آمار و احتمال مهندسی

نیم‌سال دوم ۱۴۰۳-۱۴۰۲

مدرس: دکتر امیر نجفی

تمرین سری ششم | زمان تحویل: ۲۰ خرداد

- مهلت ارسال پاسخ تا ساعت ۲۳:۵۹ روز مشخص‌شده است.
- همکاری و هم‌فکری شما در انجام تمرین مانعی ندارد اما پاسخ ارسالی هر کس حتما باید توسط خود او نوشته شده باشد.
- در صورت هم‌فکری و یا استفاده از هر منابع خارج درسی، نام هم‌فکران و آدرس منابع مورد استفاده برای حل سوال مورد نظر را ذکر کنید.
- لطفا تصویری واضح از پاسخ سوالات نظری بارگذاری کنید. در غیر این صورت پاسخ شما تصحیح نخواهد شد.

بارم‌بندی

بارم سوالات به شکل زیر است: (مجموعا ۱۰۰ نمره)

- سوالات ۱، ۲ و ۳: هر کدام ۱۵ نمره
- سوال ۴: ۲۵ نمره
- سوالات ۵، ۶ و ۷: هر کدام ۱۰ نمره

مسئله‌ی ۱. (انتظار)

مدت زمانی که صادق در ایستگاه اتوبوس منتظر اتوبوس می‌ماند، از توزیع نمایی با پارامتر Θ پیروی می‌کند. می‌دانیم توزیع پیشین پارامتر Θ به صورت زیر است:

$$f_\Theta(\Theta) = \begin{cases} ۱۰\Theta & \Theta \in [۰, \frac{۱}{\sqrt{۵}}] \\ ۰ & \text{otherwise} \end{cases}$$

- با دانستن این که در روز دوشنبه، صادق ۳۰ دقیقه در ایستگاه منتظر اتوبوس مانده است، تابع چگالی احتمال توزیع پسین را به دست آورید.
- برای این حالت، مقدار پارامتر Θ را با استفاده از روش MAP تخمین بزنید.

پاسخ:

- $$f(\theta|۳۰) = \frac{f(۳۰|\theta)f(\theta)}{f(۳۰)} = \frac{\theta e^{-۳۰\theta}۱۰\theta}{C'} = \frac{\theta^۲ e^{-۳۰\theta}}{C}$$

  بطوریکه

  $$C = \int_۰^{\frac{۱}{\sqrt{۵}}} \frac{\theta e^{-۳۰\theta}۱۰\theta \, d\theta}{۱۰} = \int_۰^{\frac{۱}{\sqrt{۵}}} \theta^۲ e^{-۳۰\theta} \, d\theta.$$

- $$\hat{\theta}_{\text{MAP}} = \arg\max_\theta f(\theta|x) = \arg\max_\theta \frac{\theta^۲ e^{-۳۰\theta}}{C} = \arg\max_\theta \theta^۲ e^{-۳۰\theta}$$

  برای بدست آوردن تتای مورد نظر مشتق آن را بدست آورده و برابر با صفر قرار می‌دهیم.

  $$\frac{d(\theta^۲ e^{-۳۰\theta})}{d\theta} = ۰ \implies ۲\theta e^{-۳۰\theta} + \theta^۲(-۳۰e^{-۳۰\theta}) = ۰$$

  فرض کنید $\theta \neq ۰$:

  $$۲\theta e^{-۳۰\theta} - ۳۰\theta^۲ e^{-۳۰\theta} = ۰ \implies ۲ - ۳۰\theta = ۰ \implies \theta = \frac{۱}{۱۵}.$$

مسئله‌ی ۲. (انتگرال)

در این مسئله می‌خواهیم مقدار انتگرال یک تابع را در بازه $[a, b]$ بدست آوریم.

$$I(f) = \int_a^b f(x)\mathrm{d}x$$

برای اینکار در نظر می‌گیریم که یک تابع توزیع داریم که احتمال PDF آن را با g نشان می‌دهیم و فقط در بازه $[a, b]$ می‌توان از آن داده گرفت. حال با n داده‌گیری از این توزیع مقدار $I(f)$ را با تخمینگر زیر تخمین می‌زنیم.

$$\hat{I}_n(f) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}$$

۱. نشان دهید که $\mathbb{E}\left[\hat{I}_n(f)\right] = I(f)$

۲. نشان دهید که در تعداد داده بالا، مقدار تخمینگر ما به خود مقدار اصلی نزدیک می‌شود.

۳. مقدار $Var\left[\hat{I}_n(f)\right]$ را بدست آورید.

