

Zadanie 1 (Wiktor K)

Wyznacz wartość oczekiwaną i wariancję rozkładu Poissona na podstawie funkcji tworzącej wyliczonej na wykładzie.

Rozwiązanie

Niech X - zmienna losowa o rozkładzie Poissona z parametrem μ

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

korzystając ze wzoru na pochodną złożenia funkcji

$$M_X^{(1)}(t) = \frac{1}{e^\mu} \cdot e^{\mu e^t} \cdot \mu e^t = \frac{\mu}{e^\mu} e^{\mu e^t + t}$$

$$M_X^{(2)}(t) = \frac{\mu}{e^\mu} \cdot e^{\mu e^t + t} (1 + \mu e^t)$$

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

$$M_X^{(2)}(0) = \frac{\mu}{e^\mu} e^\mu (1 + \mu) = \mu^2 + \mu$$

$$Var(X) = E(X^2) - E(X)^2 = \mu^2 + \mu - \mu^2 = \mu$$

Odpowiedź

$E(X) = Var(X) = \mu$

Zadanie 2 (Wiktor M)

Niech X, Y, Z będą niezależnymi zmiennymi o rozkładzie Poissona. Niech $E(X) = 1, E(Y) = 2, E(Z) = 3$. Oblicz $P(X + Y + Z \leq 2)$.

Rozwiązanie

Skoro X, Y, Z to niezależne zmienne o rozkładzie Poissona, to ich suma również jest zmienną Poissona z wartością oczekiwaną 6 (na podstawie wykładu, jest to również w książce jako Lemat 5.2). A zatem:

$$P(X + Y + Z \leq 2) = P(X + Y + Z = 0) + P(X + Y + Z = 1) + P(X + Y + Z = 2)$$

Wynikiem będzie więc:

$$\frac{e^{-6} \cdot 6^0}{0!} + \frac{e^{-6} \cdot 6^1}{1!} + \frac{e^{-6} \cdot 6^2}{2!} = 25e^{-6}$$

Odpowiedź

$25e^{-6}$

Zadanie 3 (Autor/ka)

Niech X będzie zmienną Poissona z wartością oczekiwaną $\mu \in \{1, 2, \dots\}$.

(i) Pokaż, że

$$\Pr(X = \mu + h) \geq \Pr(X = \mu - h - 1) \quad \text{dla } 0 \leq h \leq \mu - 1.$$

(ii) Na tej podstawie wykaż, że

$$\Pr(X \geq \mu) \geq \frac{1}{2}.$$

Rozwiązanie

Dla $X \sim \text{Poisson}(\mu)$ mamy

$$\Pr(X = k) = e^{-\mu} \frac{\mu^k}{k!}.$$

(i)

Dla $0 \leq h \leq \mu - 1$ rozważmy stosunek

$$R := \frac{\Pr(X = \mu + h)}{\Pr(X = \mu - h - 1)} = \frac{e^{-\mu} \mu^{\mu+h} / (\mu + h)!}{e^{-\mu} \mu^{\mu-h-1} / (\mu - h - 1)!} = \frac{\mu^{2h+1}}{(\mu - h)(\mu - h + 1) \cdots (\mu + h)}.$$

Można to zapisać jako

$$R = \prod_{s=-h}^h \frac{\mu}{\mu + s}.$$

Parując czynniki dla s i $-s$ (dla $s = 1, \dots, h$) oraz pozostawiając czynnik $s = 0$, otrzymujemy

$$\frac{\mu}{\mu - s} \cdot \frac{\mu}{\mu + s} = \frac{\mu^2}{\mu^2 - s^2} \geq 1,$$

bo $\mu^2 - s^2 \leq \mu^2$. Każda taka para jest więc nie mniejsza niż 1, a czynnik dla $s = 0$ to 1, zatem $R \geq 1$.

Stąd

$$\Pr(X = \mu + h) \geq \Pr(X = \mu - h - 1)$$

dla każdego $0 \leq h \leq \mu - 1$, co kończy część (i).

(ii)

Z części (i) dla każdego $h = 0, 1, \dots, \mu - 1$ mamy $\Pr(X = \mu + h) \geq \Pr(X = \mu - h - 1)$. Sumując po $h = 0, \dots, \mu - 1$ otrzymujemy

$$\sum_{h=0}^{\mu-1} \Pr(X = \mu + h) \geq \sum_{h=0}^{\mu-1} \Pr(X = \mu - h - 1).$$

Lewa suma to część masy rozkładu od μ wzwyż, prawa zaś – od 0 do $\mu - 1$. Zatem

$$\Pr(X \geq \mu) \geq \Pr(X \leq \mu - 1).$$

Ponieważ

$$1 = \Pr(X \geq \mu) + \Pr(X \leq \mu - 1),$$

otrzymujemy

$$\Pr(X \geq \mu) \geq \frac{1}{2}.$$

Odpowiedź

tutaj daj odpowiedź

Zadanie 4 (Pavlo Tsikalysyn)

Niech zmienna losowa $X_i^{(m)}$ oznacza liczbę kul w i -tym kubku, przy założeniu, że rzucamy m kul do n kubków w sposób jednostajny i niezależny. Niech zmienne $Y_1^{(m)}, \dots, Y_n^{(m)}$ będą niezależnymi zmiennymi z rozkładu Poissona o średniej $\frac{m}{n}$. Podczas wykładu widzieliśmy, że dla każdej nieujemnej funkcji f zachodzi

$$E\left(f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right)\right) \geq E\left(f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right) P\left(\sum_{i=1}^n Y_i^{(m)} = m\right).$$

