Wariancja, nierówność Czebyszewa i nierówność Czernowa

Zadanie 1. Algorytm bąbelkowy sortowania przegląda podaną na wejściu permutację i odwraca sąsiadujące elementy będace w inwersji. Przejścia wzdłuż permutacji są powtarzane tak długo, aż permutacja jest posortowana. Załóż, że podajemy temu algorytmowi losowo wybraną permutację zbioru $\{1, \ldots, n\}$. 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 $\Omega\left(n^2\right)$.

Zadanie 2. Każdy z n graczy rzuca niezależnie od pozostałych kostką.

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

Zadanie 3. W urnie znajduje się b białych kul i c czarnych kul. Wyciągamy kolejno wszystkie kule. Wyciągnięta kula daje *zmianę*, jeśli jej kolor jest różny od koloru kuli wyciągniętej bezpośrednio przed nią. Wyciągnięta biała kula daje *dobrą zmianę*, jeśli bezpośrednio przed nią została wyciagnięta czarna kula.

(i) Jaka jest oczekiwana liczba i wariancja zmian?
(ii) Jaka jest oczekiwana liczba i wariancja dobrych zmian?

Zadanie 4. Ze standardowej talii 52 kart wyciągamy 13 (ręka brydżowa). Niech X to liczba wyciagniętych asów, a Y to liczba wyciągniętych caro.

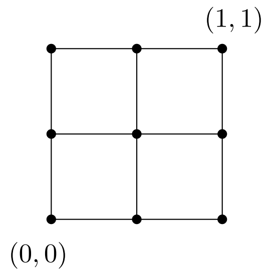
(i) Czy X i Y są niezależne?
(ii) Czy X i Y są nieskorelowane?

(Zmienne X i Y są *nieskorelowane* jeśli $\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$.)

Zadanie 5. Rzucamy kostką, aż nie zanotujemy dwóch szóstek pod rząd. Jaka jest oczekiwana liczba rzutów? Wskazówka: Odpowiedzią nie jest 36.

Zadanie 6. Wiewiórka chodzi po skończonej kracie jak na Rysunku 1 w sposób losowy. Rozpoczyna swoją trasę w wierzchołku $(0, 0)$ i w każdym kroku wybiera w sposób losowy, jednostajny i niezależny od innych wyborów jeden z sąsiednich wierzchołków aktualnego wierzchołka. W wierzchołku $(1, 1)$ znajduje się żołądź. Jaka jest oczekiwana liczba kroków w której wiewiórka znajdzie żołędzia?

Zadanie 7. Rzucamy monetą symetryczną n razy i otrzymujemy ciąg zer i jedynek 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 $X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{ij}$.



Rysunek 1: Mały świat wiewiórki.

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

Zadanie 8 (Słabe prawo wielkich liczb)**.** Niech $X_1, X_2, \ldots$ 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 $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| > \varepsilon\right) = 0$$

Zadanie 9. Wyznacz postać jawną funkcji tworzącej momenty dla rozkładów:

(i) jednostajnego na zbiorze $\{1, \ldots, n\}$,
(ii) dwumianowego $\mathrm{Binom}(n, p)$.

Na jej podstawie wyznacz wartość oczekiwaną i wariancję tych rozkładów. Dla rozkładu dwumianowego tworzącą wyznacz na dwa sposoby:

(i) oblicz $\mathrm{E}(e^{tX})$ z definicji wartości oczekiwanej,
(ii) skorzystaj z niezależności zmiennych Bernouliego.

Zadanie 10. Alicja i Bob rozgrywają pomiędzy sobą n pojedynków. Alicja jest lepsza i każdy pojedynek wygrywa niezależnie od pozostałych z prawdopodobieństwem 0.6. Zwycięzcą jest osoba która wygra więcej pojedynków. Za pomocą nierówności Chernoffa ogranicz prawdopodobieństwo tego, że zwycięzcą jest Bob.

Zadanie 11. Dana jest moneta asymetryczna z nieznanym Ci $p \in (0, 1)$. Jak wiele rzutów musisz wykonać, aby z prawdopodobieństwem co najmniej 0.9 poznać wartość p z dokładnością do 0.1?

Zadanie 12. Niech X będzie zmienną losową mówiącą ile razy spośród n niezależnych rzutów kostką, wypadła 6-tka. Ogranicz $\mathrm{P}(X \geqslant n/4)$ za pomocą nierówności Markova, Czebyszewa oraz Chernoffa. Podaj wartości numeryczne dla $n = 1000$.

Zadanie 13. Rzucamy n razy sprawiedliwą sześciościenną kostką. Podaj jak najmniejszą asymptotycznie funkcję $g(n)$ taką, że

$$P\left(\text{wypadło w sumie co najmniej } \frac{7}{2} \cdot n + g(n) \text{ oczek}\right) \leqslant \frac{c}{n},$$

dla pewnej stałej c.

Zadanie 14. Kasyno testuje nową maszynę. Aby zagrać gracz płaci \$1, zaś maszyna z prawdopodobieństwem $\frac{4}{25}$ wypłaca \$3, z prawdopodobieństwem $\frac{1}{200}$ wypłaca \$100 a w pozostałych przypadkach nic nie wypłaca. Ku zaskoczeniu pracowników kasyna w ciagu pierwszego miliona gier maszyna przegrała \$10,000. Oszacuj przy pomocy nierówności Chernoffa prawdopodobieństwo zajścia takiego zdarzenia. Numerycznie wyznacz najlepsze możliwe ograniczenie (znajdź odpowiednie t).

