دانشکده‌ی مهندسی کامپیوتر

آمار و احتمال مهندسی | بهار ۱۳۹۹

پاسخ‌نامه تمرین سری هشتم (تشخیص و تخمین)

مدرس: نعیمه امیدوار | موعد تحویل: سه‌شنبه ۳ تیر ۱۳۹۹

سؤال ۱ یک معیار مهمی که می‌تواند برای سنجیدن عملکرد تخمین به‌کار گرفته شود مقدار bias آن تخمین است. این عبارت را به‌صورت زیر تعریف می‌کنیم:

$$bias(\widehat{\theta}) = E[\widehat{\theta}] - \theta$$

($\widehat{\theta}$ تخمینگری برای پارامتر θ است.)
فرض کنید n متغیر تصادفی $x_1, x_2, ..., x_n$ داریم اگر امیدریاضی آن‌ها μ باشد برای واریانس آن از رابطه‌ی زیر استفاده می‌کنیم:

$$\sigma^2 = E[(X-\mu)^2]$$

دو رابطه‌ی زیر برای تخمین واریانس واریانس این n داده پیشنهاد می شود. این دو تخمین را بر اساس بایاس مقایسه کنید و بگویید کدام یک تخمینگر بهتری است.

$$S_1^2 = \frac{1}{n}\sum_1^n (x_i - \bar{x})^2$$

$$S_2^2 = \frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2$$

پاسخ ۱

$$E(\bar{X}^2) = E(\bar{X})^2 + Var(\bar{X}) = \frac{\sigma^2}{n} + \mu^2$$

$$E(S_1^2) = \frac{1}{n}(\sum_i^n E(x_i^2) + nE(\bar{x}^2) - 2nE(\bar{x}^2)) = \frac{1}{n}(\sum_i^n E(x_i^2) - nE(\bar{x}^2))$$

$$\rightarrow E(S_1^2) = \frac{1}{n}(n(\mu^2+\sigma^2) - n(\mu^2 + \frac{\sigma^2}{n})) = \frac{n-1}{n}\sigma^2$$

$$\Longrightarrow Bias(S_1^2) = E(S_1^2) - \sigma^2 = -\frac{\sigma^2}{n}$$

$$E(S_2^2) = E(\frac{1}{n-1}\sum_1^n (x_i - \bar{x})^2) = \frac{n-1}{n-1}\sigma^2 = \sigma^2$$

$$\Longrightarrow Bias(S_2^2) = \sigma^2 - \sigma^2 = 0$$

بنابراین تخمین دوم تخمین بهتری است.

سؤال ۲ فیزیکدانی ماهر یک آشکارساز ذرات ساخته‌است که از آن برای محاسبه‌ی کمیت‌های فیزیکی سیارات و ستاره‌ها استفاده می‌کند. فرض کنید در طی یک شبانه روز تعداد ذراتی که به این آشکارساز برخورد می‌کنند (Y) از یک توزیع احتمالی شرطی پواسونی نسبت به پارامتر x پیروی کنند. پارامتر x نامعلوم است و به صورت مقداری از متغیر تصادفی X که از توزیع نمایی با پارامتر μ تبعیت می‌کند مدلسازی می‌شود. مقدار μ مشخص است. بنابراین داریم:

$$f_X(x) = \begin{cases} \mu e^{-\mu x} & x \geq 0 \\ 0 & \text{O.W} \end{cases} \tag{۱}$$

$$P_{Y|X}(y|x) = \begin{cases} \frac{e^{-x}x^y}{y!} & y = 0, 1, 2, ... \\ 0 & \text{O.W} \end{cases} \tag{۲}$$

الف. با استفاده از تخمینگر MAP متغیر x را با استفاده از تعداد ذرات آشکارشده‌ی y تخمین بزنید.
ب. در این قسمت می‌خواهیم امیدریاضی شرطی متغیر تصادفی X را نسبت به تعداد ذرات آشکار شده‌ی y به‌دست آوریم. به عبارت بهتر با استفاده از تخمینگر MMSE این متغیر را تخمین بزنیم.

۱. نشان دهید توزیع احتمال پسین X نسبت به Y از رابطه‌ای به فرم زیر تبعیت می‌کند و پارامتر λ را نیز نسبت به پارامترهای مساله به‌دست آورید.

$$f_{X|Y}(x|y) = \frac{\lambda^{y+1}}{y!}x^y e^{-\lambda x}$$

$$Tip: \int_0^\infty \alpha^{y+1}x^y e^{-\alpha x}dx = y! \quad for\ any\ \alpha > 0$$

۲. با استفاده از نتیجه‌ی قسمت قبل امیدریاضی شرطی متغیر تصادفی X را نسبت به تعداد y ذرات مشاهده شده به‌دست آورید.