- (i) Pokaż, że jeśli $E(f(X_1^{(m)}, \dots, X_n^{(m)}))$ jest rosnąca względem m , to przy założeniu nieujemności f zachodzi:

$$E\left(f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right)\right) \geq E\left(f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right) P\left(\sum_{i=1}^n Y_i^{(m)} \geq m\right).$$

- (ii) Wykaż, że jeśli $E(f(X_1^{(m)}, \dots, X_n^{(m)}))$ jest malejąca względem m , to

$$E\left(f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right) \leq 2 \cdot E\left(f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right)\right).$$

Rozwiązanie

(i)

$$\begin{aligned} E\left(f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right)\right) &= \sum_{k=0}^{\infty} E\left(f\left(Y_1^{(m)}, \dots, Y_n^{(m)}\right) \mid \sum_{i=1}^n Y_i = k\right) P\left(\sum_{i=1}^n Y_i^{(m)} = k\right) = \\ &= \sum_{k=0}^{\infty} E\left(f\left(X_1^{(k)}, \dots, X_n^{(k)}\right)\right) P\left(\sum_{i=1}^n Y_i^{(m)} = k\right) \geq \sum_{k=m}^{\infty} E\left(f\left(X_1^{(k)}, \dots, X_n^{(k)}\right)\right) P\left(\sum_{i=1}^n Y_i^{(m)} = k\right) \end{aligned}$$

Z tego że funkcja jest rosnąca mamy że:

$$\begin{aligned} \sum_{k=m}^{\infty} E\left(f\left(X_1^{(k)}, \dots, X_n^{(k)}\right)\right) P\left(\sum_{i=1}^n Y_i^{(m)} = k\right) &\geq \sum_{k=m}^{\infty} E\left(f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right) P\left(\sum_{i=1}^n Y_i^{(m)} = k\right) = \\ &= E\left(f\left(X_1^{(m)}, \dots, X_n^{(m)}\right)\right) P\left(\sum_{i=1}^n Y_i^{(m)} \geq m\right) \end{aligned}$$

(ii) Tak samo jak w poprzednim możemy udowodnić że

$$\begin{aligned} E(f(Y_1^{(m)}, \dots, Y_n^{(m)})) &\geq E(f(X_1^{(m)}, \dots, X_n^{(m)})) P\left(\sum_{i=1}^n Y_i^{(m)} \leq m\right) \\ E(f(X_1^{(m)}, \dots, X_n^{(m)})) &\leq \frac{E(f(Y_1^{(m)}, \dots, Y_n^{(m)}))}{P\left(\sum_{i=1}^n Y_i^{(m)} \leq m\right)} \\ P\left(\sum_{i=1}^n Y_i^{(m)} \leq m\right) &\geq \frac{1}{2} \\ E(f(X_1^{(m)}, \dots, X_n^{(m)})) &\leq 2E(f(Y_1^{(m)}, \dots, Y_n^{(m)})) \end{aligned}$$

Zadanie 5 (Filip Maniżak)

Niech X będzie zmienną losową z rozkładu Poissona o wartości oczekiwanej $E(X) = \mu$. Zmienna X reprezentuje liczbę błędów w pewnej książce. Każdy błąd jest niezależnie od pozostałych błędem gramatycznym z prawdopodobieństwem p albo leksykalnym z prawdopodobieństwem $1 - p$. Niech Y określa liczbę błędów gramatycznych, zaś Z liczbę błędów leksykalnych ($X = Y + Z$). Pokaż, że Y jest zmienną losową z rozkładu Poissona o wartości oczekiwanej $p\mu$, zaś Z jest zmienną losową z rozkładu Poissona o wartości oczekiwanej $(1 - p)\mu$. Dodatkowo pokaż, że zmienne Y oraz Z są niezależne.

Rozwiązanie

Okej, no to policzmy te rozkłady (policzę $P(Y = k)$, drugi jest anal).

$$P(Y = k) = \sum_{i=k}^{\infty} \frac{\mu^i e^{-\mu}}{i!} \binom{i}{k} p^k (1-p)^{i-k}$$

Czyli, liczymy prawdopodobieństwo że $X = i$ a potem, liczymy jakie jest prawdopodobieństwo że dokładnie k z tych i błędów są naszymi błędami.

$$\begin{aligned} P(Y = k) &= \sum_{i=k}^{\infty} \frac{\mu^i e^{-\mu}}{i!} \binom{i}{k} p^k (1-p)^{i-k} = \frac{e^{-\mu} p^k \mu^k}{k!} \sum_{i=k}^{\infty} \frac{\mu^{i-k} (1-p)^{i-k}}{(i-k)!} = \\ &= \frac{e^{-\mu} p^k \mu^k}{k!} \sum_{n=0}^{\infty} \frac{(\mu(1-p))^n}{(n)!} = \frac{e^{-\mu p} (\mu p)^k}{k!} \end{aligned}$$

Ostatnie przejście, to po prostu zastosowanie wzoru na e. Czyli istotnie Y i analogicznie Z to zmienne z rozkładu Poissona. Zostało jeszcze wykazać że są niezależne. Aby $Y = k$ i $Z = m$, musi być $X = k + m$, zatem:

$$\begin{aligned} P(Y = k, Z = m) &= P(X = k + m) \binom{k+m}{k} p^k (1-p)^m \\ &= \frac{e^{-\mu} \mu^{k+m}}{(k+m)!} \binom{k+m}{k} p^k (1-p)^m. \end{aligned}$$

Ponieważ $\binom{k+m}{k} = \frac{(k+m)!}{k!m!}$, mamy:

$$P(Y = k, Z = m) = e^{-\mu} \frac{\mu^{k+m}}{k!m!} p^k (1-p)^m.$$

Teraz rozbijmy to na iloczyn dwóch składników:

$$\begin{aligned} P(Y = k, Z = m) &= \left(e^{-\mu p} \frac{(\mu p)^k}{k!} \right) \left(e^{-\mu(1-p)} \frac{(\mu(1-p))^m}{m!} \right) \\ &= P(Y = k) P(Z = m). \end{aligned}$$

Czyli zmienne są niezależne.

