СЕМ 2024

1.

1.) Дефинирайте елементите на вероятностно пространство

→ Нека Ω е мн-во от елементарни събития, $\mathcal{A}$ е σ-алгебра и $\mathbb{P}: \mathcal{A} \to [0,1]$ е вероятностна ф-я. Тогава наредената тройка $(\Omega, \mathcal{A}, \mathbb{P})$ се нарича вероятностно пространство

2.) Въведете понятието пораждаща функция

→ Нека X е целочислена, неотрицателна, дискретна сл. величина, във вероятностното пространство $V = (\Omega, \mathcal{A}, \mathbb{P})$, като $X: \Omega \to \mathbb{N} = \{0, 1, 2, \dots\}$, тогава

$$g_X(s) = \mathbb{E}\, s^X = \sum_{k=0}^{\infty} s^k \, \mathbb{P}(X = k) = \sum_{k=0}^{\infty} s^k p_k, \quad \text{за } |s| \le 1$$

се нарича пораждаща функция.

3.) Формулирайте и докажете теоремата на Поасон

→ Нека $X_n \sim Bin(n, p_n)$, $\forall n \ge 1$. Нека е в сила, че

$p_n = \frac{u_n}{n} + \frac{\lambda}{n}$, където $\lim_{n\to\infty} u_n = 0$ и $\lambda > 0$. Тогава

$$\mathbb{P}(X_n = k) \xrightarrow[n\to\infty]{} \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k \ge 0$$

$$* \quad X_n \xrightarrow[n\to\infty]{d} X \sim Poi(\lambda)$$

→ д-во: $u_n := 0 \Rightarrow p_n = \frac{\lambda}{n} \Rightarrow q_n = 1 - p_n = 1 - \frac{\lambda}{n}$

$$g_{X_n}(s) = (q_n + p_n s)^n = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n} \cdot s\right)^n = \left(1 + \frac{\lambda(s-1)}{n}\right)^n$$

$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$g_X(s) = e^{\lambda(s-1)} \qquad = \qquad \boxed{\left(1 + \frac{x}{n}\right)^n = e^x} \qquad e^{\lambda(s-1)}$$

2.

1.) Формулирайте и докажете неравенството на Чебишев

↳ $\mathbb{P}(|X-\mathbb{E}X| > \varepsilon) \le \dfrac{DX}{\varepsilon^2}$, $\forall \varepsilon > 0$

↳ доказателството следва директно от неравенството на Марков, като приложим $Y = (X-\mathbb{E}X)^2$ и $a = \varepsilon^2$

$$\Rightarrow \mathbb{P}(|X-\mathbb{E}X|^2 > \varepsilon^2) \le \frac{\mathbb{E}[|X-\mathbb{E}X|^2]}{\varepsilon^2} = \frac{DX}{\varepsilon^2}$$

$$\mathbb{P}(|X-\mathbb{E}X| > \varepsilon)$$ ↳ от свойствата (следствията) от неравенството на Марков

2) Формулирайте и докажете при крайна дисперсия слабия закон за големите числа

↳ Нека имаме редицата от еднакво разпределени и независими (i.i.d) сл. величини $(X_i)_{i \ge 1}$ с очакване съответно $\mathbb{E}X_i$. Казваме, че за X е изпълнен ЗГЧ (слаб) ако:

$$\frac{\sum_{i=1}^{n} (X_i - \mathbb{E}X_i)}{n} \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

$\rightarrow$ За $X_1, \dots, X_n$ нез. и едн. разп с $DX_1 < \infty$ е верен (слаб) ЗГЧ:

$$\frac{\sum_{i=1}^{n} (X_i - \mathbb{E}X_i)}{n} \xrightarrow[n\to\infty]{\mathbb{P}} 0 \iff \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1$$

↳ доказателство:

$$S_n = \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \mathbb{E}S_n = n \cdot \mathbb{E}X_1$$

$DS_n = n \cdot DX_1$ → понеже са независими и еднакво разпределени

искаме да докажем, че: $\frac{S_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1$

↳ Нека $\varepsilon > 0$

прилагаме Чебишев

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X_1\right| > \varepsilon\right) \le \frac{D\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\frac{1}{n^2} DS_n}{\varepsilon^2} = \frac{\frac{1}{n} DX_1}{\varepsilon^2} =$$

$$= \frac{1}{n}\left[\underbrace{\frac{DX_1}{\varepsilon^2}}_{const}\right]$$

$\Rightarrow$ при ~~...~~ $\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}X_1\right| > \varepsilon\right) \le \lim_{n\to\infty} \frac{1}{n}\left[\frac{DX}{\varepsilon^2}\right] = 0$

$$\Rightarrow \frac{S_n}{n} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}X_1 \quad //$$

$\boxed{3}$ Нека $(X_n)_{n\ge1}$ е редица от сл. вел. Нека $(e_n)_{n\ge1}$ е редица от сл. вел. с $e_n \sim Exp(\lambda_n)$, $\lambda_n > 0$, $n \ge 1$.