پاسخ ۲ الف.

$$f_{X|Y}(x|y) = \frac{P_{Y|X}(y|x)f_X(x)}{P_Y(y)} = \frac{\mu}{y!P_Y(y)}e^{-(\mu+1)x}x^y$$

$$\frac{d}{dx}f_{X|Y}(x|y) = \frac{d}{dx}(\frac{\mu}{y!P_Y(y)}e^{-(\mu+1)x}x^y)$$

$$= \frac{\mu}{y!P_Y(y)}x^{y-1}e^{-(1+\mu)x}(y - x(1+\mu))$$

$$MAP: \frac{d}{dx}f_{X|Y}(x|y) = 0 \Longrightarrow (y - x(1+\mu)) = 0$$

$$\Longrightarrow \widehat{x}_{MAP}(y) = \frac{y}{1+\mu}$$

ب. ابتدا توزیع $P_Y(y)$ را به‌دست می‌آوریم.

$$P_Y(y) = \int_0^\infty P_{Y|X}(y|x)f_X(x)dx = \int_0^\infty \frac{e^{-x}x^y}{y!}\mu e^{-\mu x}dx$$

$$P_Y(y) = \frac{\mu}{y!(1+\mu)^{y+1}}\int_0^\infty (1+\mu)^{y+1}x^y e^{-(1+\mu)x}dx$$

$$\longrightarrow P_Y(y) = \frac{\mu}{(1+\mu)^{y+1}}$$

قسمت اول:

$$f_{X|Y}(x|y) = \frac{\mu}{y!P_Y(y)}e^{-(\mu+1)x}x^y$$

$$\rightarrow f_{X|Y}(x|y) = \frac{\mu}{y!}\frac{(1+\mu)^{y+1}}{\mu}x^y e^{-(1+\mu)x} = \frac{(1+\mu)^{y+1}}{y!}x^y e^{-(1+\mu)x}$$

$$\Longrightarrow \lambda = \mu + 1$$

قسمت دوم:

$$\widehat{x} = E[X|Y=y] = \int_0^\infty x f_{X|Y}(x|y)dx$$

$$x f_{X|Y}(x|y) = \frac{(1+\mu)^{y+1}}{y!} x^{y+1} e^{-(1+\mu)x}$$

$$= \frac{y+1}{1+\mu} \frac{(1+\mu)^{y+2}}{(y+1)!} x^{y+1} e^{-(1+\mu)x} = \frac{y+1}{\mu+1} f_{X|Y}(x|y+1)$$

$$\to x f_{X|Y}(x|y) = \frac{y+1}{\mu+1} f_{X|Y}(x|y+1)$$

$$\Longrightarrow \widehat{x} = E[X|Y=y] = \int_0^\infty x f_{X|Y}(x|y)dx = \int_0^\infty \frac{y+1}{\mu+1} f_{X|Y}(x|y+1)dx = \frac{y+1}{\mu+1}$$

سؤال ۳ نمونه‌های $X_1, X_2, X_3, ...$ را درنظر بگیرید که از یک توزیع برنولی پیروی می‌کنند. احتمال موفقیت در این توزیع را q درنظر بگیرید که مقداری نامعلوم است. متغیر زمانی T_K را به صورت زیر تعریف می‌کنیم:

$$T_1 = Y_1 \quad , \quad T_k = Y_k - Y_{k-1} \quad k = 2, 3, ...$$

که متغیر Y_k زمان kمین موفقیت در این فرایند است. در این مساله سعی می‌کنیم q را با استفاده از نمونه‌های زمانی $t_1, t_2, t_3, ...$ تخمین بزنیم. توزیع پیشین (prior distribution) متغیر q را متغیرتصادفی Q درنظر بگیرید که از توزیع یکنواخت $[0, 1]$ پیروی می‌کند.
الف. تابع جرم احتمال (PMF) متغیر T_1 یعنی $P_{T_1}(t_1)$ را به‌دست آورید.
ب. تخمین MMSE متغیر q را با استفاده از اولین داده یعنی $T_1 = t$ به‌دست آورید.
تخمین MAP متغیر q را با استفاده از k داده‌ی اول یعنی $T_1 = t_1, T_2 = t_2, ..., T_k = t_k$ به‌دست آورید.
راهنمایی : از انتگرال زیر می‌توانید استفاده کنید.

$$\int_0^1 q^k (1-q)^m dq = \frac{k!m!}{(k+m+1)!}$$

پاسخ ۳ الف.

$$P_{T_1}(t_1) = \int_0^1 P_{T_1|Q}(t_1|q) f_Q(q) dq = \int_0^1 (1-q)^{t-1} q dq = \frac{1}{t(t+1)} \quad for\ t = 1, 2, 3, ...$$

برای این قسمت از این نکته که متغیر زمانی اولین موفقیت توزیع برنولی از توزیع هندسی پیروی می‌کند استفاده کردیم.
ب.