Zadanie 6 (Kuba Sułkowski)

Rzucamy m kul w sposób jednostajny i niezależny do n urn, ponumerowanych $0, \dots, n-1$. Mówimy, że w indeksie i jest k -luka, jeśli wszystkie urny o indeksach $i, i+1, \dots, i+k-1$ są puste.

- (i) Wyznacz wartość oczekiwaną liczby k -luk.
- (ii) Wyznacz nierówność Czernowa dla liczby k -luk.

Rozwiązanie (i)

Niech indykator X_i oznacza, czy w miejscu i mamy k -lukę, czyli $X_i = 1$ iff wszystkie urny $i, \dots, i+k-1$ są puste.

$E(X_i) = P(X_i = 1) = \left(\frac{n-k}{n}\right)^m$ - każda z m kul musiała trafić w inne miejsce

Mamy $(n-k+1)$ możliwych indeksów w których będzie k -luka - $0, 1, \dots, n-k$

Z liniowości wartości oczekiwanej

$$E\left(\sum_{i=0}^{n-k} X_i\right) = \sum_{i=0}^{n-k} E(X_i) = (n-k+1)E(X_0) = (n-k+1) \left(\frac{n-k}{n}\right)^m$$

Rozwiązanie (ii)

Przyjmijmy kilka oznaczeń.

$X = \sum_{i=0}^{n-k} X_i$ - liczba k -luk

$\mu = E(X) = (n-k+1) \left(\frac{n-k}{n}\right)^m$

$B = (x_0, x_1, \dots, x_{n-1})$ - losowa tupla z modelu n urn i m kul

$p = (y_0, y_1, \dots, y_{n-1})$ - niezależne zmienne o rozkładzie Poissona z współczynnikiem $\frac{m}{n}$

Niech $Z_i^{(a_0, \dots, a_{n-1})}$ będzie liczbą k -luk na indeksach $\equiv i \pmod{k}$.

Z_i^B oraz Z_i^P będą zmiennymi losowymi, oznaczającymi liczbę k -luk dla tupli z modelu z kulami i z modelu Poissona.

Nasze ograniczenie nie będzie szczególnie dokładne, więc możemy zignorować podłogi i sufity i działać bezpośrednio na ułamkach.

Zgodnie z naszymi oznaczeniami:

$$Z_i^B = Z_i^{(x_0, \dots, x_{n-1})} = \sum_{j=0}^{\frac{n}{k}} x_{i+jk}$$

$$Z_i^P = Z_i^{(y_0, \dots, y_{n-1})} = \sum_{j=0}^{\frac{n}{k}} y_{i+jk}$$

$$E(Z_i^B) = \frac{\mu}{k}$$

Ustalmy funkcję g - czy na konkretnym modulo ($0 \bmod k$) mamy więcej niż oczekiwane k -luk

$$g((a_0, \dots, a_{n-1})) = \begin{cases} 1 & Z_0^{((a_0, \dots, a_{n-1}))} \geq (1 + \delta) \frac{\mu}{k} \\ 0 & \text{wpp} \end{cases}$$

Zauważmy, że g jest monotoniczne - zmniejszając liczbę kul możemy jedynie zwiększyć liczbę k -luk. Z twierdzenia z wykładu dla funkcji monotonicznych (**Theorem 5.10** w P&C) mamy:

$$P\left(Z_0^B \geq (1 + \delta) \frac{\mu}{k}\right) = E(g(x_0, \dots, x_{n-1})) \leq E(g(y_0, \dots, y_{n-1})) \cdot 2$$

Z_i^P jest tak skonstruowane, że te k -luki są od siebie niezależne - rozważamy k -luki na pozycjach modulo k , więc są rozłączne. Można traktować Z_i^P jako $\frac{n}{k}$ niezależnych prób Bernoulliego, a wtedy skorzystać z lematu sprzed kilku zajęć (**Theorem 4.4** w P&C)

$$E(g(y_0, \dots, y_{n-1})) = P\left(Z_0^P \geq (1 + \delta) \frac{\mu}{k}\right) \leq e^{\frac{-\mu}{k} \delta^2} = \exp\left(\frac{-\mu}{k} \delta^2\right)$$

(od tego miejsca dla czytelności będę pisał $\exp(t)$ zamiast e^t)

Aby ograniczyć X (liczbę wszystkich k -luk) użyjemy upper bounda

$$P(X \geq (1 + \delta)E(X)) \leq P\left(\bigcup_{i=0}^{k-1} \left\{Z_i^B \geq (1 + \delta) \frac{\mu}{k}\right\}\right) \leq^{\text{UB}} kP\left(Z_0^B \geq (1 + \delta) \frac{\mu}{k}\right)$$

$$\leq 2k \cdot P\left(Z_0^P \geq (1 + \delta) \frac{\mu}{k}\right) \leq 2k \cdot \exp\left(\frac{-\mu}{k} \delta^2\right)$$

Licząc lewe ograniczenie postępujemy analogicznie. Warto zwrócić uwagę, że w **Theorem 4.5** mamy inny mianownik w wykładniku

$$\begin{aligned} P(X \leq (1 - \delta)E(X)) &\leq P\left(\bigcup_{i=0}^{k-1} \left\{Z_i^B \leq (1 - \delta)\frac{\mu}{k}\right\}\right) \leq^{\text{UB}} kP\left(Z_0^B \leq (1 - \delta)\frac{\mu}{k}\right) \\ &\leq 2k \cdot P\left(Z_0^B \leq (1 - \delta)\frac{\mu}{k}\right) \leq 2k \cdot \exp\left(\frac{-\frac{\mu}{k}\delta^2}{2}\right) \end{aligned}$$

Notatka od autora: Iterowanie się po ułamkach w tym zadaniu jest całkowicie nieformalne. Taka jednak metoda została zaprezentowana w czasie zajęć i taka będzie oczekiwana, gdyby podobne zadanie ewentualnie pojawiło się w czasie kolokwium.