1) Чрез $(X_n)_{n\ge1}$ въведете $X_n \xrightarrow{d} X$, където X е някаква случайна величина

↳ казваме, че редицата X_n клони към сл. вел. X по разпределение, ако за всяка точка на непрекъснатост $F_X(x) := \mathbb{P}(X \le x)$ е вярно, че:

$$F_{X_n}(x) = \mathbb{P}(X_n \le x) \xrightarrow[n\to\infty]{d} \mathbb{P}(X \le x) = F_X(x)$$ и пишем

$$X_n \xrightarrow[n\to\infty]{d} X.$$

.) Намерете функцията на моментите на e_1 и чрез нея намерете $\mathbb{E}[e_1]$ и $\mathbb{E}[e_1^2]$

$e_n \sim Exp(n\lambda) \Rightarrow n=1 \Rightarrow e_1 \sim Exp(\lambda)$

$$M_X(t) = \mathbb{E}e^{tX} = \int_{-\infty}^{\infty} e^{tx} \cdot f_X(x)\,dx$$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{иначе} \end{cases}$$

$$\Rightarrow M_X(t) = \int_0^{\infty} e^{tx} \cdot e^{-\lambda x} \cdot \lambda\,dx = \lambda \int_0^{\infty} e^{x(t-\lambda)}\,dx =$$

$$= \lambda \cdot \frac{1}{t-\lambda}\left[e^{x(t-\lambda)}\right]_0^{\infty} = \lambda \cdot \frac{1}{t-\lambda} \cdot 0-1 = \frac{\lambda}{\lambda-t},\ t<\lambda$$

$$\frac{d^k}{dt^k} M_X(t)\Big|_{t=0} = \mathbb{E}X^k$$

$$\rightarrow \frac{d}{dt} M_X(t)\Big|_{t=0} = \frac{d}{dt}\frac{\lambda}{\lambda-t} = \frac{\lambda}{(\lambda-t)^2}\Big|_{t=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$\rightarrow \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} = \frac{d^2}{dt^2}\frac{\lambda}{\lambda-t} = \frac{d}{dt}\frac{\lambda}{(\lambda-t)^2} = \frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0} =$$

$$= \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$$

$\boxed{5.}$ 1.) Дайте дефиниция за оценка по метода на моментите

Нека X има функция на ~~моментите~~ разпределение $F_X(\vec{x}, \theta)$, $\vec{x}$ е извадка над X,

$E[X^k] = \mu^{(k)}(\theta) \to$ теоретичните моменти, $k = 1, 2, \dots$

$\overline{X}^{(k)} = \frac{1}{n} \sum_{j=1}^{n} x_j^k \to$ емпиричните моменти, $k = 1, 2, \dots$

и $\hat{\theta} = (\theta_1, \dots, \theta_s)$. Тогава решение на системата

$$\mu^{(k)}(\theta) = \overline{X}_n^{(k)}, \text{ за } k = 1, \dots, s$$

θ се нарича оценка по М.М.

2.) Нека $X \sim Poi(1440\lambda)$ и $Y \sim Poi(1440\mu)$, където λ и $\mu > 0$. С X и Y се моделира броят на катастрофите съответно в Европа и Америка. Нека $\vec{X} = (x_1, \dots, x_n)$, $\vec{Y} = (y_1, \dots, y_n)$ са независими наблюдения над X и Y ($(x_1, x_2, \dots, x_n, y_1, \dots, y_n)$ са независими в съвкупност).

Нека $\vec{Z} = \vec{X} + \vec{Y}$

а) Предложете оценка по метода на моментите за λ.

$E X = 1440\lambda$ $\qquad \overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$\Rightarrow E X \approx \overline{X}$

$1400\lambda \approx \overline{X}$

$\lambda \approx \frac{\overline{X}}{1400} \Rightarrow \lambda = \frac{1}{1400} \cdot \frac{1}{n} \sum_{i=1}^{n} x_i$

5) ако $\vec{X} = (t_1, t_2)$, що считате ли че оценката за λ от а) е добра и защо?

$$\hat{\lambda}_{MM} = \frac{1}{1400} \cdot \frac{1}{n} \sum_{i=1}^{n} x_i \text{ , когато } n = 2 \text{, оценката става:}$$

$$\hat{\lambda}_{MM} = \frac{1}{1400} \cdot \frac{1}{2} (t_1 + t_2)$$

↳ оценката, базирана на две наблюдения, не е особено добра поради следните причини:

- висока променливост и нестабилност:
  малкият брой наблюдения води до значителна случайна променливост, която може да отклони оценката далеч от истинската стойност на λ.
- несигурност и ниска точност:
  малката извадка увеличава несигурността и намалява точността на оценката.