پاسخ:

۱. با توجه به اینکه داریم هر داده از تابع توزیع احتمال g می‌آید در نتیجه داریم که:

$$\mathbb{E}\left[\frac{f(X)}{g(X)}\right] = \int_a^b \frac{f(x)}{g(x)} g(x) \mathrm{d}x = \int_a^b f(x) \mathrm{d}x = I(f)$$
$$\Longrightarrow E\left[\hat{I}_n(f)\right] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{f(X_i)}{g(X_i)}\right] = I(f)$$

۲. با توجه به اینکه $\frac{f(X_i)}{g(X_i)}$ ها به هم وابسطه نیستند، طبق قانون اعداد بزرگ داریم که:

$$\hat{I}_n(f) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)} \longrightarrow \mathbb{E}\left[\frac{f(X)}{g(X)}\right] = I(f)$$

۳. با توجه به اینکه $X \sim g$ در نتیجه σ^2 را به صورت زیر تعریف می‌کنیم.

$$\sigma^2 := \mathrm{Var}\left[\frac{f(X)}{g(X)}\right] = \mathbb{E}\left[\left(\frac{f(X)}{g(X)}\right)^2\right] - \left(\mathbb{E}\left[\frac{f(X)}{g(X)}\right]\right)^2 = \int_a^b \frac{f(x)^2}{g(x)} \mathrm{d}x - (I(f))^2$$

برای بدست آوردن واریانس نیز به این صورت عمل می‌کنیم:

$$\mathrm{Var}\left[\hat{I}_n(f)\right] = \mathrm{Var}\left[\frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}\right] = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\left[\frac{f(X_i)}{g(X_i)}\right] = \frac{1}{n} \mathrm{Var}\left[\frac{f(X)}{g(X)}\right] = \frac{\sigma^2}{n}$$

مسئله‌ی ۳. (بسکتبال)

برای مسابقات بسکتبال دانشگاه‌ها، سیستم امتیازدهی جدیدی ابداع شده که در آن هر تیم می‌تواند در هر بازی امتیازی مثبت یا منفی کسب کند؛ این امتیاز، لزوما صحیح نیست و می‌تواند اعشاری نیز باشد. فرض کنید که امتیازهای

تیم بسکتبال شریف در مقابل یک تیم خاص از یک توزیع نرمال پیروی می‌کند؛ ولی ما اطلاعی در مورد میانگین (θ) و واریانس آن ($\sigma^۲$) نداریم. فرض کنید در طول ۱۰ مسابقه اخیر بین این دو تیم امتیازات تیم دانشگاه شریف به صورت زیر است (داده‌ها از هم مستقل هستند):

$$۵۹,\ ۶۲,\ ۵۹,\ ۷۴,\ ۷۰,\ ۶۱,\ ۶۲,\ ۶۶,\ ۶۲,\ ۷۵$$

- یک بازه با اطمینان ۹۵ درصد برای θ پیدا کنید.
- حال با فرض $\sigma^۲ = ۲۵$ بازه اطمینان ۹۵ درصدی را دوباره برای θ پیدا کنید. جواب این قسمت و قسمت قبل را با هم مقایسه کنید. به نظر شما دلیل تفاوت بین بازه‌ها چیست؟

پاسخ:

- به وضوح می‌توان دید که برای پیدا کردن بازه مورد نظر باید از روش $t - test$ استفاده کرد زیرا در اطلاعات داده شده مسئله، ما اطلاعی از $\sigma^۲$ نداریم. ابتدا $\bar{x}$ و $s^۲$ را حساب می‌کنیم و بعد با استفاده از جدول $t-value$، مقدار مورد نظر را برای توزیع t و درجه آزادی ۹ پیدا می‌کنیم.

$$\bar{x} = ۶۵ \quad s^۲ = ۳۵/۷۷۸ \quad t_{۹,۰/۰۲۵} = ۲/۲۶۲$$

$$-t \leqslant \frac{\bar{x}-\theta}{\frac{\sqrt{n}}{s}} \leqslant t \quad \longrightarrow \quad [\bar{x} - \frac{t_{۹,۰/۰۲۵}s}{\sqrt{n}}, \bar{x} + \frac{t_{۹,۰/۰۲۵}s}{\sqrt{n}}] = \\ [۶۵ - ۲/۲۶۳\sqrt{۳/۵۷۷۸}, ۶۵ + ۲/۲۶۲\sqrt{۳/۵۷۷۸}] = [۶۰/۷۲۱, ۶۹/۲۷۹]$$

۹۵ درصد اطمینان به این معنیست که در ۹۵ درصد از آزمایش‌ها، بازه‌ای که ارائه شده است حاوی θ واقعی است. این ۹۵ درصد احتمال قرار گرفتن θ در بازه داده شده نمی‌باشد.