Odpowiedź

$$\begin{aligned} \text{(i)} \quad & \boxed{(n - k + 1) \left(\frac{n - k}{n}\right)^m} \\ \text{(ii)} \quad & \boxed{P(X \geq (1 + \delta)E(X)) \leq 2k \cdot \exp\left(\frac{-\frac{\mu}{k}\delta^2}{3}\right)} \\ \text{(ii)} \quad & \boxed{P(X \leq (1 - \delta)E(X)) \leq 2k \cdot \exp\left(\frac{-\frac{\mu}{k}\delta^2}{2}\right)} \end{aligned}$$

Zadanie 7 (Olek Wieczorek)

Otrzymujemy w prezencie X kul, gdzie X ma rozkład Poissona z parametrem 2022. Dla każdej kuli rzucamy sześciocienną kostką. Jeśli wyrzuciliśmy 1, to malujemy kulę na białą. Jeśli wyrzuciliśmy 2, to malujemy kulę na czarno. Jeśli zaś wyrzuciliśmy coś innego, to wyrzucamy kulę do śmieci.

- Z jakim prawdopodobieństwem będziemy mieć więcej czarnych kul niż białych?
- Jaki jest rozkład liczby czarnych kul przy założeniu, że nie uzyskaliśmy żadnej białej kuli?
- Jaka jest oczekiwana liczba białych kul przy założeniu, że czarnych kul jest dwa razy więcej od białych?

Rozwiązanie

Problem jest bardzo podobny do tego z zadania 5. Będę używał wniosków tego zadania jako blackboxa i oznaczał to gwiazdką (*)

Rozbijmy problem na prostszy, aby sprawnie policzyć wszystkie podpunkty. Rozbijmy najpierw zmienną X .

X' – kule które zostawiamy

W – kule które wyrzucamy

Zostawiamy kulę, jeśli wyrzucimy 1 lub 2 (z $p = \frac{1}{3}$).

Z (*) X' ma rozkład Poissona z parametrem $\frac{2022}{3} = 674$. Wynik W nie będzie nas dalej interesował.

Analogicznie rozbijamy X' :

C – kule które pomalowano na czarno

B – kule które pomalowano na białe

Zauważmy, że zostawiona kula z prawdopodobieństwem $\frac{1}{2}$ będzie biała, albo będzie czarna z takim samym prawdopodobieństwem.

Jeszcze raz z (*) B oraz C są zmiennymi Poissona z parametrem $\frac{674}{2} = 337$, co więcej te zmienne są niezależne (również z (*)). Mając to wszystko możemy przejść do podpunktów.

(i) Z jakim prawdopodobieństwem będziemy mieć więcej czarnych kul niż białych?

Wystarczy policzyć $P(B > C)$. Zauważmy, że jako, że te zmienne mają ten sam rozkład to mamy symetrycznie $P(B > C) = P(C > B)$. Stąd dostaniemy $P(B > C) = \frac{1 - P(B=C)}{2}$

$$P(B = C) = \sum_{n=0}^{\infty} P(B = n \cap C = n) = \sum_{n=0}^{\infty} P(B = n) \cdot P(C = n) =$$

$$\sum_{n=0}^{\infty} \left(\frac{e^{-337} (337)^n}{n!} \right)^2$$

Stąd mamy odpowiedź:

$$P(B > C) = \frac{1 - \sum_{n=0}^{\infty} \left(\frac{e^{-337} (337)^n}{n!} \right)^2}{2}$$

(ii) Jaki jest rozkład liczby czarnych kul przy założeniu, że nie uzyskaliśmy żadnej białej kuli?

Liczmy $P(C|B=0)$ z definicji

$$P(C = k|B = 0) = \frac{P(C = k \cap B = 0)}{P(B = 0)} = \frac{P(C = k) \cdot P(B = 0)}{P(B = 0)} = P(C = k)$$

Rozkład jest taki sam jak dla samej zmiennej C (warunek się skraca z niezależności), czyli:

$$P(C = k) = \frac{e^{-337} (337)^k}{k!}$$

(iii) Jaka jest oczekiwana liczba białych kul przy założeniu, że czarnych kul jest dwa razy więcej od białych? Liczymy wartość oczekiwaną z definicji:

$$E(B|C = 2B) = \sum_{k=0}^{\infty} k P(B = k|C = 2B) = \sum_{k=0}^{\infty} \frac{k P(B = k \cap C = 2k)}{P(C = 2B)} =$$

$$\frac{\sum_{k=0}^{\infty} k P(B = k) \cdot P(C = 2k)}{\sum_{j=0}^{\infty} P(B = j) \cdot P(C = 2j)}$$

To już nam się tak ładnie nie skróci. Możemy jedynie podstawić wzór na rozkład Poissona:

$$E(B|C = 2B) = \sum_k \frac{k \frac{e^{-337} (337)^k}{k!} \frac{e^{-337} (337)^{2k}}{(2k)!}}{\sum_{j=0}^{\infty} \frac{e^{-337} (337)^j}{j!} \frac{e^{-337} (337)^{2j}}{(2j)!}} =$$

$$\frac{\sum_{k=0}^{\infty} k \frac{(337)^{3k}}{k! (2k)!}}{\sum_{j=0}^{\infty} \frac{(337)^{3j}}{j! (2j)!}}$$

Odpowiedź

$$\begin{aligned} \text{(i)} \quad & \frac{1 - P(B = C)}{2} \\ \text{(ii)} \quad & P(C = k) \\ \text{(iii)} \quad & \frac{\sum_{k=0}^{\infty} k P(B = k) \cdot P(C = 2k)}{\sum_{j=0}^{\infty} P(B = j) \cdot P(C = 2j)} \end{aligned}$$

Zadanie 8 (Dominik)

Dostaliśmy od Świętego Mikołaja B białych kul, gdzie B jest zmienną o rozkładzie Poissona z parametrem $\mu = 2023$.

