Изпит
Вероятности и статистика (ВиС)

05.07.2024

Време за работа: 180 минути

Указания за работа

При работата върху задачите може да реферирате към теореми и твърдения, които помагат за извеждането на някоя стъпка. Точките имат индикативен характер. За събеседване е необходимо ≥ 1 точка.

Задачи

Задача 1.

(1) *Дефинирайте дискретна случайна величина и въведете нейното математическо очакване и дисперсия.* **(0.2 т.)**

(2) *Дефинирайте независимост на две дискретни случайни величини* X, Y. **(0.1 т.)**

(3) *За горните* X, Y *намерете възможните стойности на* $X+Y$ *и изразете техните вероятности чрез вероятностите на стойностите на* X, Y. **(0.2 т.)**

(4) *Докажете вероятностно, че за всеки* $n \geq 2$ *различни положителни числа* $(x_1, x_2, \ldots, x_n)$

$$\frac{\sum_{i=1}^{n} x_i}{n^{\frac{3}{4}}} \leq \sqrt[4]{\sum_{j=1}^{n} x_j^4}. \text{ (0.5 т.)}$$

Задача 2.

(1) *Въведете схема на Бернули.* **(0.1 т.)**

(2) *Чрез схемата на Бернули с параметър* p *въведете геометрична случайна величина с параметър* p *и разпишете нейната таблица.* **(0.1 т.)**

(3) *Въведете* $X \sim NB(r, p), r \geq 1$, *и пресметнете нейната пораждаща функция.* **(0.2 т.)**

(4) *Нека* $X \sim NB(r_1, p), Y \sim NB(r_2, p), r_1, r_2 \geq 1$, *са независими случайни величини. Намерете разпределението на* $X+Y$. **(0.2 т.)**

(5) *Нека* $(X_n)_{n\geq 1}$ *е редица от независими в съвкупност геометрични случайни величини с параметър* $9/10$, *където* X_n *описва нивото на замърсяване през фиксиран час, индексиран с* n. *Часът* n *се счита с опасно замърсяване, ако* $X_n \geq 3$. *Има проблем със замърсяването, ако измежду първите* 1000 *часа има поне* 5 *часа с опасни нива на замърсяването. Намерете приблизително вероятността да има проблем със замърсяването.* **(0.4 т.)**

Задача 3.

(1) *Формулирайте централната гранична теорема и илюстрирайте нейно приложение.* **(0.2 т.)** → илюстрация? какво се има предвид?

(2) *Докажете централната гранична теорема.* **(0.2 т.)**

(3) *Двама души играят две различни тактови игри*[1], *които независимо една от друга и независимо от различните тактове на игрите, дават случаен брой точки, означени с* $(X_n)_{n\geq 0}, (Y_n)_{n\geq 0}$. *Кумулативният точков резултат след* n *игри на първия човек е* $V_n :=$

[1] Първият играе първата игра, а вторият втората.

$\sum_{j=0}^{n-1} X_j$, *а на втория* $W_n := \sum_{j=0}^{n-1} Y_j$. *Нека* $X_1 \sim Exp(a), a > 0$, *и* $Y_1 \sim U(0,b), b > 0$, *и* $A_{n,m} = \{V_n > W_m\}$. *Намерете всички* a, b, *такива, че за* $n = 1000000$ *приблизително:*

(а) $P(A_{n,2n}) = 1$; ***(0.2 т.)***

(б) $P(A_{n,4n}) = \frac{1}{2}$; ***(0.2 т.)***

(в) $P(A_{2n,3n}) = \frac{1}{4}$. ***(0.2 т.)***

Задача 4.