- حال که دیگر واریانس داده توزیع مورد نظر را داریم باید از $z - test$ استفاده بکنیم. مانند قسمت قبل با استفاده از $z - table$ مقدار $z_{۰/۰۲۵}$ را پیدا می‌کنیم و در رابطه قرار می‌دهیم. دقت کنید که فرق این قسمت و دلیل تفاوت در نوع آزمون، مشخص بودن یا نبودن واریانس جامعه بود.

$$\bar{x} = ۶۵ \quad \sigma^۲ = ۲۵ \quad z_{۰/۰۲۵} = ۱/۹۶$$

$$-t \leqslant \frac{\bar{x}-\theta}{\frac{\sqrt{n}}{\sigma}} \leqslant t \quad \longrightarrow \quad [\bar{x} - \frac{z_{۰/۰۲۵}\sigma}{\sqrt{n}}, \bar{x} + \frac{z_{۰/۰۲۵}\sigma}{\sqrt{n}}] = [۶۵ - ۱/۹۶ \cdot ۵, ۶۵ + ۱/۹۶ \cdot ۵] = \\ [۶۱/۹۰۱, ۶۸/۰۹۹]$$

حال به دلیل تفاوت بین بازه‌های دو قسمت سوال می‌پردازیم. می‌توان دید که بازه قسمت دوم کوچک‌تر از بازه قسمت اول است. این موضوع به دو دلیل اتفاق افتاده است. یک، به دلیل آن که واریانس واقعی توزیع، $\sigma^۲$، از واریانس داده‌های بدست آمده، $s^۲$، کمتر است، بازه تولید شده در قسمت دوم از قسمت اول کوچکتر می‌شود. همچنین می‌دانیم که توزیع نرمال، دم (tail) محدودتری نسبت به توزیع t دارد پس بازه ۹۵ درصدی آن کوچکتر می‌شود.

مسئله‌ی ۴. (رگرسیون)

مسئله‌ی رگرسیون خطی ساده را در نظر بگیرید که در آن ورودی‌های $\{(x_i, y_i)\}_{i=1}^{n}$ داده شده است. مقادیر $\{x_i\}_{i=1}^{n}$ یقینی و مشخص هستند. اما به ازای هر x_i مقدار y_i از طریق رابطه‌ی زیر به دست می‌آید:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

که $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. اگر بدانیم که β_0، β_1 و σ نامنفی هستند، به پرسش‌های زیر پاسخ هید.

آ) اثبات کنید که تخمین بیشینه درست‌نمایی دو پارامتر β_0 و β_1 معادل انتخاب مقادیری برای β_0 و β_1 است که میانگین مربعات خطا را کمینه می‌کند.

ب) اثبات کنید که تخمین‌های بدست آمده در بخش پیشین نااُریب بوده و از توزیع‌های زیر پیروی می‌کنند:

$$\hat{\beta}_1 \sim \mathcal{N}\left(\beta_1, \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}\right), \quad \hat{\beta}_0 \sim \mathcal{N}\left(\beta_0, \frac{\sigma^2 \sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2}\right)$$

پ) بررسی کنید که آیا تخمینگر بیشینه درست‌نمایی عضوی از خانواده‌ی خطی تخمینگرهای زیر است یا نه؟ اگر هست رابطه‌ی γ_i را برحسب داده‌های ورودی بدست آورید.

$$\tilde{\beta}_1 = \frac{\sum \gamma_i y_i}{\sum \gamma_i x_i} \quad \text{such that} \quad \sum_i \gamma_i = 0$$

ت) اثبات کنید هر تخمینگری که عضو خانواده فوق است نااُریب می‌باشد.

ث) اثبات کنید به ازای هر انتخابی از مقادیر γ_i در خانواده فوق داریم $Var(\hat{\beta}_1) \leqslant Var(\tilde{\beta}_1)$. نتیجه بدست آمده را توضیح دهید.

پاسخ:

آ)

$$f_L = f(y_1, \cdots, y_n | \beta_0, \beta_1) = \prod_{i=1}^{n} p(y_i | \beta_0, \beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

$$\rightarrow f_L = (2\pi)^{\frac{-n}{2}} \sigma^{-n} \exp\left(- \frac{\sum_i (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

$$\mathcal{L} = \log(f_L) = c - \frac{1}{2\sigma^2} \sum (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial \mathcal{L}}{\partial \beta_0} \propto \sum (y_i - \beta_0 - \beta_1 x_i) = 0 \rightarrow \beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\frac{\partial \mathcal{L}}{\partial \beta_1} \propto \sum (y_i - \beta_0 - \beta_1 x_i) x_i = 0 \rightarrow \beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

ب) رابطه را برای $\hat{\beta}_1$ اثبات می‌کنیم و باقی روند برای $\hat{\beta}_0$ به همین صورت می‌باشد.