Każda kula z prawdopodobieństwem p ($0 < p < 1$) jest pomalowana na czerwono niezależnie od innych kul. Niech C będzie zmienną losową oznaczającą liczbę czerwonych kul w tym momencie.

Każda czerwona kula z prawdopodobieństwem p jest pomalowana na niebiesko niezależnie od innych kul. Niech D będzie zmienną losową oznaczającą liczbę niebieskich kul w tym momencie.

Każda niebieska kula z prawdopodobieństwem p jest pomalowana na żółto niezależnie od innych kul. Niech E będzie zmienną losową oznaczającą liczbę żółtych kul w tym momencie.

1. Dla $p = \sqrt[3]{1 - \frac{\ln 3}{2023}}$ sprawdź, czy wartość $P(B > E)$ jest większa, mniejsza, czy równa $\frac{2}{3}$.
2. Dla jakiego p wartość oczekiwana E jest równa 1?
3. Znajdź $\mathbb{E}[E \mid B]$ i $\mathbb{E}[B \mid E]$.

Rozwiązanie podpunktu (1)

$P(B > E)$ to prawdopodobieństwo zdarzenia, że **nie wszystkie** kule zostały pomalowane na żółto (a więc wcześniej też na czerwono i na niebiesko).

Policzmy prawdopodobieństwo zdarzenia przeciwnego, czyli $P(B = E)$.

$$\begin{aligned} P(B = E) &= \sum_{i=0}^{\infty} P(B = E \cap B = i) \\ &= \sum_{i=0}^{\infty} P(B = E \mid B = i) \cdot P(B = i) \end{aligned}$$

W sumowanym wyrażeniu $P(B = E \mid B = i)$ to prawdopodobieństwo zdarzenia, że wszystkie z i wylosowanych kul zostały pokolorowane na czerwono, na niebiesko i na żółto, więc:

$$P(B = E \mid B = i) = p^{3i}.$$

B ma rozkład Poissona, więc $P(B = i) = \frac{e^{-\mu} \cdot \mu^i}{i!}$.

Podstawmy te wartości w powyższej równości i przekształćmy ją tak, aby po prawej stronie znaleźć prawdopodobieństwo wartości w rozkładzie Poissona z parametrem $p^3 \cdot \mu$:

$$\begin{aligned} P(B = \textcolor{brown}{E}) &= \sum_{i=0}^{\infty} p^{3i} \cdot \frac{e^{-\mu} \cdot \mu^i}{i!} \\ &= \sum_{i=0}^{\infty} \frac{e^{-\mu} \cdot (p^3 \cdot \mu)^i}{i!} \\ &= \sum_{i=0}^{\infty} \frac{e^{-\mu} \cdot e^{(p^3 \cdot \mu)} \cdot e^{-(p^3 \cdot \mu)} \cdot (p^3 \cdot \mu)^i}{i!} \\ &= e^{(p^3 - 1)\mu} \cdot \sum_{i=0}^{\infty} \frac{e^{-(p^3 \cdot \mu)} \cdot (p^3 \cdot \mu)^i}{i!} \end{aligned}$$

Powiedzmy, że X to zmienna losowa o rozkładzie Poissona z parametrem $p^3 \cdot \mu$. Wtedy:

$$\sum_{i=0}^{\infty} \frac{e^{-(p^3 \cdot \mu)} \cdot (p^3 \cdot \mu)^i}{i!} = \sum_{i=0}^{\infty} P(X = i) = 1.$$

Mamy więc, że $P(B = \textcolor{brown}{E}) = e^{(p^3 - 1)\mu}$. Podstawiając wartości z zadania, czyli $p = \sqrt[3]{1 - \frac{\ln 3}{2023}}$ i $\mu = 2023$, mamy:

$$P(B = \textcolor{brown}{E}) = e^{\left(\left(\sqrt[3]{1 - \frac{\ln 3}{2023}}\right)^3 - 1\right) \cdot 2023} = e^{-\frac{\ln 3}{2023} \cdot 2023} = e^{-\ln 3} = \frac{1}{3},$$

zatem $P(B > \textcolor{brown}{E}) = 1 - P(B = \textcolor{brown}{E}) = \frac{2}{3}$.

Rozwiązanie podpunktu (2)

Intuicyjnym jest, że oczekiwana liczba kul pokolorowanych na żółto będzie iloczynem oczekiwanej liczby wszystkich kul i prawdopodobieństwa tego, że dana kula zostanie pokolorowana na żółto, czyli: $\mathbb{E}[\textcolor{brown}{E}] = p^3 \cdot \mathbb{E}[B]$. Poniższy fragment jedynie formalizuje tę intuicję:

Niech $\textcolor{brown}{E}_j$ to indykator zdarzenia, że j -ta kula została pomalowana na żółto:

$$\mathbb{E}[\textcolor{brown}{E}_j] = P(\textcolor{brown}{E}_j = 1) = p^3.$$

W zależności od wartości zmiennej losowej B :

$$\textcolor{brown}{E} = \sum_{j=1}^B \textcolor{brown}{E}_j.$$

Wartość oczekiwana E to $\mathbb{E}[\textcolor{brown}{E}] = \sum_{i=0}^{\infty} \left(\mathbb{E}[\textcolor{brown}{E} \mid B = i] \cdot P(B = i) \right)$.