(1) *Въведете кумулативната функция на разпределение на случайна величина* X *и разпишете нейните свойства.* ***(0.1 т.)***

(2) *Какво означава* $X \overset{d}{=} Y$, *където* X, Y *са случайни величини.* ***(0.1 т.)***

(3) *Нека*

$$\left\{\begin{array}{ll} F(x) = 1 + \sin(x), & \text{ако } x \in (a,b), b > a \geq 0, \\ F(x) = 0, & \text{ако } x \leq a, \\ F(x) = 1, & \text{ако } x \geq b. \end{array}\right\}$$

Вярно ли е, че съществуват a, b, *такива, ча* F *е функция на разпределение на случайна величина и, ако да, намерете всички такива* a, b? ***(0.2 т.)***

(4) *Докажете, че сходимост по вероятност влече сходимост по разпределение.* ***(0.3 т.)***

(5) *Нека* $S_n := \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}_n^{(1)}\right)^2$, *където* $(X_n)_{n\geq 1}$ *е редица от независими и еднакво разпределени случайни величини с* $DX_1 = \sigma^2 < \infty$ *и* $\bar{X}_n^{(1)} = \frac{1}{n}\sum_{j=1}^{n} X_j$. *Докажете, че при* $n \to \infty$, S_n *се схожда по вероятност към* σ^2. ***(0.3 т.)***

Задача 5. *Нека* X *е произволна случайна величина.*

(1) *Формулирайте неравенството на Чебишев.* ***(0.1 т.)***

(2) *Ако* X *е неотрицателна и* $\mathbb{E}[X] < \infty$, *покажете, че за* $a > 0$,

$$\mathbb{P}(X \geq a\mathbb{E}[X]) \leq \frac{1}{a}. \textit{ (0.2 т.)}$$

(3) *Ако* $X \sim Be(p), p \in (0,1)$, *покажете, че за всяко* $b > 0$ *и всяко* $\lambda > 0$

$$\mathbb{P}(X \geq p + b) \leq \left(pe^{\lambda} + 1 - p\right)e^{-\lambda(p+b)}. \textit{ (0.2 т.)}$$

(4) *Ако* $X \sim Bin(n,p), p \in (0,1), n \geq 1$, *покажете, че за всяко* $\lambda > 0, \epsilon > 0$

$$\mathbb{P}(X > n(p+\epsilon)) \leq f(\epsilon, n, \lambda)^2$$

и намерете λ, *което минимизира оценката отгоре. Как би могло да се използва това?* ***(0.7 т.)***

Задача 6. *Нека* $\overrightarrow{X} = (X_1, X_2, \ldots, X_n)$ *са наблюдения над случайна величина* X *с параметризирана вероятностна плътност* $f(x;\theta) = C_\theta x^{\theta-1}, x \in (0,1), \theta > 0$. *Истинската стойност е* $\theta_0 > 0$, *но не я знаем.*

(1) *Намерете* C_θ *за* $\theta > 0$. ***(0.1 т.)***

(2) *Въведете понятието максимално правдоподобна оценка и намерете максимално правдоподобната оценка за* θ_0. ***(0.2 т.)***

(3) *Ако* $L(\overrightarrow{X}, \theta)$ *е функцията на правдоподобие, покажете, че за всяко* $\theta > 0$, $\left(L(\overrightarrow{X}, \theta)\right)^{\frac{1}{n}}$ *се схожда почти сигурно при* $n \to \infty$. ***(0.3 т.)***

(4) *Намерете горната граница в зависимост от* θ *и покажете, че тя се максимизира за* $\theta = \theta_0$. ***(0.4 т.)***

[2]Използвайте горното.

ВиС 2024

1.

1. Дефинирайте дискретна сл. вел и въведете нейното математическо очакване и дисперсия.

↳ Нека V е вероятностно пространство и $\bar{X}$ и H са дадени. Тогава $X = \sum_{i=1}^{n} x_i \cdot 1_{H_i}$ $\left(X = \sum_{i=1}^{\infty} x_i \cdot 1_{H_i} \right)$ се нарича дискретна сл. величина.

