Metody Probabilistyczne i Statystyka

LISTA 7

Informatyka algorytmiczna (I st.) WIT – 2024 / 2025

ZADANIA PODSTAWOWE

Zadanie 1. Pokaż przykład dyskretnych zmiennych losowych X i Y, które nie są niezależne, ale zachodzi $\mathbf{E}(XY) = \mathbf{E}(X)\,\mathbf{E}(Y)$.

Zadanie 2. Pokaż, że $\mathbf{var}(aX + b) = a^2\,\mathbf{var}(X)$ dla $a, b \in \mathbb{R}$.

Zadanie 3. Wyznacz wariancję

(a) zmiennych losowych X i Y z zadania 1 z listy 6,

(b) zmiennej losowej o rozkładzie geometrycznym $\mathrm{Geo}(p)$,

(c) zmiennej losowej o rozkładzie Poissona $\mathrm{Po}(\lambda)$.

Zadanie 4. Niezależnie powtarzamy próbę ethernetową z optymalną wartością parametru p aż do pierwszego momentu, w którym dowolne z urządzeń uzyska dostęp do zasobu (patrz zadanie 8 z listy 3). Jaki rozkład ma zmienna losowa reprezentująca liczbę wykonanych powtórzeń? Korzystając z wcześniej wyznaczonych wzorów, oblicz jej wartość oczekiwaną i wariancję.

Zadanie 5. Źródło generuje losową liczbę N pakietów zgodnie z rozkładem Poissona z parametrem λ. Każdy pakiet dociera do celu niezależnie z prawdopodobieństwem p. Wyznacz rozkład liczby pakietów, które dotrą do celu.

Zadanie 6. Niech $X \sim \mathrm{Geo}(p)$ oraz $Y \sim \mathrm{Geo}(r)$ będą niezależnymi zmiennymi losowymi o rozkładzie geometrycznym. Wyznacz rozkład zmiennej losowej $Z = \min\{X, Y\}$.

Zadanie 7. (*Własność braku pamięci, ang. memoryless property*) Niech $X \sim \mathrm{Geo}(p)$.

(a) Pokaż, że dla dowolnych $m, n \in \mathbb{N}$ zachodzi $P(X = n + m \mid X > m) = P(X = n)$.

(b) Zaproponuj *rozsądną* interpretację powyższego faktu.

Zadanie 8. Niech R oznacza IQ losowo spotkanej osoby na ulicy. Załóżmy, że wiadomo, iż $\mathbf{E}(R) = 100$ oraz $\mathbf{var}(R) = 100$. Zastosuj nierówność Markowa oraz Czebyszewa do oszacowania z góry $P(R \geqslant 200)$.

ZADANIA DODATKOWE

W zadaniach 9, 10 i 11 zakładamy, że kule wrzucane są niezależnie i jednostajnie losowo do n ponumerowanych urn (rozważamy model jak w zadaniu domowym 2).

Zadanie 9. Niech U_n^m oznacza liczbę pustych urn po wrzuceniu m kul.

(a) Wyznacz $\mathbf{E}(U_n^m)$.

(b) Oznaczmy $u_1(n) = \mathbf{E}\left(U_n^n\right)$ oraz $u_2(n) = \mathbf{E}\left(U_n^m\right)$ dla $m = n \ln n$. Wyznacz $\lim\limits_{n\to\infty} \frac{u_1(n)}{n}$ oraz $\lim\limits_{n\to\infty} \frac{u_2(n)}{n}$.

Zadanie 10. (*Birthday paradox*) Niech B_n oznacza moment pierwszej kolizji.

(a) Wyznacz wzór na $\mathbf{E}\left(B_n\right)$.

HINT: Wyznacz $P(B_n > k)$ dla $k \in \mathbb{N}$ i skorzystaj z zadania 6 z listy 6.

(b) Korzystając z podobnych oszacowań jak w zadaniu 5 z listy 0, wyprowadź ograniczenie górne na $\mathbf{E}\left(B_n\right)$ postaci $\alpha + \sum_k e^{f(n,k)}$, gdzie α jest pewną stałą, a f funkcją n i k. Następnie oszacuj z góry otrzymaną sumę przed odpowiednią całkę i wyznacz ograniczenie górne na wartość oczekiwaną momentu pierwszej kolizji postaci $\mathbf{E}\left(B_n\right) \leqslant an^b + c$ dla odpowiednich stałych a, b i c.

HINT: Do obliczenia całki, która pojawi się we wzorze, użyj wybranego pakietu matematycznego lub skorzystaj z faktu, że $\int_0^\infty e^{-x^2}\mathrm{d}x = \frac{\sqrt{\pi}}{2}$.

UWAGA: Stosując nieco bardziej subtelne techniki pozwalające na uzyskanie dokładniejszych oszacowań i przeprowadzając żmudne obliczenia można pokazać, że

$$\mathbf{E}\left(B_n\right) = \sqrt{\frac{\pi n}{2}} + \frac{2}{3} + \frac{1}{12}\sqrt{\frac{\pi}{2n}} + \mathrm{O}\left(\frac{1}{n}\right).$$

★ (c) Załóżmy, że każda z m osób ma urodziny niezależnie z jednakowym prawdopodobieństwem w każdym z 365 dni (przyjmijmy, że nie ma lat przestępnych). Niech A_m oznacza zdarzenie, że co najmniej dwie osoby mają urodziny tego samego dnia. Wyznacz $P(A_m)$ i oblicz jego wartość dla m równego liczbie osób obecnych na dzisiejszych zajęciach. Jaka jest najmniejsza wartość m, dla której $P(A_m) > 0{,}5$?

Zadanie 11. (*Coupon collector's problem*) Niech C_n oznacza minimalną liczbę rzutów, po której nie ma już pustych urn.

(a) Wyznacz $\mathbf{E}\left(C_n\right)$ (podaj możliwie zwartą postać wzoru) oraz asymptotykę $\mathbf{E}\left(C_n\right)$ (podaj kolejne składniki w rozwinięciu asymptotycznym aż do $\mathrm{O}\left(\frac{1}{n}\right)$).

HINT: Podziel proces na n rund, gdzie k-ta runda rozpoczyna się w pierwszym momencie, w którym mamy $k-1$ niepustych urn i trwa do momentu, w którym wrzuciliśmy kulę do pewnej pustej urny (czyli do momentu zapełnienia kolejnej urny). Niech X_k będzie liczbą kul wrzuconych podczas k-tej rundy. Wyznacz rozkład zmiennych losowych X_k i wyraź C_n przy pomocy X_k.

(b) Wyznacz dokładny wzór na $\mathbf{var}\left(C_n\right)$ i pokaż, że $\mathbf{var}\left(C_n\right) \leqslant \frac{\pi^2}{6} n^2$.

HINT: Zauważ, że zmienne losowe X_k z podpunktu (a) są niezależne.

(c) Korzystając z nierówności Czebyszewa oraz ograniczenia górnego na wariancję z podpunktu (b) oszacuj $P(|C_n - \mathbf{E}\left(C_n\right)| \geqslant \alpha \cdot n\, H_n)$, gdzie α jest pewną dodatnią stałą, a H_n jest n-tą liczbą harmoniczną. Co na tej podstawie możesz powiedzieć o koncentracji zmiennej losowej C_n wokół wartości oczekiwanej?