$$\widehat{q} = E[Q|T_1 = t] = \int_0^1 P_{Q|T_1}(q|t) q dq = \int_0^1 \frac{P_{T_1|Q}(t|q) f_Q(q)}{P_{T_1}(t)} q dq$$

$$= \int_0^1 t(t+1) q (1-q)^{t-1} q dq = \int_0^1 t(t+1) q^2 (1-q)^{t-1} dq = t(t+1) \frac{2(t-1)!}{(t+2)!} = \frac{2}{t+2}$$

پ.

$$f_{Q|T_1,T_2,...,T_n}(q|t_1, t_2, ..., t_n) = \frac{f_Q(q) \Pi_i^k P_{T_i}(T_i = t_i\, Q = q)}{\int_0^1 f_Q(q) \Pi_i^k P_{T_i}(T_i = t_i\, Q = q) dq}$$

$$\rightarrow f_{Q|T_1,T_2,...,T_n}(q|t_1, t_2, ..., t_n) = \frac{q^k (1-q)^{\sum_i^k t_i - k}}{c}$$

$$MAP : \frac{d}{dq} f_{Q|T_1,T_2,...,T_n}(q|t_1, t_2, ..., t_n) = 0 \rightarrow k q^{k-1} (1-q)^{\sum_i^k t_i - k} - (\sum_i^k t_i - k) q^k (1-q)^{(\sum_i^k t_i - k) - 1} = 0$$

$$\rightarrow k(1-q) - (\sum_i^k t_i - k) q = 0 \Longrightarrow \widehat{q} = \frac{k}{\sum_{i=1}^k t_i}$$

سؤال ۴ فرض کنید توزیع توام دو متغیر تصادفی X و Y از رابطه‌ی زیر تبعیت کند:

$$f_{XY}(x,y) = \begin{cases} cxy & if\ 0 < x \leq 1\ ,\ 0 < y \leq 1 \\ 0 & \text{O.W} \end{cases} \tag{۳}$$

الف. مقدار ضریب نرمالسازی c را به‌دست آورید.
ب. تخمین MMSE متغیر تصادفی X را براساس داده‌ی مشاهده شده‌ی $Y = y$ به‌دست آورید.
پ. آیا تخمینی که در قسمت قبل به‌دست آوردید با حالتی که هیچ اطلاعی از داده‌ی $Y = y$ داشتیم متفاوت می‌بود یا نه؟ توضیح دهید.
ت. دو قسمت قبل را این بار برای تخمینگر MAP متغیر X انجام دهید.

پاسخ ۴ الف.

$$\int_0^1 \int_0^1 f_{XY}(x,y)dydx = 1 \rightarrow c\int_0^1 \int_0^1 xydydx = c\int_0^1 \frac{1}{2}xdx = \frac{c}{4}$$
$$\implies c = 4$$

ب.

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{4xy}{\int_0^1 4xydx} = \frac{4xy}{2y} = 2x\ , \quad x \in (0,1]$$

نتیجه‌ی مهمی که به‌دست آمد این است که توزیع شرطی X نسبت به Y تابعیتی از $Y = y$ ندارد و بنابراین دو متغیر تصادفی X و Y از هم مستقل هستند!

$$\implies \widehat{x}_{MSE} = E[X|Y=y] = E[X] = \int_0^1 x2xdx = \frac{2}{3}$$

پ. همانطور که در قسمت قبل به‌دست آوردیم دو متغیر تصادفی X و Y از هم مستقل هستند بنابراین داشتن اطلاعاتی از متغیر تصادفی Y تاثیری روی مقدار و درنتیجه تخمین متغیر تصادفی X نخواهد داشت!
ت. از آنجا که دو متغیر تصادفی مستقل هستند بنابراین هیچ تخمینی نمی‌تواند از مقدار مشاهده شده‌ی $Y = y$ برای تخمین متغیر تصادفی X استفاده کند که این موضوع شامل تخمینگر MAP نیز می‌شود. بنابراین تخمینگر MAP متغیر تصادفی X برابر ۱ است زیرا مستقل از هرمقداری که y اتخاذ کند احتمال شرطی یا احتمال مارجینال بیشینه خواهد شد.

سؤال ۵ در این سوال قصد داریم مساله‌ی رگرسیون خطی را با رویکرد Bayesian بررسی کنیم. فرض کنید n جفت داده‌ی (x_1, y_1) و (x_2, y_2) و ... (x_n, y_n) داریم. ساده ترین مدلی که به ذهنمان می‌رسد این است که یک رابطه‌ی خطی میان این داده‌ها برقرار است. یعنی Y یک تابع خطی از X است. به عبارت دیگر:

$$y_i = ax_i + b$$

در این مساله قصد داریم با رویکرد بیزین دو پارامتر a و b را تخمین بزنیم. در این صورت برای یک داده‌ی x' که قبلا مشاهده نشده می‌توانیم y متناظر را تخمین بزنیم. برای سادگی فرض کنید که x_i و y_i یک بعدی هستند.
معادله‌ی $y_i = ax_i + b$ را به شکل دیگری نیز می‌توانیم بنویسیم:

$$y_i = w^T X_i \quad , \quad w = \begin{bmatrix} a \\ b \end{bmatrix} \quad , \quad X_i = \begin{bmatrix} x_i \\ 1 \end{bmatrix}$$