Obliczmy $\mathbb{E}[\textcolor{brown}{E} \mid B = i]$:

$$\mathbb{E}[\textcolor{brown}{E} \mid B = i] = \mathbb{E} \left[\sum_{j=1}^B \textcolor{brown}{E}_j \mid B = i \right] = \sum_{j=1}^i \mathbb{E}[\textcolor{brown}{E}_j] = i \cdot p^3$$

Podstawiając tę wartość do $\mathbb{E}[E]$:

$$\begin{aligned}\mathbb{E}[E] &= \sum_{i=0}^{\infty} \left(\mathbb{E}[E \mid B = i] \cdot P(B = i) \right) \\ &= \sum_{i=0}^{\infty} \left(i \cdot p^3 \cdot P(B = i) \right) \\ &= p^3 \cdot \mathbb{E}[B]\end{aligned}$$

B jest zmienną losową z rozkładem Poissona, więc jej wartość oczekiwana to $\mathbb{E}[B] = \mu = 2023$.

Aby $\mathbb{E}[E] = 1$: $p^3 = \frac{\mathbb{E}[E]}{\mathbb{E}[B]} = \frac{1}{2023}$, czyli $p = \frac{1}{\sqrt[3]{2023}}$.

Rozwiązanie podpunktu (3)

Część pierwsza: $\mathbb{E}[E \mid B]$

$\mathbb{E}[E \mid B]$ jest funkcją $\mathbb{E}[E \mid B]: \mathbb{R} \rightarrow \mathbb{R}$ zdefiniowaną jako:

$$\mathbb{E}[E \mid B](i) := \mathbb{E}[E \mid B = i].$$

Wartość $\mathbb{E}[E \mid B = i]$ policzyliśmy już w podpunkcie (2).

Dla $i \in \mathbb{N}$: $\mathbb{E}[E \mid B](i) = i \cdot p^3$, dla pozostałych i $\mathbb{E}[E \mid B](i) = 0$, bo $P(B = i) = 0$.

Część druga: $\mathbb{E}[B \mid E]$

$\mathbb{E}[B \mid E = i]$ to oczekiwana liczba wszystkich kul zakładając, że i spośród tych kul zostało pokolorowanych na żółto.

Mogłoby się wydawać, że $\mathbb{E}[B \mid E = i] = i \cdot p^{-3}$, ale tak nie jest, bo $\mathbb{E}[B \mid E = 0] > 0$.

Przeprowadźmy inne doświadczenie losowe. Losujemy niezależnie:

- Pewną liczbę żółtych kul, gdzie E ma rozkład Poissona z parametrem $\mu_E = \mu \cdot p^3$.
- Pewną liczbę F kul w innych kolorach, gdzie F ma rozkład Poissona z parametrem $\mu_F = \mu \cdot (1 - p^3)$.

Zmienna losowa B to liczba wszystkich kul, więc $B = E + F$.

Lematu z wykładu: Suma zmiennych o rozkładzie Poissona z parametrami μ_1 i μ_2 jest zmienną o rozkładzie Poissona z parametrem $\mu_1 + \mu_2$.

Blef alert: „Utożsamianie zmiennych” poniżej wymaga lepszego dowodu. Używamy tutaj związku zmiennych o rozkładzie Poissona ze zmiennymi o rozkładzie dwumianowym, ale ja ten związek niezbyt rozumiem.

Na mocy tego lematu B jest zmienną o rozkładzie Poissona z parametrem $\mu_E + \mu_F = \mu$, co zgadza się z B z treści zadania — utożsammy te zmienne ze sobą.

Dla ustalonego B zmienna E ma rozkład dwumianowy z parametrem p^3 , co również jest zgodne z rozkładem E z treści zadania, więc je także możemy utożsamiać.

$$\begin{aligned}\mathbb{E}[B \mid E = i] &= \mathbb{E}[(E + F) \mid E = i] \\ &= \mathbb{E}[E \mid E = i] + \mathbb{E}[F \mid E = i] \\ &= i + \mathbb{E}[F \mid E = i]\end{aligned}$$

Zmienne E i F są niezależne, więc $\mathbb{E}[F \mid E = i] = \mathbb{E}[F] = \mu_F = \mu \cdot (1 - p^3)$.

$$\mathbb{E}[B \mid E = i] = i + \mu(1 - p^3) = i + 2023 \cdot (1 - p^3).$$

Ponownie z definicji $\mathbb{E}[B \mid E](i) = \mathbb{E}[B \mid E = i]$ dostajemy odpowiedź.

Odpowiedź

Podpunkt (1): $P(B > E) = \frac{2}{3}$.

Podpunkt (2): $p = \frac{1}{\sqrt[3]{2023}}$

Podpunkt (3): $\mathbb{E}[E \mid B](B) = \begin{cases} B \cdot p^3 & \text{gdy } B \in \mathbb{N}, \\ 0 & \text{wpp.} \end{cases}$

oraz $\mathbb{E}[B \mid E](E) = \begin{cases} E + 2023 \cdot (1 - p^3) & \text{gdy } E \in \mathbb{N}, \\ 0 & \text{wpp.} \end{cases}$

Zadanie 9

Przeczytaj Sekcję 5.5.1 Chain Hashing i 5.5.2 Hashing: Bit Strings (przykład zastosowania modelu kul i urn).

Zadanie 10 (Kacper Orszulak)

Rozwiąż Zadanie 5.22 z książki.

