RP dla bioinformatyki, pisemne zadania domowe II

Termin: 21.12.2018, godz. 17.00

1. Poniżej a, b, c, d, e, f to kolejne cyfry numeru indeksu osoby rozwiązującej zadanie.
Wektor losowy (X, Y) ma średnią $[c, d]$ i macierz kowariancji

$$\begin{bmatrix} e+1 & 1 \\ 1 & f+1 \end{bmatrix}.$$

a) Czy X i Y są niezależne? Obliczyć $\mathbb{E}XY$ oraz $\mathrm{Var}(X + 2Y + 3)$.
b) Załóżmy dodatkowo, że wektor losowy (X, Y) ma rozkład normalny (o powyższych parametrach).
   b1) Jaki rozkład ma wtedy $X + 2Y + 3$? Wypisać gęstość tej zmiennej losowej.
   b2) Znaleźć taką stałą $\alpha \in \mathbb{R}$, żeby zmienne X oraz $X + \alpha Y$ były niezależne.

2. Zmienne losowe X, Y są niezależne i mają rozkład Poissona z parametrami $\lambda > 0$ i $\mu > 0$ odpowiednio.

a) Wyznaczyć rozkład wektora losowego $(X, X + Y)$ – ile jest równe $\mathbb{P}\big((X, X + Y) = (k, n)\big)$ dla nieujemnych całkowitych k, n?
b) Czy zmienne losowe X i $X + Y$ są niezależne?
c) Ustalmy nieujemną liczbę całkowitą n. Wyznaczyć rozkład warunkowy X pod warunkiem $X + Y = n$, tzn. znaleźć

$$\mathbb{P}(X = k | X + Y = n) \quad \text{dla } k = 0, 1, 2, \ldots$$

d) Obliczyć warunkową wartość oczekiwaną X pod warunkiem $X + Y = n$, tzn. obliczyć

$$\sum_{k=0}^{\infty} k\mathbb{P}(X = k | X + Y = n).$$

Wskazówka. Sprawdzaliśmy kiedyś, jaki rozkład ma $X + Y$.

3. Załóżmy, że zmienne losowe X i Y są całkowalne z kwadratem (i że znamy łączny rozkład wektora losowego (X, Y) lub przynajmniej niektóre jego parametry). Przypuśćmy, że łatwo jest obserwować zmienną X, a zmienna Y jest trudna do zmierzenia. Chcemy jak najlepiej przybliżać wartości zmiennej losowej Y funkcją liniową postaci $aX + b$, $a, b \in \mathbb{R}$, a błąd będziemy mierzyć w sensie średniokwadratowym. Innymi słowy: szukamy stałych $a, b \in \mathbb{R}$, dla których wartość funkcji

$$f(a, b) = \mathbb{E}\big[(Y - aX - b)^2\big]$$

jest najmniejsza. Poniżej można przyjąć, że $\mathrm{Var}(X) > 0$, tzn. że zmienna losowa X nie jest stała p.n. (bo inaczej zadanie się trywializuje (jak?)).

a) Uzasadnić, że przy ustalonym a funkcja $b \mapsto f(a, b)$ osiąga swoje minimum w punkcie $b = \mathbb{E}(Y - aX)$.

b) Niech więc

$$h(a) := f\big(a, \mathbb{E}(Y - aX)\big).$$

Sprawdzić, że $h(a)$ to funkcja kwadratowa zmiennej a i wyrazić jej współczynniki w terminach $\mathrm{Var}(X)$, $\mathrm{Cov}(X, Y)$, $\mathrm{Var}(Y)$.

c) Dla jakiego a jest przymowane minimum funkcji h?