در واقع یک سطر به x اضافه کردیم تا این رابطه را به‌صورت ضرب ماتریسی بنویسیم. سپس برای سادگی فرض کنید

$$P(y \mid X, w) \sim N(w^T X, \sigma^2)$$

و فرض کنید مقدار σ را می‌دانیم. هدف نهایی این است که توزیع پسین posterior distribution ماتریس w را حساب کنیم. به عنوان توزیع پیشین برای این متغیر فرض کنید

$$P(w) \sim N(\mu_0 = 0, \sigma_0^2) \propto exp(-\frac{w^2}{2\sigma_0^2})$$

الف. حال با استفاده از نکات بالا رابطه‌ی زیر را برای توزیع پسین w به‌دست آورید:

$$log(P(w \mid y, X)) \propto -\frac{1}{2\sigma^2} \sum (y_i - w^T x_i)^2 - \frac{1}{2\sigma_0^2} w^2$$

ب. حال می‌خواهیم رابطه‌ی بالا نسبت به پارامتر w یعنی a و b بیشینه شود. بنابراین با مشتق‌گیری نسبت به این پارامترها a و b را برحسب رابطه‌ای از x_iها و y_iها تخمین بزنید.

بخش عملی سوال:

یک فایل data.csv در کنار این فایل قرار گرفته‌است که شامل دو ستون X و Y است. با استفاده از زبان R یا Python یک تابع بنویسید که با دریافت این فایل به عنوان ورودی
الف. دو مقدار a و b را محاسبه کرده و آن‌ها را خروجی دهد.
ب. یک نمودار رسم کند که داده‌های x و y به‌صورت scatter روی آن رسم شده‌باشند و خط $y = ax + b$ نیز روی این داده‌ها رسم شده باشد تا بتوانیم دقت خروجی‌مان را با استفاده از نمودار بسنجیم.
(منظور از توان ۲ی بردار یا ماتریس درواقع جمع توان ۲ی تمام درایه‌های آن است همچنین برای سادگی فرض کنید $\sigma_0 = \sigma = 1$)

پاسخ ۵

$$P(w|y, X) = P(y|x, y)P(w) \propto exp(-\frac{(y - w^T X)^2}{2\sigma^2})exp(-\frac{w^2}{2\sigma_0^2})$$

$$\longrightarrow log(P(w \mid y, X)) \propto -\frac{1}{2\sigma^2} \sum (y_i - w^T x_i)^2 - \frac{1}{2\sigma_0^2} w^2$$

$$f := \frac{1}{2\sigma^2} \sum (y_i - w^T x_i)^2 + \frac{1}{2\sigma_0^2} w^2 = \frac{1}{2}(\sum (y_i - ax_i - b)^2) - \frac{1}{2}(a^2 + b^2)$$

$$\frac{df}{da} = -\sum x_i(y_i - ax_i - b) + a = 0$$

$$\Longrightarrow a = \frac{\sum xy - b \sum x}{1 + \sum (x^2)}$$

$$\frac{df}{db} = -\sum (y - ax - b) + b \longrightarrow \sum y - a \sum x - nb = b$$

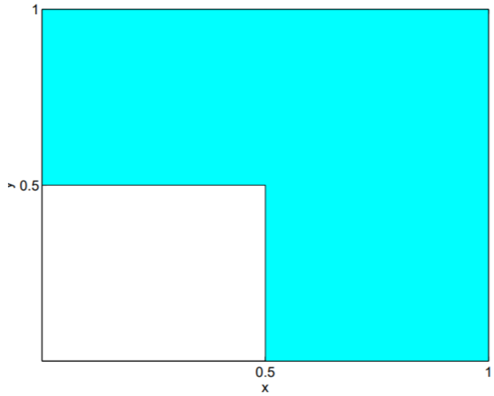
$$\Longrightarrow b = \frac{\sum y - a \sum x}{n+1}$$

با جایگذری b در a داریم:

$$a = \frac{\frac{\sum xy}{1+\sum x^2} - \frac{\sum x \sum y}{(n+1)(1+\sum x^2)}}{1 - \frac{(\sum x)^2}{(n+1)(1+\sum x^2)}}$$

کد قسمت عملی این سوال در فایل src.R قرار گرفته‌است.

سؤال ۶ در شکل زیر نقاط تصادفی (X,Y) به صورت یکنواخت روی ناحیه آبی پخش شده‌اند.



(الف) تخمینگر MMSE ، $\hat{y}_{mmse}(x)$ را محاسبه کنید.

(ب) تخمینگر LMMSE ، $\hat{y}_{lmmse}(x)$ را محاسبه کنید.