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x}) y_i}{\sum (x_i - \bar{x})^2} = \sum_i \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} y_i$$

از آنجایی که y_i خود یک توزیع گاوسی دارد، $\hat{\beta}_1$ نیز جمع تعدادی جمله‌ی گاوسی می‌شود و در نتیجه توزیع آن نیز گاوسی می‌شود. پس داریم:

$$\mathbb{E}[\hat{\beta}_1] = \sum_i \frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} (\beta_0 + \beta_1 x_i) = \beta_1$$

$$Var(\hat{\beta}_1) = \sum_i \left(\frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right)^2 Var(y_i) = \frac{\sigma^2}{\sum (x_i - \bar{x})}$$

پ) بدیهیست با جایگذاری $\gamma_i = x_i - \bar{x}$ به همیان تخمین کمینه‌ی مربعات خطا می‌رسیم.

ت)

$$\mathbb{E}[\tilde{\beta}_1] = \sum_i \frac{\gamma_i}{\sum \gamma_i x_i} \mathbb{E}[y_i] = \sum_i \frac{\gamma_i}{\sum \gamma_i x_i} (\beta_0 + \beta_1 x_i) = 0 + \frac{\sum \gamma_i x_i}{\gamma_i x_i} \beta_1 = \beta_1$$

ث)

$$\frac{\gamma_i}{\sum \gamma_i x_i} := d_i \rightarrow Var(\tilde{\beta}_1) = \sigma^2 \sum d_i^2$$

$$\sigma^2 \sum d_i^2 \sum (x_i - \bar{x})^2 \geqslant \sigma^2 \Big(\sum d_i (x_i - \bar{x}) \Big)^2 = \sigma^2$$

$$\rightarrow \sigma^2 \sum d_i^2 = Var(\tilde{\beta}_1) \geqslant \frac{\sigma^2}{\sum (x_i - \bar{x})^2} = Var(\hat{\beta}_1)$$

در این خانواده از تخمینگرها همه نااُریب هستند و تخمین بیشینه درست‌نمایی در میان آن‌ها از کمترین واریانس برخوردار است پس می‌توان ادعا کرد که این تخمین بهترین است.

مسئله‌ی ۵. (تخمین)

اگر $X_1, X_2, \cdots, X_n$ یک نمونه از توزیع نمایی با پارامتر λ باشند و این پارامتر خود از یک توزیع با تابع چگالی احتمال $f_X(x) = \beta^\alpha x^{\alpha-1} e^{-\beta x}$ پیروی کند، تخمینگر MAP را برای λ با توجه به n وقوع مستقل X_i محاسبه کنید.

(راهنمایی: پاسخ شما باید بر مبنای پارامتر های β, α, n و X_i باشد.)

پاسخ: طبق قانون بیز برای تخمینگر MAP داریم

$$\hat{\lambda}_{\text{MAP}} = \arg\max \left(\frac{f(\lambda) \prod_{i=1}^{n} f(X_i|\lambda)}{f(X)} \right)$$

حال طبق فرضیات سوال داریم:

$$\prod_{i=۱}^{n} f(X_i|\lambda) = \prod_{i=۱}^{n} \lambda e^{-\lambda X_i} = \lambda^n e^{-\lambda \sum_{i=۱}^{n} X_i}$$

$$f(\lambda) = \beta^\alpha \lambda^{\alpha-۱} e^{-\beta\lambda}$$

$$f(X) = \int_{-\infty}^{\infty} f(X|\lambda) f(\lambda) d\lambda = C(X)$$

همان طور که مشاهده میکنید مقدار $f(x)$ مستقل از است. حال با جایگذاری داریم:

$$\hat{\lambda}_{\text{MAP}} = \arg\max \frac{\lambda^{\alpha+n-۱} e^{-(\beta+\sum_{i=۱}^{n} X_i)\lambda} \beta^\alpha}{C(X)}$$

با لگاریتم گیری و حذف عبارات مستقل از λ داریم:

$$\hat{\lambda}_{\text{MAP}} = \arg\max \left((n+\alpha-۱)\ln\lambda - (\beta + \sum_{i=۱}^{n} X_i)\lambda \right)$$

