دانشکده‌ی مهندسی کامپیوتر

آمار و احتمال مهندسی

پاسخ‌نامه تمرین سری پنجم

مدرس: دکتر مطهری

سوال ۱

الف

برای حساب کردن ضریب نرمال‌سازی کافی است مقدار انتگرال زیر را حساب کنیم:

$$\int_{-\infty}^{\infty} p(\theta|\alpha,\tau)d\theta = \int_{-\infty}^{\tau} p(\theta|\alpha,\tau)d\theta + \int_{\tau}^{\infty} p(\theta|\alpha,\tau)d\theta = \int_{\tau}^{\infty} \theta^{-(\alpha+1)}d\theta = \frac{1}{\alpha}\tau^{-\alpha}$$

ب

اگر $q(x|\theta) \sim \theta^{-x}$ ، در این صورت $p(\theta|\alpha+x,\tau) = Prob(\theta|x) \sim q(x|\tau)p(\theta|\alpha,\tau)$ برقرار است. دقت کنید که تابع q یکتا نیست و می‌توانیم آن را با هر تابعی به فرم $q(x|\theta) \sim \theta^{f(x)}$ جایگزین کنیم.

سوال ۲

$$\begin{aligned} Prob(X_1, X_2, \ldots, X_n|\theta) &= \prod_{i=1}^{n} Prob(X_i|\theta) \\ &= \prod_{i=1}^{n}(\frac{1}{\theta})\mathbb{I}(0 \leq X_i \leq \theta) = (\frac{1}{\theta})^n \prod_{i=1}^{n}\mathbb{I}(0 \leq X_i \leq \theta) \end{aligned}$$

الف

با توجه به فرمول درستنمایی که در بالا نوشته شده است، $\hat{\theta}_{ML} = \max(X_1, X_2, \ldots, X_n)$ است.

ب

کافی است $\mathbb{E}[\hat{\theta}_{ML}|\theta] = \mathbb{E}[max(X_1, X_2, \ldots, X_n)|\theta]$ را حساب کنیم. برای این ابتدا توزیع تجمعی $\hat{\theta}_{ML}$ را حساب می‌کنیم. به سادگی می‌توان دید:

$$Prob(\hat{\theta}_{ML} \leq \tau) = \begin{cases} 0, & \tau < 0, \\ (\frac{\tau}{\theta})^n, & 0 \leq \tau \leq \theta, \\ 1, & \tau > \theta \end{cases}$$

حال برای حساب کردن تابع چگالی، کافی است از توزیع تجمعی مشتق بگیریم:

$$p(\hat{\theta}_{ML} = \tau) = \begin{cases} 0, & \tau < 0, \\ \frac{n}{\tau}(\frac{\tau}{\theta})^n, & 0 \leq \tau \leq \theta, \\ 0, & \tau > \theta \end{cases}$$

حال داریم:

$$\begin{aligned} &\mathbb{E}[\hat{\theta}_{ML}|\theta] = \int_0^\theta \tau p(\hat{\theta}_{ML} = \tau) d\tau \\ &= \int_0^\theta \tau \frac{n}{\tau} (\frac{\tau}{\theta})^n d\tau = \int_0^\theta n (\frac{\tau}{\theta})^n d\tau \\ &= \frac{n}{n+1} \theta \end{aligned}$$

با توجه به این‌که $\hat{\theta}_{UB} = \alpha \hat{\theta}_{ML}$، باید $\alpha = \frac{n+1}{n}$ باشد.

ج

با توجه به این‌که در قسمت ب، تابع چگالی $\hat{\theta}_{ML}$ را حساب کردیم، به رابطه‌ی زیر برسیم:

$$Var(\hat{\theta}_{UB}) = \alpha^2 Var(\hat{\theta}_{ML}) = (\frac{n+1}{n})^2 (\frac{n}{n+2} \theta^2 - (\frac{n}{n+1} \theta)^2) = \frac{\theta^2}{n(n+2)}$$

مقدار Fisher Information به صورت زیر محاسبه می‌شود:

$$\begin{aligned} &I(\theta) = \mathbb{E}[(\frac{\log Prob(X_1, X_2, \ldots, X_n|\theta)}{d\theta})^2] \\ &= \mathbb{E}[(\frac{-n \log \theta}{d\theta})^2] = (\frac{n}{\theta})^2 \end{aligned}$$