پاسخ ۶ (الف) می‌دانیم $\hat{y}_{mmse}(x)=E[Y|X=x]$. برای محاسبه توزیع $Y|X=x$ طبق شکل می‌دانیم هنگامی که $x\in[0,0.5]$ باشد توزیع شرطی Y یکنواخت بین $[0.5,1]$ خواهد بود ($f_{Y|X=x}(y)=\frac{1}{0.5}=2$). و وقتی $x\in[0.5,1]$ باشد توزیع شرطی Y یکنواخت بین $[0,1]$ است ($f_{Y|X=x}(y)=\frac{1}{1}=1$).

$$\hat{y}_{mmse}(x)=E[Y|X=x]=\begin{cases}\int_{0.5}^{1} y\times 2=2\times\frac{1}{2}(1-\frac{1}{4})=\frac{3}{4}, & x\in[0,0.5)\\ \int_{0}^{1} y\times 1=1\times\frac{1}{2}(1-0)=\frac{1}{2}, & x\in[0.5,1]\end{cases}$$

(ب) می‌دانیم:

$$\hat{y}_{LMMSE}(x)=E[Y]+\frac{Cov(X,Y)}{Var(X)}(x-E[X])$$

حال باید پارامترهای مورد نیاز را محاسبه کنیم. با توجه به اینکه توزیع یکنواخت است، تابع چگالی احتمال $f_{X,Y}(x,y)$ برابر $\frac{1}{\text{مساحت ناحیه}}$ است، یعنی داریم : $f_{X,Y}(x,y)=\frac{4}{3}$. با توجه به شکل نیز می‌توانیم دریابیم که:

$$f_X(x)=\int_y f_{X,Y}(x,y)dy=\begin{cases}\int_{0.5}^{1} dy\frac{4}{3}=\frac{2}{3}, & x\in[0,\frac{1}{2})\\ \int_{0}^{1} dy\frac{4}{3}=\frac{4}{3}, & x\in[\frac{1}{2},1]\\ 0, & \end{cases}=\begin{cases}\frac{2}{3}, & x\in[0,\frac{1}{2})\\ \frac{4}{3}, & x\in[\frac{1}{2},1)\\ 0, & \end{cases}$$

بنابراین داریم :

$$E[X]=\int_{-\infty}^{\infty} xf_X(x)dx=\int_{0}^{\frac{1}{2}} x\frac{2}{3}dx+\int_{\frac{1}{2}}^{1} x\frac{4}{3}dx=\frac{7}{12}$$

$$E[X^2]=\int_{-\infty}^{\infty} x^2f_X(x)dx=\int_{0}^{\frac{1}{2}} x^2\frac{2}{3}dx+\int_{\frac{1}{2}}^{1} x^2\frac{4}{3}dx=\frac{5}{12}$$

$$Var[X]=E[X^2]-E[X]^2=\frac{11}{144}.$$

با توجه به تقارن نیز می‌دانیم Y هم توزیعی مشابه X دارد و $E[Y]=\frac{7}{12}, Var[Y]=\frac{11}{144}$. حال باید مقدار کواریانس دو متغیر را محاسبه کنیم، داریم:

$$E[XY]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy=\int_{0}^{0.5}\int_{0.5}^{1} xy\frac{4}{3}dxdy+\int_{0.5}^{1}\int_{0}^{1} xy\frac{4}{3}dxdy=\frac{5}{8}$$

$$Cov(X,Y)=E[XY]-E[X]E[Y]=\frac{5}{8}-(\frac{7}{12})^2=\frac{41}{144}.$$

بنابراین در نهایت داریم :

$$\hat{y}_{LMMSE}(x) = \frac{7}{12} + \frac{\frac{41}{144}}{\frac{11}{144}}(x - \frac{7}{12}) = \frac{7}{12} + \frac{41}{11}(x - \frac{7}{12}).$$

سؤال ۷ فرض کنید تخمینگر زیر را برای پارامتر θ داریم که مبنی بر مشاهده یک نمونه (x_0) است.

$$\hat{\theta} = ax_0^2 + bx_0 + c$$

ضرایب a, b, c را چنان بیابید که خطای MSE مربوط به این تخمینگر کمینه شود. میدانیم x_0 از توزیع یکنواخت در بازه $[-0.5, 0.5]$ است و $\theta = cos(2\pi x_0)$، تخمینگر LMMSE را نیز برای θ بیابید. MSE کمینه هر دو تخمینگر را یافته و با هم مقایسه کنید.