↳ За по-надеждна оценка на λ, би било по-добре да имаме повече наблюдения.

г) Ако разполагаме със $\vec{z} = (z_1, z_2)$, то намерете оценки за λ, μ и аргументирано дискутирайте тяхното качество

$$X \sim Poi(1400\lambda), \quad Y \sim Poi(1400\mu) \Rightarrow Z = X + Y \Rightarrow Z \sim Poi(1400\lambda + 1400\mu) = \sim Poi(1400(\mu+\lambda))$$

$\hookrightarrow$ $\mathbb{E}Z = 1400(\lambda+\mu)$

$$\bar{Z} = \frac{1}{2}(z_1 + z_2)$$

MM

$\hookrightarrow$ приравняване теоретичните към емпиричните моменти

$$1400(\lambda+\mu) \approx \frac{1}{2}(z_1+z_2)$$

$$\Rightarrow (\widehat{\lambda+\mu})_{MM} = \frac{1}{2\cdot 1400}(z_1+z_2)$$

$\Rightarrow$ ако искаме да ги намерим отделно $\hat{\lambda}_{MM} = \frac{\widehat{\lambda+\mu}}{2}$?

$$\hat{\mu}_{MM} = \frac{\widehat{\lambda+\mu}}{2}$$

$\Rightarrow$ $L(\vec{z}, \lambda, \mu) = \prod_{i=1}^{2} \frac{e^{-1400(\lambda+\mu)}(1400(\lambda+\mu))^{z_i}}{z_i!}$ $\rightarrow$ ф-я на правдоподобие

$\rightarrow$ логаритмична ф-я на правдоподобност

$$\ln L(\vec{z}, \lambda, \mu) = \sum_{i=1}^{2}\left[-1400(\lambda+\mu) + z_i \ln(1400(\lambda+\mu)) - \ln(z_i!)\right] =$$

$$= -2\cdot 1400(\lambda+\mu) + (z_1+z_2)\ln(1400(\lambda+\mu)) - const$$

$\rightarrow$ $$\frac{\partial \ln L}{\partial(\mu+\lambda)} = -2\cdot 1400 + \frac{z_1+z_2}{1400(\lambda+\mu)}\cdot 1400 = 0$$ MLE

$$1400(\lambda+\mu) = \frac{z_1+z_2}{2} \Rightarrow \widehat{\lambda+\mu}_{MLE} = \frac{1}{2\cdot 1400}(z_1+z_2)$$

⇒ При малък брой наблюдения, оценките могат ненадеждни и неточни

Нека $X \sim Exp(\lambda)$, $\lambda > 0$ и имаме наблюдения $\vec{x} = (x_1, x_2, \ldots, x_n)$. В контекста на тестване на хипотези разгледайте въпросите:

1.) Дефинирайте грешка от първи и грешка от втори род

↳ Грешка от първи род: $P(\vec{x} \in \omega | H_0) = \alpha = \alpha_\omega$

⇒ нашите наблюдения $(\vec{x})$ да попаднат в ω, т.е. да отхвърлим нулевата хипотеза, при положение, че тя е вярна

↳ Грешка от втори род: $P(\vec{x} \notin \omega | H_1) = \beta = \beta_\omega$

⇒ приели сме нулевата хипотеза, но е вярна алтернативната

2.) Дефинирайте оптимална критична област при зададена грешка от първи род

↳ Казваме, че $\omega^* \in \mathbb{R}^n$ е оптимална критична област (О.К.О) за фиксирана грешка от първи род, ако $\beta_{\omega^*} = \min \beta_\omega$

$\alpha_\omega = \alpha_{\omega^*} = \alpha$

бележка

⇒ т.е. фиксирали сме грешка от първи род и търсим такава критична област, така че грешката от втори род да е минимална

3.) Формулирайте лемата на Нейман-Пирсън

Нека X, $f(x|x,\theta)$, $L(\vec{x},\theta)$ е извадка $\vec{x}$.

Ако $w^* \subseteq \mathbb{R}^n$ е к.о. с грешка от първи род α и

$$w^* \subseteq \{L_1(\vec{x}) \geq k L_0(\vec{x})\}$$
$$(w^*)^c \subseteq \{L_1(\vec{x}) \leq k L_0(\vec{x})\}$$
, за някое $k > 0$.

Тогава w^* е о.к.о. за $H_0: \theta = \theta_0$
$H_1: \theta = \theta_1$.