بنابراین با توجه به کران کرامر-راو، مقدار واریانس تمام تخمین‌گرهای ناااریب θ حداقل $\frac{1}{I(\theta)} = (\frac{\theta}{n})^2$ است، اما داریم:

$$Var(\hat{\theta}_{UB}) = \frac{\theta^2}{n(n+2)} < \frac{1}{I(\theta)} = (\frac{\theta}{n})^2$$

علت این‌که کران کرامر–راو برقرار نیست این است که تابع توزیع و تخمین‌گر پیشنهادی در شرایط مورد نیاز این کران صدق نمی‌کنند. برای برقرار کرامر–راو باید دو شرط زیر برقرار باشند (برای توضیحات بیشتر می‌توانید به صفحه‌ی ویکیپدیا [1] مراجعه کنید):

۱. برای هر x که $p(x|\theta) > 0$ است، مقدار $\frac{d}{d\theta}\log p(x|\theta)$ تعریف شده و محدود باشد.

۲. تساوی زیر برقرار باشد که معادل است با این‌که بتوانیم مشتق را به داخل انتگرال ببریم:

$$\begin{aligned}
&\frac{d}{d\theta}[\int \hat{\theta}_{UB}(X_1, X_2, \ldots, X_n)p(X_1, X_2, \ldots, X_n|\theta)dX] \\
&= \int \hat{\theta}_{UB}(X_1, X_2, \ldots, X_n)\frac{dp(X_1, X_2, \ldots, X_n|\theta)}{d\theta}dX
\end{aligned}$$

دقت کنید که سمت چپ تساوی برای تمام تخمین‌گرهای ناااریب θ برابر ۱ است اما می‌توانیم کرامر–راو را برای هر تابعی از θ هم به کار ببریم که در این صورت سمت چپ تساوی می‌تواند ۱ نباشد.

می‌توانید بررسی کنید که برای $\hat{\theta}_{UB}$ شرط اول برقرار است ولی شرط دوم برقرار نیست.

سوال ۳

می‌دانیم که اگر $X \sim Poisson(\lambda)$، در این صورت:

$$\begin{aligned}
\mathbb{E}[X] &= \lambda \\
\mathbb{E}[X^2] &= \lambda^2 + \lambda
\end{aligned}$$

بنابراین باید مقدار $\hat{\lambda}_1$ را طوری انتخاب کنیم که اگر $Y \sim \text{Poisson}(\hat{\lambda}_1)$ در این صورت $\mathbb{E}[\text{Poisson}(\hat{\lambda}_1)] = \frac{\sum_{i=1}^n X_i}{n}$، پس $\hat{\lambda}_1 = \frac{\sum_{i=1}^n X_i}{n}$ است.

برای $\hat{\lambda}_2$، اگر $Y \sim \text{Poisson}(\hat{\lambda}_2)$ در این صورت $\mathbb{E}[Y^2] = \frac{\sum_{i=1}^n X_i^2}{n}$ باشد، پس:

$$\begin{aligned}
&\hat{\lambda}_2^2 + \hat{\lambda}_2 - \frac{\sum_{i=1}^n X_i^2}{n} = 0 \Rightarrow \hat{\lambda}_2 = \frac{-1 \pm \sqrt{1 + 4\frac{\sum_{i=1}^n X_i^2}{n}}}{2} \\
&\xRightarrow{\hat{\lambda}_2 > 0} \hat{\lambda}_2 = \frac{-1 + \sqrt{1 + 4\frac{\sum_{i=1}^n X_i^2}{n}}}{2}
\end{aligned}$$

[1] https://en.wikipedia.org/wiki/Cramér-Rao_bound

سوال ۴

الف

می‌توانیم به جای این‌که مقدار تابع درستنمایی را بیشینه کنیم، مقدار لگاریتم تابع درستنمایی را بیشینه کنیم (زیرا لگاریتم یک تابع اکیدا صعودی است). پس:

$$\begin{aligned}
&\hat{w}_{ML} = \arg\max_{w} \log Prob(Y_1, Y_2, \ldots, Y_n | X_1, X_2, \ldots, X_n, w_0, w_1) \\
&= \sum_{i=1}^{n} \log \mathcal{N}(Y_i | w_1 X_i + w_0, \sigma^2) \\
&\propto \sum_{i=1}^{n} \frac{-(Y_i - (w_1 X_i + w_0))^2}{2\sigma^2} \\
&\propto \sum_{i=1}^{n} -(Y_i - (w_1 X_i + w_0))^2
\end{aligned}$$

بنابراین $\hat{w}_{ML} = \hat{w}_{LMS}$ است.