پاسخ ۷ ابتدا برای تخمینگر داده شده محاسبات را انجام میدهیم.

$$MSE[\hat{\Theta}] = E[(\Theta - \hat{\Theta})^2] = E[(\Theta - aX_0^2 - bX_0 - c)^2].$$

برای کمینه کردن این خطا نسبت به ضرایب مشتق میگیریم. در شیوه نوشتار زیر به خطی بودن امید ریاضی و مشتق توجه کنید.

$$\begin{gathered}
0 = \frac{\partial MSE[\hat{\Theta}]}{\partial a} \\
0 = \frac{\partial MSE[\hat{\Theta}]}{\partial b} \\
0 = \frac{\partial MSE[\hat{\Theta}]}{\partial c} \\
0 = E[2(\Theta - aX_0^2 - bX_0 - c)(-X_0^2)] \\
0 = E[2(\Theta - aX_0^2 - bX_0 - c)(-X_0)] \\
0 = E[-2(\Theta - aX_0^2 - bX_0 - c)] \\
E[\Theta X_0^2] = E[(aX_0^4 + bX_0^3 + cX_0^2)] \\
E[\Theta X_0] = E[(aX_0^3 + bX_0^2 + cX_0)] \\
E[\Theta] = E[(aX_0^2 + bX_0 + c)]
\end{gathered}$$

حال باید مقادیر مورد نیاز را با توجه به $X_0 \sim uniform[-0.5, 0.5]$ و $\theta = cos(2\pi x_0)$ با استفاده از تعاریف محاسبه کنیم، داریم :

$$E[X^n] = \int_{-0.5}^{0.5} x^n F_X(x)dx = \int_{-0.5}^{0.5} x^n dx = \frac{1}{n+1}((0.5)^{n+1} - (-0.5)^{n+1}) = \begin{cases} \frac{(0.5)^n}{n+1}, & n: even \\ 0, & else \end{cases}$$

$$E[X_0] = 0, \quad E[X_0^2] = \frac{1}{12}, \quad E[X_0^3] = 0, \quad E[X_0^4] = \frac{1}{80}$$

$$E[\Theta X^2] = \int_{-0.5}^{0.5} \theta x^2 F_x(x)dx = \int_{-0.5}^{0.5} cos(2\pi x)x^2 dx = \frac{-1}{2\pi^2}$$

$$E[\Theta X] = \int_{-0.5}^{0.5} \theta x F_x(x)dx = \int_{-0.5}^{0.5} cos(2\pi x)x dx = 0$$

$$E[\Theta] = \int_{-0.5}^{0.5} cos(2\pi x)dx = 0$$

$$E[\Theta^2] = \int_{-0.5}^{0.5} cos(2\pi x)^2 dx = \frac{1}{2}$$

با استفاده از مقادیر بالا میتوانیم معادلات را به صورت دستگاه زیر بازنویسی کنیم.

$$\begin{bmatrix} \frac{-1}{2\pi^2} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{80} & 0 & \frac{1}{12} \\ 0 & \frac{1}{12} & 0 \\ \frac{1}{12} & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

با حل دستگاه خواهیم داشت:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \frac{-90}{\pi^2} \\ 0 \\ \frac{15}{2\pi^2} \end{bmatrix}$$

بنابراین داریم :

$$\hat{\theta} = \frac{-90}{\pi^2} x_0^2 + \frac{15}{2\pi^2}$$

حال باید MSE مربوط به این تخمینگر را محاسبه کنیم.

$$MSE[\hat{\theta}] = E[(\Theta - \hat{\Theta})^2] = E[\Theta^2 - 2\Theta(\frac{-90}{\pi^2} X_0^2 + \frac{15}{2\pi^2}) + (\frac{-90}{\pi^2} X_0^2 + \frac{15}{2\pi^2})^2] =$$
$$E[\Theta^2] + \frac{180}{\pi^2} E[\Theta X_0^2] - \frac{15}{\pi^2} E[\Theta] + (\frac{15}{2\pi^2})^2 (144 E[X_0^4] - 24 E[X_0^2] + 1) =$$
$$0.5 - \frac{180}{\pi^2} \times (\frac{-1}{2\pi^2}) + 0 + (\frac{15}{2\pi^2})^2 \times 0.8 = 0.5 - \frac{45}{\pi^2}$$

حال باید تخمینگر LMMSE را محاسبه میکنیم.

$$\hat{\theta}_L = E[\hat{\Theta}] + \frac{Cov(\Theta, X_0)}{Var(X_0)} (x_0 - E[X_0])$$
$$\hat{\theta}_L = 0 + \frac{E[\Theta X_0] - E[\Theta] E[X_0]}{Var(X_0)} (X_0 - 0) = \frac{0 - 0}{\frac{1}{12}} x_0 = 0$$

برای محاسبه خطای MSE مربوط به تخمینگر خطی نیز داریم:

$$MSE(\theta_L) = E[(\Theta - \hat{\Theta})^2] = E[\Theta^2] = \frac{1}{2}$$

مشاهده میشود خطای تخمینگر مربعی از تخمینگر خطی کمتر است.

سؤال ۸ فرض کنیم توزیع های زیر را می‌دانیم.

$$X \sim N(\mu, \tau^2) \quad , \quad Y|X = x \sim N(x, \sigma^2)$$

(الف) توزیع $X|Y = y$ را محاسبه کنید.

(ب) فرض کنید که ما n وقوع تصادفی $(Y_1, ..., Y_n)$ از توزیع $Y|X = x$ را مشاهده کرده‌ایم. حال توزیع $X|\bar{Y}$ را محاسبه کنید که در آن $\bar{Y}$ میانگین نمونه $(\bar{Y} = \frac{\sum_i Y_i}{n})$ است. (میتوانید از نتایج الف استفاده کنید.)