* $\bar{X} = \begin{cases} \{x_1, x_2, \dots, x_n\} & \text{— n различни числа} \\ (x_1, x_2, \dots) & \text{— изброимо много различни числа} \end{cases}$

$H = \begin{cases} \{H_1, \dots, H_n & \text{— пълна група от събития във V} \\ (H_i)_{i \geq 1}, & \text{където } H_i \in A,\ H_i \cap H_j = \emptyset,\ i \neq j,\ \bigcup_{i=1}^{\infty} H_i = \Omega \end{cases}$

↳ Нека X е дискретна сл. величина. Тогава под очакване на X или EX разбираме:

$$EX = \sum_{i \geq 1} x_i \cdot p_i \quad \text{ако} \quad \sum_i |x_i| \cdot p_i < \infty$$

↳ Нека X е дискретна сл. величина с очакване EX. Тогава $DX := \sum_i |x_i - EX|^2 p_i$, която $DX < \infty$, е дисперсия на X.

2. Дефинирайте независимост на две дискретни случайни величини X и Y

↳ Нека X и Y са две дискретни случайни величини в едно вероятностно пространство V.
Тогава $X \perp\!\!\!\perp Y \iff P(X = x_i \cap Y = y_j) = P(X = x_i) P(Y = y_j)$
за всеки две възможни стойности x_i и y_j

3. За горните X, Y намерете възможните стойности на $X+Y$ и изразете техните вероятности чрез вероятностите на X, Y

↳ Ако X и Y са независими, тогава съвместната вероятност се изразява като произведение на индивидуалните вероятности:

$$\mathbb{P}(X=x_i, Y=y_j) = \mathbb{P}(X=x_i)\,\mathbb{P}(Y=y_j)$$

???

$$\Rightarrow \mathbb{P}(X+Y=k) = \sum_{(i,j):\, x_i+y_j=k} \mathbb{P}(X=x_i)\,\mathbb{P}(Y=y_j)$$

$\boxed{2.}$ 1. Въведете схема на Бернули

↳ $(X_i)_{i\geq 1}$ са независими в съвкупност и еднакво разпределени дискретни случайни величини, т.е.

$$\mathbb{P}(X_1=1) = p = \mathbb{P}(X_j=1),\ j\geq 1$$
$$\mathbb{P}(X_1=0) = 1-p = \mathbb{P}(X_j=0),\ j\geq 1.$$

Тази схема се нарича схема на Бернули.

2. Чрез схемата на Бернули с параметър p въведете геометрична случайна величина с параметър p и разпишете нейната таблица.

$$X \sim Ge(p) \text{ и } X = \min\left\{\sum_{i=1}^{k} X_i = 1\right\} - 1,\ k\geq 1$$

↳ т.е. най-малкото k, за което сумата $\sum_{i=1}^{k} x_i$ става единица, като от нея вадим последната единица, за да извадим последната стъпка.

↳ С геометричното разпределение се моделира броят неуспехи до първия успех

→ таблицата на разпределение:

X	0	1	2	...	k	...
$\mathbb{P}$	p	$p \cdot q$	$p \cdot q^2$	...	$p \cdot q^k$	...

$\mathbb{P}(X = k) = p \cdot q^k$

3. Въведете $X \sim NB(r, p)$, $r \geq 1$, и пресметнете нейната пораждаща функция

↳ $X \sim NB(r, p)$ и $X = \min \{ j \geq 1 \mid \sum_{j=1} x_j = r \} - r$.

p → вероятността за успех в схемата на Бернули

↳ С отрицателно биномно се моделира брой неуспехи до r-тия успех.

$$\Rightarrow g_X(s) \overset{\text{тв.}}{=} \prod_{j=1}^{r} g_{Y_j}(s) = \prod_{j=1}^{r} \frac{p}{1-qs} = \left(\frac{p}{1-qs} \right)^r$$

↳ Тук използване твърдението, което казва:

Ако $X \sim N(r, p)$, то $X = \sum_{i=1}^{r} Y_i$, където $Y_i \sim Ge(p)$

$\Rightarrow Y_i$ са независими в съвкупност и тогава пораждащата ф-я има следното свойство:

$$Y = \sum_{i=1}^{n} x_i \Rightarrow g_Y(s) = \prod_{i=1}^{n} g_{x_i}(s)$$

→ ще понеже са еднакво разпределени, то пораждащата от $g_X(s)$ става r пъти геометричното

4. Нека $X \sim NB(r_1, p)$, $Y \sim NB(r_2, p)$, $r_1, r_2 \geq 1$ са независими случайни величини. Намерете разпределението на $X+Y$.

$X \sim NB(r_1, p)$ $\quad X \perp\!\!\!\perp Y$

$Y \sim NB(r_2, p)$ $\quad X+Y \sim$

$$g_{X+Y}(s) = g_X(s)\, g_Y(s) = \left(\frac{p}{1-qs}\right)^{r_1} \left(\frac{p}{1-qs}\right)^{r_2} =$$

$$= \left(\frac{p}{1-qs}\right)^{r_1+r_2} \Rightarrow X+Y \sim NB(r_1+r_2, p)$$

3.

1. Формулирайте централна гранична теорема и илюстрирайте нейното приложение

Нека $(X_n)_{n\geq 1}$ е редица от независими и еднакво разпределени случайни величини, дефинирани в едно вероятностно пространство. Нека $EX_1^2 < \infty$ и $EX_1 = \mu$. Нека $DX = \sigma^2$, тогава

$$Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} Z \sim N(0,1), \text{ където}$$

$$S_n = \sum_{i=1}^{n} X_i.$$

долми илюстрация ?

1. Въведете кумулативната функция на разпределение на случайна величина X и различете нейните свойства

Нека X е сл. величина във вероятностно пространство U. Тогава $F_X(x) = \mathbb{P}(X \le x)$, $\forall x \in \mathbb{R}$ се нарича функция на разпределение.

св-ва само на X. сл. вел. $\lim_{x \to \infty} F_X = 1$ $\lim_{x \to -\infty} F_X = 0$

2. Какво означава $X \stackrel{d}{=} Y$, където X,Y са случайни величини

$X \stackrel{d}{=} Y \iff g_X = g_Y$ → но ако дискретни ?

5. Нека X е произволна случайна величина.

1.) Формулирайте неравенството на Чебишев.

$$\mathbb{P}(|X - \mathbb{E}X| > \varepsilon) \leq \frac{DX}{\varepsilon^2}, \quad \forall \varepsilon > 0$$

2.) Ако X е неотрицателна и $\mathbb{E}X < \infty$, покажете, че за $a > 0$, $\mathbb{P}(X \geq a\,\mathbb{E}X) \leq \frac{1}{a}$

6. Нека $\vec{x} = (x_1, x_2, \dots, x_n)$ са наблюдения над сл. вел. X с параметризирана ~~[illegible]~~ вероятностна плътност $f(x;\theta) = C_\theta x^{\theta-1}$, $x \in (0,1)$, $\theta > 0$. Истинската стойност е $\theta_0 > 0$, но не я знаем.

1.) Намерете C_θ за $\theta > 0$

↳ плътността трябва да се интегрира до 1 в интервала $(0,1)$

$$\Rightarrow \int_0^1 f(x,\theta)\,dx = \int_0^1 C_\theta x^{\theta-1}\,dx = 1$$

$$C_\theta \int_0^1 x^{\theta-1}\,dx = C_\theta \left[\frac{x^\theta}{\theta}\right]_0^1 = C_\theta \cdot \frac{1}{\theta} = 1$$

$$\Rightarrow C_\theta = \theta$$

2. Въведете понятието максимално правдоподобна оценка и намерете максимално правдоподобната оценка за θ_0

↳ Статистиката $\hat{\theta}(\vec{x})$ се нарича оценка по метода на М.П. ако $L(\vec{x}, \hat{\theta}) = \sup_{\theta \in \Theta} L(\vec{x}, \theta)$, където Θ е множеството от всички допустими стойности за θ.