ب

مشابه قسمت قبل، داریم:

$$\begin{aligned}
&\hat{w}_{MAP} = \arg\max_{w} \log Prob(w_0, w_1 | X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n) \\
&= \log Prob(Y_1, Y_2, \ldots, Y_n | X_1, X_2, \ldots, X_n, w_0, w_1) + \log Prob(w_0, w_1) \\
&= \sum_{i=1}^{n} \log \mathcal{N}(Y_i | w_1 X_i + w_0, \sigma^2) + \log \mathcal{N}(w_0 | 0, 1) + \log \mathcal{N}(w_1 | 0, 1) \\
&\propto \sum_{i=1}^{n} \frac{-(Y_i - (w_1 X_i + w_0))^2}{2\sigma^2} - \frac{w_0^2 + w_1^2}{2} \\
&\propto \sum_{i=1}^{n} -(Y_i - (w_1 X_i + w_0))^2 - \frac{w_0^2 + w_1^2}{\sigma^2}
\end{aligned}$$

برای این‌که $\hat{w}_{MAP} = \hat{w}_{RLMS}$ باشد، کافی است که $\lambda = \frac{1}{\sigma^2}$ باشد.

سوال ۵

الف

$$
\begin{aligned}
&P(\sigma^2|X_1, X_2, \ldots, X_n) \propto P(X_1, X_2, \ldots, X_n|\sigma^2)P(\sigma^2)\\
&= \prod_{i=1}^{n} \mathcal{N}(X_i|0, \sigma^2)\text{InverseGamma}(\sigma^2|2, 1)\\
&\propto (\frac{1}{\sqrt{\sigma^2}})^n exp\{-\frac{\sum_{i=1}^{n} X_i^2}{2\sigma^2}\}(\sigma^2)^{-3} exp\{-\frac{1}{\sigma^2}\}\\
&\propto (\sigma^2)^{-\frac{n}{2}-3} exp\{-\frac{\frac{\sum_{i=1}^{n} X_i^2}{2}+1}{\sigma^2}\}\\
&\Rightarrow P(\sigma^2|X_1, X_2, \ldots, X_n) = \text{InverseGamma}(\frac{n}{2}+2, \frac{\sum_{i=1}^{n} X_i^2}{2}+1)
\end{aligned}
$$

ب

با توجه به این‌که تابع فاصله، مجذور فاصله اقلیدسی است، تخمین‌گر بیز برابر میانگین توزیع پسین است (در واقع باید $\hat{\theta}$ را پیدا کنیم که عبارت $\mathbb{E}[(\theta - \hat{\theta})^2]$ کمینه شود. در صورتی که مشتق این عبارت بر حسب $\hat{\theta}$ را مساوی صفر قرار دهید، مشخص می‌شود که $\hat{\theta} = \mathbb{E}[X]$ است) و با توجه به این‌که میانگین توزیع $\text{InverseGamma}(\alpha, \beta) = \frac{\beta}{\alpha-1}$، بنابراین $\hat{\theta}_{Bayes} = \frac{\frac{\sum_{i=1}^{n} X_i^2}{2}+1}{\frac{n}{2}+1}$ است.

ج

$$
\begin{aligned}
&P(X_{n+1}|X_1, X_2, \ldots, X_n)\\
&= \int_0^\infty P(X_{n+1}|\sigma^2)P(\sigma^2|X_1, X_2, \ldots, X_n)d\sigma^2\\
&= \int_0^\infty \mathcal{N}(X_{n+1}|0, \sigma^2)\text{InverseGamma}(\frac{n}{2}+2, \frac{\sum_{i=1}^{n} X_i^2}{2}+1)d\sigma^2
\end{aligned}
$$