(پ) تخمینگر MAP و MMSE را برای X بر حسب مشاهده $\bar{Y}$ بیابید.

پاسخ ۸ (الف) از قاعده بیز استفاده میکنیم.

$$f_{X|Y}(X|Y) = \frac{f_{Y|X}(Y|X)P(X)}{P(Y)} = \frac{f_{Y|X}(Y|X)P(X)}{\int_{-\infty}^{\infty} f_{Y|X}(Y|X)P(X)dx}$$

$$f_{X|Y}(X|Y) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-x)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\tau^2}}e^{-\frac{(x-\mu)^2}{2\tau^2}}}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-x)^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\tau^2}}e^{-\frac{(x-\mu)^2}{2\tau^2}}}$$

$$f_{X|Y}(X|Y) = \frac{e^{-\frac{(y-x)^2}{2\sigma^2}} e^{-\frac{(x-\mu)^2}{2\tau^2}}}{\int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{2\sigma^2}} e^{-\frac{(x-\mu)^2}{2\tau^2}}}$$

با محاسبه انتگرال داریم :

$$\int_{-\infty}^{\infty} e^{-\frac{(y-x)^2}{2\sigma^2}} e^{-\frac{(x-\mu)^2}{2\tau^2}} = \frac{\sqrt{2\pi}\sigma\tau e^{-\frac{y^2-2\mu y+\mu^2}{2\tau^2+2\sigma^2}}}{\sqrt{\sigma^2+\tau^2}}$$

$$f_{X|Y}(X|Y) = \frac{e^{-\frac{(y-x)^2}{2\sigma^2}} e^{-\frac{(x-\mu)^2}{2\tau^2}}}{\frac{\sqrt{2\pi}\sigma\tau e^{-\frac{y^2-2\mu y+\mu^2}{2\tau^2+2\sigma^2}}}{\sqrt{\sigma^2+\tau^2}}} = \frac{1}{\sqrt{2\pi\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}} e^{-\frac{(x-\frac{y/\sigma^2+\mu/\tau^2}{1/\sigma^2+1/\tau^2})}{1/\sigma^2+1/\tau^2}}$$

$$\rightarrow X|Y = y \sim N(\frac{y/\sigma^2+\mu/\tau^2}{1/\sigma^2+1/\tau^2}, 1/\sigma^2+1/\tau^2)$$

(ب)

$$Y|X = x \sim N(x, \sigma^2)$$
$$\bar{Y}|X = x \sim N(x, \frac{\sigma^2}{n})$$

با توجه به بخش الف داریم:

$$X|\bar{Y} \sim N(\frac{n\bar{Y}/\sigma^2+\mu/\tau^2}{n/\sigma^2+1/\tau^2}, \frac{1}{n/\sigma^2+1/\tau^2})$$

(پ) از آنجا که توزیع‌ها نرمال هستند مقدار ماکسیمم آنها در میانگین رخ می‌دهد و تخمینگر MAP به صورت زیر است.

$$\hat{X}_{MAP} = \arg\max f_{X|\bar{Y}}(x) = \frac{n\bar{Y}/\sigma^2+\mu/\tau^2}{n/\sigma^2+1/\tau^2}$$

تخمینگر MMSE نیز به صورت زیر خواهد بود.

$$\hat{X}_{MMSE} = E[X|\bar{Y}] = \frac{n\bar{Y}/\sigma^2+\mu/\tau^2}{n/\sigma^2+1/\tau^2}$$

سؤال ۹ فرض کنید $X \sim Uniform(1,2)$ و همچنین میدانیم توزیع $Y|X=x$ نمایی با پارامتر $\lambda = \frac{1}{x}$ است.

(الف) تخمین LMMSE ، X بر حسب Y را بیابید. ($\hat{X}_L$)

(ب) خطای MSE مربوط به تخمین بخش الف را بیابید.

(پ) نشان دهید $E[(X-\hat{X}_L)Y]=0$.

پاسخ ۹ (الف)

$$\hat{X}_L = \frac{Cov(X,Y)}{Var(Y)}(Y - E[Y]) + E[X]$$

حال با محاسبه پارامترهای مورد نظر داریم:

$$E[X] = \frac{1+2}{2} = \frac{3}{2}$$
$$E[Y] = E[E[Y|X]] = E[X] = \frac{3}{2}$$
$$E[Y^2] = E[E[Y^2|X]] = E[2X^2] = \int_1^2 2x^2 dx = \frac{14}{3}$$
$$Var(Y) = E[Y^2] - (E[Y])^2 = \frac{29}{12}$$
$$E[XY] = E[E[XY|X]] = E[XE[Y|X]] = E[X.X] = E[X^2] = \int_1^2 x^2 dx = \frac{7}{3}$$
$$Cov(X,Y) = E[XY] - E[X]E[Y] = \frac{7}{3} - \frac{3}{2}\frac{3}{2} = \frac{1}{12}.$$

