Kule i urny; Aproksymacja Poissona

Zadanie 1. Rozkład *Poissona* z parametrem λ dany jest funkcją $\mathbf{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ dla $k \in \{0, 1, \ldots\}$. Wyznacz funkcję tworzącą momenty tego rozkładu i na jej bazie wartość oczekiwaną i wariancję.

Zadanie 2. Niech X będzie zmienną losową z rozkładu Poissona o wartości oczekiwanej $\mathbf{E}[X] = \mu$. 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.

Zadanie 3. Rzucamy m kul w sposób jednostajny i niezależny do n urn, ponumerowanych $0, \ldots, n-1$. Mówimy, że w indeksie i jest k-luka, jeśli wszystkie urny o indeksach $i, i+1, \ldots, i+k-1$ są puste.

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

Zadanie 4. Niech X jest zmienną Poissona z wartością oczekiwaną $\mu \in \{1, 2, \ldots\}$.

(i) Pokaż, że $\mathbf{P}(X = \mu + h) \geqslant \mathbf{P}(X = \mu - h - 1)$ dla $0 \leqslant h \leqslant \mu - 1$.
(ii) Na bazie poprzedniego podpunktu, wyargumentuj że $\mathbf{P}(X \geqslant \mu) \geqslant \frac{1}{2}$.

Zadanie 5. Rzucamy n kul jednostajnie i niezależnie do n kubełków. Rozważ prawdopodobieństwo tego, że w każdym kubełku jest dokładnie jedna kula.

(i) Na bazie aproksymacji Poissona podaj ograniczenie górne na to prawdopodobieństwo.
(ii) Wyznacz dokładny wynik.
(iii) Pokaż, że wartości wyznaczone w poprzednich podpunktach różnią się od siebie stałą multiplikatywną, która jest równa prawdopodobieństwu tego, że zmienna Poissona o parametrze n przyjmuje wartość n.

Zadanie 6. Rzucamy n kul jednostajnie i niezależnie do n kubełków.

(i) Wyznacz prawdopodobieństwo tego, że w pierwszym kubełku jest 1 kula, przy założeniu, że dokładnie jedna kula znajduje się w pierwszych trzech kubełkach.
(ii) Wyznacz oczekiwaną liczbę kul w pierwszym kubełku, przy założeniu, że drugi kubełek jest pusty.
(iii) Wyznacz prawdopodobieństwo tego, że w pierwszym kubełku jest więcej kul niż w drugim kubełku.

Zadanie 7. Niech zmienna $X_i^{(m)}$ oznacza liczbę kul w i-tym kubełku, przy założeniu, że rzucamy m kul do n kubełków w sposób jednostajny i niezależny. Niech zmienne $Y_1^{(m)}, \ldots, Y_n^{(m)}$ będą niezależnymi zmiennymi z rozkładu Poissona o średniej $\frac{m}{n}$. Podczas wykładu widzieliśmy, że dla każdej nieujemnej funkcji f zachodzi

$$\mathbf{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right] \geqslant \mathbf{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \mathbf{P}\left(\sum_{i=1}^{n} Y_i^{(m)} = m\right)$$

(i) Pokaż, że jeśli $\mathbf{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ jest rosnąca względem m, to przy założeniu nieujemnośći f zachodzi:

$$\mathbf{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right] \geqslant \mathbf{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \mathbf{P}\left(\sum_{i=1}^{n} Y_i^{(m)} \geqslant m\right)$$

(ii) Wykorzystując poprzedni podpunkt wykaż, że jeśli $\mathbf{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ jest monotoniczna względem m to

$$\mathbf{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \leqslant 2 \cdot \mathbf{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right].$$

Zadanie 8. Przeczytaj sekcje 5.5.1 *Chain Hashing* i 5.5.2 *Hashing: Bit Strings* (przykład zastosowania modelu kul i urn). Rozwiąż zadanie 5.22 z książki. Uwaga: w podpunkcie (b) powinno być $\lfloor \log n \rfloor$ zamiast $\log n$!

Zadanie 9. Przeczytaj sekcję 5.5.3 *Filtry Blooma*. Rozwiąż zadania 5.23 i 5.24 z książki.