)}(0) = np$$

$$M_X^{(2)}(t) = npe^t(1-p + npe^t)(1-p + pe^t)^{n-2}$$

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