4. a) Rzucamy sześcienną kostką aż wyrzucimy wszystkie możliwe wyniki. Jaka jest wartość oczekiwana liczby rzutów?

b) A wariancja?

c) Rzucamy kostką o stu ściankach aż wyrzucimy wszystkie możliwe wyniki. Jaka jest wartość oczekiwana liczby rzutów? Wypisać odpowiedni wzór i podać przybliżoną wartość liczbową.

d) Rzucamy kostką o stu ściankach aż wyrzucimy 95 różnych wyników (dowolnych, nie ustalamy z góry liczb, na które czekamy). Jaka jest wartość oczekiwana liczby rzutów? Wypisać odpowiedni wzór i podać przybliżoną wartość liczbową.

5. Niech n będzie ustaloną liczbą naturalną. Zmienne losowe $X_1, \ldots, X_n$ są niezależne i każda z nich ma rozkład $\mathcal{N}(0,1)$, czyli $X = (X_1, \ldots, X_n)^T$ to wektor losowy o rozkładzie $\mathcal{N}(0, I)$ (transponujemy, bo chemy żeby był wektorem pionowym). Oznaczmy

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n},$$

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\Big(\sum_{i=1}^{n} X_i^2\Big) - \bar{X}^2$$

(są to tzw. średnia i wariancja z próby). Niech A będzie macierzą ortonormalną o wymiarach $n \times n$ taką, że jej pierwszy wiersz to $(\frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}})$. Niech $Y = (Y_1, \ldots, Y_n)^T := AX$.

a) Korzystając z wiedzy z ćwiczeń (wykładu?) uświadomić sobie, że wektor losowy Y też ma rozkład $\mathcal{N}(0, I)$ i w szczególności jego współrzędne są niezależne.

b) Sprawdzić, że $\sqrt{n}\bar{X} = Y_1$ i wyrazić nS^2 w terminach $Y_1, Y_2, \ldots, Y_n$.

c) Wywnioskować, że zmienne losowe $\bar{X}$ oraz S^2 niezależne, a także że nS^2 ma rozkład χ^2_{n-1} (chi-kwadrat o $n-1$ stopniach swobody, patrz seria V).

Uwaga+wskazówka. Macierz A jest ortonormalna normalna wtedy i tylko wtedy, gdy $A^T A = AA^T = I$, tzn. gdy wiersze są prostopadłe i mają długość jeden (i tak samo dla kolumn). Ponadto wtedy dla każdego wektora $v \in \mathbb{R}^n$ mamy $\|Av\| = \|v\|$.

6. Poniżej K_n oznacza graf pełny o n wierzchołkach i $\binom{n}{2}$ krawędziach.

Ustalmy $n \geqslant 2$. Każdą krawędź grafu K_n malujemy (niezależnie) losowo wybranym kolorem: na niebiesko lub na czerwono.

Ustalmy też $k \in \{2, ..., n\}$. Poniżej rozważamy tylko podgrafy G grafu K_n, które mają dokładnie k wierzchołków i wszystkie $\binom{k}{2}$ krawędzi między nimi (a więc G też jest grafem pełnym). Dla takiego grafu G niech X_G będzie zmienną losową zadaną przez

$$X_G = \begin{cases} 1 & \text{jeśli wszystkie krawędzie } G \text{ są w jednym kolorze,} \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

a) Uzasadnić, że $\mathbb{E}X_G = 2 \cdot 2^{-\binom{k}{2}}$.

b) Niech Y będzie liczbą podgrafów K_n o k wierzchołkach, których wszystkie krawędzie są jednego koloru. Wyznaczyć $\mathbb{E}Y$.

c) Korzystając z faktu, że Y przyjmuje tylko wartości całkowite nieujemne, uzasadnić, że jeśli $\mathbb{E}Y < 1$, to *musi istnieć* co najmniej jedno kolorowanie krawędzi, takie że żaden podgraf G o k wierzchołkach nie ma wszystkich krawędzi tego samego koloru.

d) Jaś narysował K_{11} i chce pokolorować każdą krawędź jednym z dwóch kolorów (czerwonym lub niebieskim), tak żeby żaden podgraf powstały poprzez wybranie 5 wierzchołków i wszystkich krawędzi między nimi nie miał wszystkich krawędzi w tym samym kolorze. Czy może to zrobić?