بنابراین داریم:

$$\hat{X}_L = \frac{Cov(X,Y)}{Var(Y)}(Y - E[Y]) + E[X] = \frac{Y}{29} + \frac{42}{29}.$$

(ب)

$$MSE = E[(\hat{X}_L - X)^2] = E[(\frac{Y}{29} + \frac{42}{29} - X)^2] = \frac{7}{87}$$

(پ)

$$E[(X - \frac{Y}{29} - \frac{42}{29})Y] = E[XY] - \frac{E[Y^2]}{29} - \frac{42}{29}E[Y] = \frac{7}{3} - \frac{14}{3.29} - \frac{42}{29}.\frac{3}{2} = 0.$$

سؤال ۱۰ فرض کنید یک پارامتر تصادفی $\Theta \in R$ داریم که تابع PMF آن به صورت $P(\theta) = \frac{1}{2}(\delta(\theta-1)+\delta(\theta+1))$ است (δ بیانگر تابع دلتای دیراک ست که در صورتی که $x=0$ باشد مقدار $\delta(x)=1$ و در غیر این صورت مقدار صفر به خود میگیرد). فرض کنیم که ما خروجی نویزی Y را مشاهده کرده‌ایم.

$$Y = \Theta + W$$

که در آن $W \sim N(0,\sigma^2)$ است و مقدار σ^2 در آن معین است. تخمینگر MMSE و MAP را برای پارامتر Θ محاسبه کنید و بحث کنید تحت چه شرایطی این دو تخمین تقریبا یکی هستند.

پاسخ ۱۰ ابتدا توزیع شرطی $Y|\Theta=\theta$ را محاسبه میکنیم.

$$f_{(Y|\Theta)}(Y|\Theta=\theta) = \begin{cases} N(1,\sigma^2), & \theta = 1 \\ N(=1,\sigma^2), & \theta=-1 \end{cases}$$

حال توزیع حاشیه ای Y را محاسبه میکنیم.

$$f_Y(y) = \int_\theta f(Y|\Theta=\theta)p(\theta)d\theta = \frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y-1)^2}{2\sigma^2}} + \frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y+1)^2}{2\sigma^2}}$$

حال توسط قاعده بیز توزیع $\Theta|Y$ را محاسبه میکنیم.

$$P(\Theta|Y) = \frac{P(Y|\Theta)P(\Theta)}{P(Y)} = \begin{cases} \dfrac{\frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y-1)^2}{2\sigma^2}}\frac{1}{2}}{\frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y-1)^2}{2\sigma^2}}+\frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y+1)^2}{2\sigma^2}}} & if \quad \theta=1 \\ \dfrac{\frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y+1)^2}{2\sigma^2}}\frac{1}{2}}{\frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y-1)^2}{2\sigma^2}}+\frac{1}{2}\frac{1}{\sqrt{2\sigma^2}}e^{-\frac{(y+1)^2}{2\sigma^2}}} & if \quad \theta=-1 \end{cases}$$

$$P(\Theta|Y) = \begin{cases} \frac{1}{1+e^{-\frac{2y}{\sigma^2}}}, & if \quad \theta=1 \\ \frac{1}{1+e^{\frac{2y}{\sigma^2}}}, & if \quad \theta=1 \end{cases}$$

حال تخمینگر MMSE را محاسبه میکنیم.

$$\hat{\theta}_{MMSE}(y) = E[\Theta|Y] = P(\Theta=1|y) - P(\Theta=-1|y) = \frac{e^{\frac{y}{\sigma^2}} - e^{\frac{-y}{\sigma^2}}}{e^{\frac{y}{\sigma^2}} + e^{\frac{-y}{\sigma^2}}} = tanh(\frac{y}{\sigma^2})$$

برای تخمینگر MAP نیز داریم.

$$\hat{\theta}_{MAP}(y) = argmax_{\theta=\pm1}P(\Theta|Y) =$$
$$\begin{cases} 1 & if \quad \frac{1}{1+e^{-\frac{2y}{\sigma^2}}} \geq \frac{1}{1+e^{+\frac{2y}{\sigma^2}}} \\ -1 & if \quad \frac{1}{1+e^{-\frac{2y}{\sigma^2}}} < \frac{1}{1+e^{+\frac{2y}{\sigma^2}}} \end{cases}$$
$$= \begin{cases} 1 & if \quad y \geq 0 \\ -1 & if \quad y<0 \end{cases} = sgn(y)$$

میدانیم اگر x به سمت $+\infty$ برود، مقدار $tanh(x)$ به ۱ نزدیک میشود و اگر x به سمت $-\infty$ برود، مقدار $tanh(x)$ به -1 میرود بنابراین لازم است که $\sigma << 1$ یعنی بسیار کوچک باشد تا دو تخمینگر تقریبا یکی شوند.

موفق باشید