Uwaga: w podpunkcie (b) powinno być $\lfloor \log n \rfloor$ zamiast $\log n$.

Excercise 5.22 (P&C 2nd Edition): In hashing with open addressing, the hash table is implemented as an array and there are no linked lists or chaining. Each entry in the array either contains one hashed item or is empty. The hash function defines, for each key k , a probe sequence $h(k, 0), h(k, 1), \dots$ of table locations. To insert the key k , we first examine the sequence of table locations in the order defined by the key's probe sequence until we find an empty location; then we insert the item at that position. When searching for an item in the hash table, we examine the sequence of table locations in the order defined by the key's probe sequence until either the item is found or we have found an empty location in the sequence. If an empty location is found, this means the item is not present in the table.

An open-address hash table with $2n$ entries is used to store n items. Assume that the table location $h(k, j)$ is uniform over the $2n$ possible table locations and that all $h(k, j)$ are independent.

- (a) Show that, under these conditions, the probability of an insertion requiring more than k probes is at most 2^{-k} .
- (b) Show that, for $i = 1, 2, \dots, n$, the probability that the i th insertion requires more than $2 \lfloor \log n \rfloor$ probes is at most $1/n^2$.

Let the random variable X_i denote the number of probes required by the i th insertion. You have shown in part (b) that $\Pr(X_i > 2 \lfloor \log n \rfloor) \leq 1/n^2$. Let the random variable $X = \max_{1 \leq i \leq n} X_i$ denote the maximum number of probes required by any of the n insertions.

- (c) Show that $\Pr(X > 2 \lfloor \log n \rfloor) \leq 1/n$.
- (d) Show that the expected length of the longest probe sequence is $\mathbb{E}[X] = O(\log n)$.

Rozwiązanie

(a)

Założmy pesymistycznie, że n komórek tablicy jest już zajęte. Prawdopodobieństwo, że dowolna komórka jest zajęta jest równe zatem $(1 - p) = \frac{n}{n/2} = \frac{1}{2}$ (w ogólności oczywiście $(1 - p) \leq \frac{1}{2}$). Jest tak dlatego, ponieważ każda komórka występuje w ciągu $h(k, *)$ z równym prawdopodobieństwem, a co za tym idzie, jeżeli wstawiamy i -ty element do tablicy, to $i - 1$ elementów jest zajętych, a skoro wybieramy pierwszą niezajętą komórkę w ciągu $h(k, *)$, to z równym prawdopodobieństwem będzie to dowolna niezajęta komórka.

Z tego powodu X_i (definicja wyżej) ma rozkład geometryczny. Dostajemy

$$\mathcal{P}(X_i = k) = (1 - p)^{k-1} p \stackrel{(1-p) \leq \frac{1}{2}}{\leq} \frac{1}{2^k}, \text{ a co za tym idzie } \mathcal{P}(X_i > k) = (1 - p)^k \leq \frac{1}{2^k}. \quad \square$$

(b)

Korzystamy z podpunktu (a) dla $k := 2\lfloor \log n \rfloor$. Dostajemy $\mathcal{P}(X_i > 2\lfloor \log n \rfloor) \leq \frac{1}{2^{2\lfloor \log n \rfloor}} \leq \frac{1}{2^{2 \log n}} = \frac{1}{n^2}$. $\square$

(c)

Obserwacja. Dla dowolnego k

$$\mathcal{P}(X \geq k) = \mathcal{P}\left(\bigcup_{i \in [n]} X_i \geq k\right) \stackrel{\text{Union Bound}}{\leq} \sum_{i \in [n]} \mathcal{P}(X_i \geq k) \stackrel{(a)}{\leq} \frac{n}{2^k}.$$

W szczególności dla $k := 2\lfloor \log n \rfloor$:

$$\mathcal{P}(X \geq 2\lfloor \log n \rfloor) \stackrel{(b)}{\leq} n \cdot \frac{1}{n^2} = 1/n.$$

$\square$

(d)

Niech $T := 2\lfloor \log n \rfloor$.

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=1}^{\infty} k \cdot \mathcal{P}(X = k) \stackrel{\text{Lemma 2.9}}{=} \sum_{k=1}^{\infty} \mathcal{P}(X \geq k) \\ &= \sum_{k=1}^T \mathcal{P}(X \geq k) + \sum_{k=T+1}^{\infty} \mathcal{P}(X \geq k) \end{aligned}$$

$$\sum_{k=1}^T \mathcal{P}(X \geq k) \leq T \cdot \mathcal{P}(X \geq T) \leq T \cdot 1 = T = O(\log n)$$

$$\begin{aligned} \sum_{k=T+1}^{\infty} \mathcal{P}(X \geq k) &\stackrel{\text{Obserwacja}}{\leq} \sum_{k=T+1}^{\infty} \frac{n}{2^k} \\ &= n \sum_{k=T+1}^{\infty} \frac{1}{2^k} \\ &= n \cdot \frac{1}{2^{T+1}} \cdot \frac{1}{1 - 1/2} \\ &= 2n \cdot \frac{1}{2^{2\lfloor \log n \rfloor + 1}} \\ &\leq 2n \cdot \frac{1}{2n^2} = \frac{1}{n} \end{aligned}$$

Sumarycznie dostajemy

$$\mathbb{E}[X] = O(\log n) + \frac{1}{n} = O(\log n).$$

$\square$

Zadanie 11(Mateusz Wojaczek)

Przeczytaj Sekcję 5.6.1 *Random Graph Models* (modele losowych grafów, podobieństwo do modelu kul i urn). Wypowiedz i udowodnij Lemat 5.14 dla malejących własności grafów.

Rozwiązanie

Graf jest **monotonicznie malejący**, dla dowolnych właściwości, gdy zachodzi dla

$$G = (V, E),$$

to zachodzi też dla

$$G' = (V, E'),$$

gdzie $E' \subset E$.