با مشتق‌گیری داریم:

$$\frac{n+\alpha-۱}{\lambda} - (\beta + \sum_{i=۱}^{n} X_i) = ۰ \Rightarrow \hat{\lambda}_{\text{MAP}} = \frac{n+\alpha-۱}{\beta + \sum_{i=۱}^{n} X_i}$$

مسئله‌ی ۶. (هندسی)

متغیر تصادفی پیوسته X با تابع چگالی احتمالاتی زیر را در نظر بگیرید:

$$f_X(x) = \begin{cases} ۳x^۲ & x \in [۰, ۱] \\ ۰ & o.w. \end{cases}$$

اگر $Y|X = x \sim \text{Geometric}(x)$ باشد، تخمین MAP متغیر تصادفی X به شرط $Y = ۳$ را بدست آورید.

پاسخ:
با توجه به تعریف توزیع هندسی برای $y = ۱, ۲, \cdots$ داریم:

$$P_{Y|X}(y|x) = x(۱-x)^{y-۱}$$

پس با جایگزینی $y = ۳$ خواهیم داشت:

$$P_{Y|X}(۳|x) = x(۱-x)^۲$$

اکنون باید به دنبال $x \in [۰, ۱]$ بگردیم که مقدار زیر را بیشینه کند:

$$P_{Y|X}(y|x) f_X(x) = x(۱-x)^۲ \cdot ۳x^۲ = ۳x^۳(۱-x)^۲$$

برای یافتن بیشینه این تابع از آن برحسب x مشتق گرفته و حاصل را برابر با صفر قرار می‌دهیم.

$$\frac{d}{dx}\left[۳x^۳(۱-x)^۲\right] = ۹x^۲(۱-x)^۲ - ۶x^۳(۱-x) = ۰ \rightarrow \begin{cases} x = ۰ \\ x = \frac{۳}{۵} \quad \max \\ x = ۱ \end{cases}$$

با بررسی مقادیر فوق تخمینگر MAP را بدست می‌آوریم و داریم:

$$\hat{x}_{MAP} = \frac{۳}{۵}$$

مسئله‌ی ۷. (رگرسیون تک‌متغیره)

فرض کنید یک مدل داریم که برای ورودی x_i خروجی y_i را با استفاده از رابطه‌ی زیر تولید می‌کند.

$$y_i = ax_i + \epsilon_i$$

به صورتی که ϵ_i در واقع نویز سیستم است و از توزیع $\epsilon_i \sim N(۰, \sigma^۲)$ با پارامتر ثابت σ تبعیت می‌کند. با فرض داشتن n داده به صورت $\{(x_۱, y_۱), (x_۲, y_۲), ..., (x_n, y_n)\}$، با استفاده از MLE، a را تخمین بزنید.

پاسخ:
با توجه به اینکه ارور از توزیع نرمال پیروی می‌کند داریم که

$$\epsilon_i \sim \frac{۱}{\sqrt{۲\pi\sigma^۲}} \exp\left(-\frac{\epsilon_i^۲}{۲\sigma^۲}\right)$$

Likelihood برای یک مشاهده (x_i, y_i) برابر است با:

$$L(a, \sigma^۲ \mid x_i, y_i) = \frac{۱}{\sqrt{۲\pi\sigma^۲}} \exp\left(-\frac{(y_i - ax_i)^۲}{۲\sigma^۲}\right)$$

با توجه به اینکه هر مشاهده مستقل از دیگری است، در نتیجه برای Likelihood داریم:

$$L(a, \sigma^۲ \mid \mathbf{x}, \mathbf{y}) = \prod_{i=۱}^{n} \frac{۱}{\sqrt{۲\pi\sigma^۲}} \exp\left(-\frac{(y_i - ax_i)^۲}{۲\sigma^۲}\right)$$

با گرفتن لگاریتم از تابع بالا به Log Likelihood می‌رسیم که برابر است

$$l(a, \sigma^۲ \mid \mathbf{x}, \mathbf{y}) = -\frac{n}{۲} \log\left(۲\pi\sigma^۲\right) - \frac{۱}{۲\sigma^۲} \sum_{i=۱}^{n} (y_i - ax_i)^۲$$

با گرفتن مشتق از تابع بالا مینیمم آن را بدست می‌اوریم تا $\hat{a}$ را بدست آوریم.

$$\frac{\partial}{\partial a}\left(l(a, \sigma^۲ \mid \mathbf{x}, \mathbf{y})\right) = \frac{۱}{\sigma^۲} \sum_{i=۱}^{n} x_i\left(y_i - ax_i\right) \Longrightarrow \hat{a} = \frac{\sum_{i=۱}^{n} x_i y_i}{\sum_{i=۱}^{n} x_i^۲}$$