Zadanie 15. Niech $X = X_1 + \ldots + X_n$, gdzie X_i są niezależnymi zmiennymi losowymi z rozkładu jednostajnego na zbiorze $\{0, 1, 2\}$. Wyprowadź nierówność Chernoffa dla zmiennej X, tj. ogranicz od góry następujące prawdopodobieństwa: $\mathrm{P}\left(X \geqslant (1+\delta)\, n\right)$ oraz $\mathrm{P}\left(x \leqslant (1-\delta)\, n\right)$, dla $0 < \delta < 1$.

Zadanie 16. Rzucamy n kul w sposób jednostajny i niezależny do n urn. Niech X_i będzie indykatorem na zdarzenie - i-ta urna jest pusta. Niech $Y_1, \ldots, Y_n$ będą niezależnymi zmiennymi wskaźnikowymi z prawdopodobieństwem sukcesu $p = \left(1 - \frac{1}{n}\right)^n$. Niech $X = \sum_{i=1}^{n} X_i$ oraz $Y = \sum_{i=1}^{n} Y_i$.

(i) Pokaż, że dla każdego $k \geqslant 1$, zachodzi $\mathrm{E}(X_1 X_2 \cdots X_k) \leqslant \mathrm{E}(Y_1 Y_2 \cdots Y_k)$.
(ii) Pokaż, że dla każdego $t \geqslant 0$, zachodzi $\mathrm{E}(e^{tX}) \leqslant \mathrm{E}(e^{tY})$.
(iii) Wyznacz ograniczenie Chernoffa dla $\mathrm{P}(X \geqslant (1+\delta)\,\mathrm{E}(X))$.

Zadanie 17 (Randomizowany QuickSort)**.** Spójrzmy na QuickSorta z innej strony. Każdy element x należący do sortowanej tablicy jest wybierany na *pivota* dokładnie raz podczas działania algorytmu. Zaczynamy od nieposortowanej n-elementowej tablicy. Losujemy element x, który będzie pełnił rolę pivota podziału. Niech ten element x stanie się korzeniem drzewa binarnego budowanego 'na boku'. Jego lewym poddrzewem będzie drzewo binarne wytworzone rekurencyjnie z elementów mniejszych od x w podziale, zaś prawym poddrzewem będzie analogiczne drzewo powstałe z elementów większych od x w podziale. Tak więc konkretnemu wykonaniu algorymu QuickSort odpowiada dobrze określone drzewo binarne kodujące historię wyborów elementów na pivoty. Mówimy, że element x jest *zły* jeśli w momencie wyboru na pivota dzieli tablicę na dwie części z których jedna jest conajmniej dwa razy większa od drugiej. W języku drzewa przekłada się to na fakt, że jedno z poddrzew jest conajmniej 2 razy większe na liczność od drugiego. Dopełniając opisu x jest *dobry* jeśli nie jest *zły*.

(i) Pokaż, że na konkretnej ścieżce z korzenia do liścia drzewa, liczba *dobrych* węzłów jest niewiększa niż $c \log_2 n$ dla pewnego $c > 0$.

(ii) Pokaż, że z prawdopodobieństwem co najmniej $1 - \frac{1}{n^2}$ liczba węzłów na konkretnej ścieżce od korzenia do liścia jest niewiększa niż $c' \log_2 n$ dla pewnego $c' > 0$.

(iii) Pokaż, że z prawdopodobieństwem co najmniej $1 - \frac{1}{n}$ liczba węzłów na najdłuższej ścieżce od korzenia do liścia jest niewiększa niż $c' \log_2 n$.

(iv) Uzasadnij, że z prawdopodobieństwem co najmniej $1 - \frac{1}{n}$ czas działania QuickSorta to $\mathcal{O}(n \log n)$.

Zadanie bonusowe do spisania

Rozważamy algorytm trasowania pakietów na hiperkostce. Hiperkostka jest to graf złożony z zerojednykowych ciągów długości n, gdzie krawędź łączy dwa ciągi, gdy różnią się one dokładnie na jednej pozycji. W każdym wierzchołku kładziemy pakiet i dla każdego pakietu wybieramy unikalny cel. W jednej jednostce czasu tylko jeden pakiet może używać danej krawędzi. Najprostszy algorytm polega na sukcesywnym "naprawianiu" bitów tak aby z każdym krokiem zbliżać sie do celu (BitFixing). Okazuje się, że taki algorytm nie jest optymalny i dużo lepiej działa jego następująca nieoczywista modyfikacja. Dla każdego pakietu losujemy wierzchołek pośredni, wysyłamy pakiet do niego używając BitFixingu, a potem z wierzchołka pośredniego wysyłamy pakiet do wierzchołka końcowego również używając BitFixingu.

Problem 1. Przeczytaj i zrozum analizę powyżej przedstawionego algorytmu w książce (sekcja 4.6.1). W ramach weryfikacji należy napisać (indywidualne) oświadczenie, że się przeczytało i zrozumiało materiał.

Problem 2. Znajdź instancję problemu (cele kazdego pakietu), gdzie algorytm BitFixing (formalna definicja w książce) wykonuje $\Omega(\sqrt{2^n})$ kroków.

Problem 3. Rozważ następującą modyfikację algorytmu BitFixing. Zamiast naprawiać bity w kolejności 1 do n, każdy pakiet wybiera swoją lokalną losową kolejność, niezależnie od pozostałych. Pokaż, że istnieje permutacja (wybór celów), dla której algorytm wymaga $\Omega(2^n)$ kroków z dużym prawdopodobieństwem.

Każdy z trzech problemów jest warty $\frac{1}{3}$ możliwych punktów za to zadanie. Przypominamy, że rozwiązania spisane **na komputerze** należy wysyłać na adres podany na stronie internetowej kursu w terminie do środy 8.11.2023 do godziny 23:59.