Niech:

$P(n, N)$ – prawdopodobieństwo, że własność zachodzi dla grafu $G_{n,N}$,

$P(n, p)$ – prawdopodobieństwo, że własność zachodzi dla grafu $G_{n,p}$.

Niech

$$p^+ = \frac{(1 + \epsilon)N}{\binom{n}{2}}, \quad p^- = \frac{(1 - \epsilon)N}{\binom{n}{2}},$$

dla $0 < \epsilon < 1$.

Lemat. Dla monotonicznie malejących grafów właściwości zachodzi:

$$P(n, p^+) - e^{-O(N)} \leq P(n, N) \leq P(n, p^-) + e^{-O(N)}.$$

(i) **Dowód:** $P(n, p^+) - e^{-O(N)} \leq P(n, N)$

Niech X będzie zmienną losową oznaczającą liczbę krawędzi w grafie G_{n,p^+} .

$$P(n, p^+) = \sum_{k=0}^{\binom{n}{2}} P(n, k) \cdot \Pr(X = k).$$

Rozbijamy sumę:

$$P(n, p^+) = \sum_{k < N} P(n, k) \Pr(X = k) + \sum_{k \geq N} P(n, k) \Pr(X = k).$$

Dla monotonicznie malejących własności mamy, że $P(n, k) \leq P(n, N)$ dla $k \geq N$, zatem

$$P(n, p^+) \leq \Pr(X < N) + P(n, N) \Pr(X \geq N) \leq \Pr(X < N) + P(n, N).$$

Ponadto:

$$\Pr(X < N) = \Pr\left(X < \frac{1}{1+\epsilon}\mathbb{E}[X]\right) \leq \Pr\left(X < (1 + \frac{\epsilon}{2})\mathbb{E}[X]\right) \leq e^{-\frac{(1+\epsilon)\epsilon^2 N}{8}} = e^{-O(N)}.$$

Zatem:

$$P(n, p^+) \leq P(n, N) + e^{-O(N)},$$

czyli

$$P(n, p^+) - e^{-O(N)} \leq P(n, N).$$

ckd

(ii) Dowód: $P(n, N) \leq P(n, p^-) + e^{-O(N)}$

Niech X będzie zmienną losową liczby krawędzi w grafie G_{n, p^-} .

$$P(n, p^-) = \sum_{k=0}^{\binom{n}{2}} P(n, k) \Pr(X = k) = \sum_{k \leq N} P(n, k) \Pr(X = k) + \sum_{k > N} P(n, k) \Pr(X = k).$$

Z monotoniczności mamy:

$$P(n, p^-) \geq P(n, N) \Pr(X \leq N) \geq P(n, N) - \Pr(X > N).$$

Z kolei:

$$\Pr(X > N) = \Pr\left(X > \frac{1}{1-\epsilon}\mathbb{E}[X]\right) \leq \Pr(X > (1 + \epsilon)\mathbb{E}[X]) \leq e^{-\frac{(1-\epsilon)\epsilon^2 N}{3}} = e^{-O(N)}.$$

Zatem:

$$P(n, p^-) \geq P(n, N) - e^{-O(N)},$$

czyli

$$P(n, N) \leq P(n, p^-) + e^{-O(N)}.$$

ckd

Zadanie 12 (Kacper Poneta)

Dla jakiego $p \in [0, 1]$ oczekiwana liczba kopii podgrafu indukowanego $K_{3,3}$ w grafie G wylosowanym z modelu $G_{n,p}$ (definicja w książce) jest największa? Możesz założyć, że n jest wystarczająco duże, wynik może zależeć od n .

Rozwiązanie

Niech $G \sim G_{n,p}$ będzie grafem losowym w modelu Erdős–Rényiego, tzn. grafem prostym o n wierzchołkach, w którym każda z $\binom{n}{2}$ możliwych krawędzi jest niezależnie obecna z prawdopodobieństwem $p \in [0, 1]$.

Oznaczmy przez X liczbę kopii podgrafu indukowanego izomorficznego z $K_{3,3}$ w G . Dla ustalonego zbioru 6 wierzchołków, prawdopodobieństwo, że podgraf indukowany na nich jest dokładnie $K_{3,3}$, wynosi

$$p^9(1-p)^6,$$

ponieważ $K_{3,3}$ ma 9 krawędzi (między częściami dwudzielnymi) i 6 niekrawędzi (wewnątrz części).

Liczba różnych uporządkowanych podziałów 6 wierzchołków na dwie trójki wynosi

$$\binom{n}{3} \binom{n-3}{3},$$

ale każda nieuporządkowana kopia $K_{3,3}$ jest liczona dwa razy (zamiana części), więc

$$\mathbb{E}X = \frac{1}{2} \binom{n}{3} \binom{n-3}{3} p^9(1-p)^6 = 10 \binom{n}{6} p^9(1-p)^6.$$

Dla ustalonego n maksymalizujemy funkcję

$$f(p) = p^9(1-p)^6.$$

Licząc pochodną logarytmu:

$$\frac{d}{dp} \ln f(p) = \frac{9}{p} - \frac{6}{1-p} = 0 \implies 9(1-p) = 6p \implies p = \frac{3}{5}.$$

Zatem oczekiwana liczba kopii indukowanego $K_{3,3}$ w grafie $G_{n,p}$ jest największa dla

$$p = \frac{3}{5}.$$

Wartość maksymalna to

$$\mathbb{E}X_{\max} = 10 \binom{n}{6} \left(\frac{3}{5}\right)^9 \left(\frac{2}{5}\right)^6.$$

Odpowiedź

$$\boxed{p = \frac{3}{5}}$$