Bayesian Machine Learning: Assignment #1

Due on September 29, 2020 at 11:59pm

Prof. Juho Lee

Hyunsu Kim

Problem 1

(a) Using Integration by parts,

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x = \int_{-\infty}^{\theta} x \cdot 0\,\mathrm{d}x + \int_{\theta}^{\infty} xe^{\theta-x}\,\mathrm{d}x = 1+\theta$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \int_{\theta}^{\infty} x^2 e^{\theta-x}\,\mathrm{d}x - (1+\theta)^2 = \theta^2 + 2(1+\theta) - (1+\theta)^2 = 1$$

(b) To show unbiasedness, compare θ with the exepectation. Note X_i's are i.i.d., and $\mathrm{E}[X_i] = 1+\theta$ in (a):

$$\begin{aligned}
\mathrm{E}[\hat{\theta}_n] &= \mathrm{E}[\frac{1}{n}\sum_{i=1}^{n}(X_i - 1)] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i - 1] \\
&= \frac{1}{n}\sum_{i=1}^{n}(\mathrm{E}[X_i] - 1) \\
&= \frac{1}{n}\sum_{i=1}^{n}((1+\theta) - 1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\theta \\
&= \frac{1}{n}(n\theta) \\
&= \theta
\end{aligned}$$

(c) Let $Y_i = X_i - 1$. Then, $\mathrm{E}[Y_i] = \theta$ and $\mathrm{Var}[Y_i] = 1$. Now, according to (b) and Central Limit Theorem,

$$\frac{\hat{\theta}_n - \theta}{1/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

holds. Then, $100(1-\alpha)\%$ confidence interval of θ is $[\hat{\theta}_n - \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}}, \hat{\theta}_n - \frac{\Phi^{-1}(\frac{\alpha}{2})}{\sqrt{n}}]$ where Φ is Cumulative Density Function (CDF) of unit normal distribution.

(d) From observations, we can easily compute $\hat{\theta}_3 = \frac{1}{3}\{(10.0-1) + (12.0-1) + (15.0-1)\} = 11.33$. According to given fact and that unit normal distribution is an even function, $\Phi^{-1}(0.025) = -1.96$ and $\Phi^{-1}(0.975) = 1.96$. Thus, 95% confidence interval of θ is $[11.33 - \frac{1.96}{\sqrt{3}}, 11.33 + \frac{1.96}{\sqrt{3}}] = [10.20, 12.46]$. It is weird that the observed data 10.0 is contradictory for any θ in the obtained confidence interval according to the PDF of X. Such odd situation is indeed expected to happen because of the small sample size. ($n = 3$) Also, the computed confidence interval is based on frequentist approach, of which "95%" stands for how frequently would θ be contained in the interval as we repeat the procedure of computing confidence interval.

Problem 2

1. The proof is straightforward:

$$\begin{aligned}
\mu(B) &= \mu(B \cap (A \cup A^c)) \\
&= \mu((B \cap A) \cup (B \cap A^c)) && \because \text{ distributive law} \\
&= \mu(B \cap A) + \mu(B \cap A^c) && \because \text{ countable additivity of } \mu \\
&= \mu(A) + \mu(B \cap A^c) && \because A \subset B \implies B \cap A = A \\
&\geq \mu(A) && \because \mu(B \cap A^c) \geq 0 : \text{ nonnegativity of } \mu
\end{aligned}$$

2. Use induction. (Base case) We first show it holds for $n = 1$:

$$\begin{aligned}
\mu(\bigcup_{i=1}^{1} A_i) &= \mu(A_1) \\
&= \sum_{i=1}^{1} \mu(A_i) \\
&\leq \sum_{i=1}^{1} \mu(A_i)
\end{aligned}$$

(Inductive case) Assume it holds for $n = k$. Want to show it holds for $n = k + 1$:

$$\begin{aligned}
\mu(\bigcup_{i=1}^{k+1} A_i) &= \mu(\bigcup_{i=1}^{k} A_i \cup A_{k+1}) \\
&= \mu(\bigcup_{i=1}^{k} A_i \cup (A_{k+1} - \bigcup_{i=1}^{k} A_i)) && \because \text{ simple Venn diagram argument} \\
&= \mu(\bigcup_{i=1}^{k} A_i) + \mu(A_{k+1} - \bigcup_{i=1}^{k} A_i) && \because \text{ countable additivity of } \mu \\
&\leq \mu(\bigcup_{i=1}^{k} A_i) + \mu(A_{k+1}) && \because (A_{k+1} - \bigcup_{i=1}^{k} A_i) \subseteq A_{k+1} \text{ and 1.} \\
&\leq \sum_{i=1}^{k} \mu(A_i) + \mu(A_{k+1}) && \because \text{ inductive hypothesis} \\
&= \sum_{i=1}^{k+1} \mu(A_i)
\end{aligned}$$