↳ Ако $f_X(\vec{x}, \theta)$ е диф. по θ, то $\hat{\theta}$ е решение на

$$\frac{\partial L}{\partial \theta} \iff \frac{\partial \ln L}{\partial \theta} = 0$$

→ правдоподобната функция

$$L(\vec{x},\theta) = \prod_{i=1}^{n} f_X(x_i,\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1}$$

$$\ln L(\vec{x},\theta) = \sum_{i=1}^{n} \ln(\theta x^{\theta-1}) = n\ln(\theta) + (\theta-1)\sum_{i=1}^{n} \ln x_i$$

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i = 0$$

$$n + \theta\sum_{i=1}^{n} \ln x_i = 0$$

$$\theta = -\frac{n}{\sum_{i=1}^{n} \ln x_i} \to \hat{\theta}$$

???

$$f_X(\vec{x},\theta) = \theta \cdot x^{\theta-1}, \quad x \in (0,1)$$

$$L(\vec{x},\theta) = \prod_{i=1}^{n} \theta \cdot x^{\theta-1}$$

$$\ln L(\vec{x},\theta) = \ln\left(\prod_{i=1}^{n} \theta \cdot x_i^{\theta-1}\right) = \sum_{i=1}^{n} \ln(\theta \cdot x_i^{\theta-1}) =$$

$$= \sum_{i=1}^{n} \ln(\theta) + \sum_{i=1}^{n} (\theta-1)\ln(x_i) =$$

$$= n\ln(\theta) + (\theta-1)\sum_{i=1}^{n} \ln(x_i)$$

$$\Rightarrow \frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln(x_i) = 0$$

$$\frac{n}{\theta} = -\sum_{i=1}^{n} \ln(x_i)$$

$$n = -\sum_{i=1}^{n} \ln(x_i) \cdot \theta$$

$$\theta = \frac{n}{-\sum_{i=1}^{n} \ln(x_i)}$$

2a

5)

→ X_n описва нивото на замърсяване в часа n

$X_n \sim Ge\left(\frac{9}{10}\right)$

→ вероятността час n да има замърсяване $\mathbb{P}(X_n \geq 3)$

$$\Rightarrow \mathbb{P}(X_n \geq 3) = \mathbb{P}(X_n > 2) = (1-p)^2 = \left(\frac{1}{10}\right)^2 = \frac{1}{100}$$

↓ т.е. замърсяването ще е поне 3

⇒ Y = „# замърсявания ~~през~~ измежду първите 1000 часа"

$Y \sim Bin\left(1000, \frac{1}{100}\right)$

т.е. $EY = n \cdot p = 1000 \cdot \frac{1}{100} = 10$

$DY = n \cdot p \cdot q = 1000 \cdot \frac{1}{100} \cdot \frac{99}{100} = \frac{99}{10}$

$$\mathbb{P}(Y \geq 5) = \mathbb{P}\left(\frac{Y-np}{\sqrt{np(1-p)}} \geq \frac{5-np}{\sqrt{np(1-p)}}\right) = \mathbb{P}\left(\frac{Y-10}{\sqrt{\frac{99}{10}}} \geq \frac{5-10}{\sqrt{\frac{99}{10}}}\right) =$$

$\sqrt{\frac{99}{10}} = \sqrt{9,9} \approx 3,15$

$$= \mathbb{P}\left(\frac{Y-10}{3,15} \geq \frac{-5}{3,15}\right) \overset{\text{Моавър-Лаплас}}{\sim} \mathbb{P}\left(Z \geq \frac{-5}{3,15}\right) =$$

$$= \mathbb{P}(Z \geq -1,59) = 1 - \mathbb{P}(Z \leq -1,59) = 1 - 0,0559 = 0,9441$$

$\approx 94,41\%$

???