□

Problem 3

To show the asserted convergence in probability, want to show

$$\lim_{n \to \infty} \mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| > \epsilon) = 0$$

for any $\epsilon > 0$. Note $\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$ is a sample mean of i.i.d. random variables so that $\mathrm{E}[\overline{X}_n] = \mathrm{E}[X_1]$ and $\mathrm{Var}[\overline{X}_n] = \frac{\mathrm{Var}[X_1]^2}{n}$. Using Chebyshev's inequality for $k > 0$:

$$\mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| \geq k \frac{\mathrm{Var}[X_1]}{\sqrt{n}}) \leq \frac{1}{k^2} \tag{1}$$

holds. Now, let $k = \frac{\sqrt{n}\epsilon}{\mathrm{Var}[X_1]}$. Then, (1) becomes as follows:

$$\mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| \geq \epsilon) \leq \frac{Var[X_1]^2}{n\epsilon^2}$$

Then, as we take $n \to \infty$ on both sides of inequality:

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X}_n - \mathrm{E}[X_1]| > \epsilon) \leq \lim_{n\to\infty} \frac{Var[X_1]^2}{n\epsilon^2} = 0$$

holds for any $\epsilon > 0$. $\square$

Problem 4

For any $\epsilon > 0$, the following holds:

$$\begin{aligned}
\mathbb{P}(|X_n + Y_n - (X+Y)| > \epsilon) &\leq \mathbb{P}(|X_n - X| > \frac{\epsilon}{2} \text{ or } |Y_n - Y| > \frac{\epsilon}{2}) && \because \text{contraposition and triangular ineq.} \\
&\leq \mathbb{P}(|X_n - X| > \frac{\epsilon}{2}) + \mathbb{P}(|Y_n - Y| > \frac{\epsilon}{2}) && \because \text{union bound} \\
&\to 0 \text{ as } n \to \infty. && \because X_n \xrightarrow{p} X \text{ and } Y_n \xrightarrow{p} Y
\end{aligned}$$

Thus, $X_n + Y_n \xrightarrow{p} X + Y$. $\square$

Problem 5

Use the change of variable

$$(Y_1, \cdots, Y_{n-1}, Z) = g(X_1, \cdots, X_n)$$

to compute f_Y from f_X by

$$f_Y(y_1, \cdots, y_n) = f_X(g^{-1}(y_1, \cdots, y_n))|J_{g^{-1}}|$$

where

$$f_X(x_1, \cdots, x_n) = \textstyle\prod_{i=1}^n \frac{x_i^{\alpha_i - 1} e^{x_i}}{\Gamma(\alpha_i)}$$

is the joint density of X_i's. (i.i.d. random variables) We need to compute the determinant of Jacobian of g^{-1}. Note that

$$X_i = Y_i Z, \quad i = 1, \cdots, n-1$$
$$X_n = Z - (X_1 + \cdots + X_{n-1}) = Z - (Y_1 Z + \cdots + Y_{n-1} Z) = Z(1 - Y_1 - \cdots - Y_{n-1}).$$

Then,

$$\begin{aligned}
|J_{g^{-1}}| &= \begin{vmatrix} z & 0 & \cdots & 0 & y_1 \\ 0 & z & \cdots & 0 & y_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & z & y_{n-1} \\ -z & -z & \cdots & -z & 1 - y_1 - \cdots - y_{n-1} \end{vmatrix} \\
&= \begin{vmatrix} z & 0 & \cdots & 0 & y_1 \\ 0 & z & \cdots & 0 & y_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & z & y_{n-1} \\ 0 & 0 & \cdots & 0 & 1 \end{vmatrix} && \because \text{invariant under elementary row operations} \\
&= z^{n-1}. && \because \text{upper triangular matrix}
\end{aligned}$$

Now, we get the joint density of $Y_1, \cdots, Y_{n-1}, Z$:

$$f(y_1, \cdots, y_{n-1}, z) = \frac{y_1^{\alpha_1 - 1} \cdots y_{n-1}^{\alpha_{n-1}-1}(1-y_1-\cdots-y_{n-1})^{\alpha_n - 1}}{\prod_{i=1}^n \Gamma(\alpha_i)} z^{\sum_{i=1}^n \alpha_i - 1} e^{-z}.$$

Finally, marginalize out z together with the definition of Gamma function (Γ), to get f_Y:

$$f(y_1, \cdots, y_{n-1}) = \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} y_1^{\alpha_1 - 1} \cdots y_{n-1}^{\alpha_{n-1}-1}(1 - y_1 - \cdots - y_{n-1})^{\alpha_n - 1}$$

or,

$$f_Y(y_1, \cdots, y_n) = \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n y_i^{\alpha_i - 1}$$

by abuse of notation. (since y_n is dependent to other y_i's) □

Problem 6

TODO □

Problem 7

The procedure is very similar with that in Problem 5.

$$\begin{aligned}
\mathbb{P}(|X_n + Y_n - (X+Y)| > \epsilon) &\leq \mathbb{P}(|X_n - X| > \frac{\epsilon}{2} \text{ or } |Y_n - Y| > \frac{\epsilon}{2}) && \because \text{contraposition and triangular ineq.} \\
&\leq \mathbb{P}(|X_n - X| > \frac{\epsilon}{2}) + \mathbb{P}(|Y_n - Y| > \frac{\epsilon}{2}) && \because \text{union bound} \\
&\to 0 \text{ as } n \to \infty. && \because X_n \xrightarrow{p} X \text{ and } Y_n \xrightarrow{p} Y
\end{aligned}$$

Thus, $X_n + Y_n \xrightarrow{p} X + Y$. □