STA 35C – Homework 1

Submission due: Tue, April 7 at 11:59 PM PT

Instructor: Dogyoon Song

Instructions: Upload a single PDF file to Canvas ("Homework 1" under "Assignments"). Name the file using the prefix of your UC Davis email ID and the homework number (e.g., `dgsong_hw1.pdf`). Include "STA 35C," your name, and the last four digits of your student ID on the front page. No late submissions will be accepted; any submission received after the deadline will receive 0 points. For full submission requirements and the late submission policy, see the syllabus.

Problem 1 (25 points in total).

Consider the experiment of tossing a fair coin twice. Let

$$\Omega = \{HH, HT, TH, TT\},$$

where, for example, HT denotes "Head on the first toss and Tail on the second toss." Define the events

$$A = \{HH, HT\} \qquad \text{and} \qquad B = \{HH, TH\}.$$

Thus, A is the event that the first toss is Head, and B is the event that the second toss is Head.

(a) **(7 points)** Draw a simple Venn diagram (or any clear schematic) for the events A and B inside Ω by placing the four outcomes HH, HT, TH, TT in the appropriate regions. Then list the outcomes in each of the following events:

$$A \cup B, \qquad A \cap B, \qquad A \cap B^c, \qquad B \cap A^c, \qquad (A \cup B)^c.$$

(b) **(9 points)** Let P be any probability law on a sample space Ω. Using only the probability axioms, prove the following:

(i) $P(A^c) = 1 - P(A)$ for every event $A \subseteq \Omega$.

(ii) If $A \subseteq B$, then $P(A) \le P(B)$.

(iii) $P\big((A \cap B^c) \cup (B \cap A^c)\big) = P(A) + P(B) - 2P(A \cap B)$.

(c) **(9 points)** Another student suggests using the alternative sample space

$$\widetilde{\Omega} = \{0, 1, 2\},$$

where the outcome records only the total number of Heads in the two tosses.

(i) Write down the probability law on $\widetilde{\Omega}$ induced by two fair coin tosses.

(ii) Can the event $A \cup B = \{\text{at least one Head}\}$ be represented as a subset of $\widetilde{\Omega}$? If yes, write that subset.

(iii) Can the event $A = \{\text{first toss is Head}\}$ be represented as a subset of $\widetilde{\Omega}$? Briefly explain why or why not.

Problem 2 (25 points in total).

(a) (5 points) Let X be a random variable with probability density function f_X defined by

$$f_X(x) = \begin{cases} 3x^2, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

(b) (12 points) Suppose X and Y are two random variables with

$$\mathbb{E}[X] = 2, \quad \mathrm{Var}(X) = 4, \quad \mathbb{E}[Y] = 5, \quad \mathrm{Var}(Y) = 1.$$

(i) If X and Y are *independent*, compute $\mathbb{E}[X + 2Y]$ and $\mathrm{Var}(X + 2Y)$.

(ii) Now assume $\mathrm{corr}(X, Y) = 0.5$. Recompute $\mathbb{E}[X + 2Y]$ and $\mathrm{Var}(X + 2Y)$.

(iii) Compare these two results, and briefly comment on why knowledge of correlation matters.

(c) (8 points) Let S denote the event "the email is spam," and let F denote the event "the filter flags the email as spam."

(i) Suppose $\Pr(S) = 0.05$, $\Pr(F \mid S) = 0.90$, and $\Pr(F^c \mid S^c) = 0.95$. If an email is flagged, what is $\Pr(S \mid F)$, the conditional probability that the flagged email is actually spam?

(ii) Briefly explain why this conditional probability $\Pr(S \mid F)$ can be much smaller than $\Pr(F \mid S)$. (*Hint:* consider the roles of the base rate $\Pr(S)$ and the false-positive rate $\Pr(F \mid S^c)$)

Problem 3 (30 points in total).

(a) (10 points) Let X be a discrete random variable taking values in $\{1, 2, 3\}$ with probability mass function

$$P(X = k) = \frac{1}{Z}\theta^k, \quad k = 1, 2, 3,$$

for some parameter $\theta > 0$ and unknown constant Z.

(i) Find the normalizing constant $Z = Z(\theta)$ so that $P(X = k)$ is a valid probability distribution, and write $p_X(k)$ in closed form.

(ii) Using your expression for $p_X(k)$, compute the cumulative distribution function $F_X(k) = \Pr(X \leq k)$ and sketch its graph.

(iii) Compute $\mathbb{E}[X]$ and $\text{Var}(X)$.

(b) (10 points) Let $\alpha > 0$, and suppose that Y is a continuous random variable with density

$$f_Y(y) = \begin{cases} c\, e^{-\alpha y}, & y \geq 0, \\ 0, & y < 0, \end{cases}$$

for some constant $c > 0$. (*Hint*: Your answer may contain α)

(i) Find the value of c so that $f_Y(y)$ integrates to 1 over $[0, \infty)$.

(ii) Derive $F_Y(y) = \Pr(Y \leq y)$ for all $y \in \mathbb{R}$ from $f_Y(y)$, and sketch its graph.

(iii) Compute $\mathbb{E}[Y]$ and $\text{Var}(Y)$.

(c) (10 points) For this part, first assume that X and Y are independent, and define $Z = 2X + Y$.

(i) Sketch (draw a graph) and briefly describe the shape of the distribution of Z in relation to the distributions of X and Y.

(ii) Compute $\mathbb{E}[\,Z\,]$ and $\text{Var}(Z)$ (still assuming X and Y are independent).

(iii) Now drop the independence assumption and instead assume that $\text{corr}(X, Y) = 0.4$. Recompute $\mathbb{E}[Z]$ and $\text{Var}(Z)$. Briefly explain why this information is enough to determine the expectation and variance of Z, but not the full distribution of Z.

Problem 4 (20 points in total + 5 bonus points).

Suppose that we have a coin with unknown head probability $p_{\text{true}} = \Pr(\text{Head}) \in [0, 1]$, which *we believe* can only take one of three possible values:

$$p \in \{0.2,\ 0.5,\ 0.8\}.$$

Let us denote this model parameter by θ, and place a *uniform prior* on these three values:

$$P(\theta = 0.2) \ = \ P(\theta = 0.5) \ = \ P(\theta = 0.8) \ = \ \frac{1}{3}. \tag{1}$$

This reflects no prior preference, or lack of any prior information, among the three values.

(a) **(5 points)** Suppose we flip the coin once and observe $X \in \{0, 1\}$, where 1 denotes a Head and 0 denotes a Tail. Using the law of total probability, compute the marginal probability of observing a head:

$$P(X = 1) \ = \sum_{k \in \{0.2,\, 0.5,\, 0.8\}} P(X = 1 \mid \theta = k)\, P(\theta = k).$$

Then compute $P(X = 0)$ similarly. Give numerical values under the uniform prior as in (1).

(b) **(5 points)** The *posterior* probability for each $\theta \in \{0.2, 0.5, 0.8\}$ if the observed flip is a head ($X = 1$) is derived as:

$$\underbrace{P(\theta = k \ | \ X = 1)}_{\text{posterior}} \ = \ \frac{P(X = 1 \mid \theta = k)\ \overbrace{P(\theta = k)}^{\text{prior}}}{P(X = 1)}.$$

Compute these three posterior probabilities for $k = 0.2,\ 0.5,\ 0.8$. Repeat for $X = 0$. Check that each set of posteriors (conditioned on $X = 0$ and $X = 1$) sums to 1.

(c) **(10 points)** Now consider a *stream of n coin flips* generated from Bernoulli(p_{true}), where the true value p_{true} may or may not belong to the candidate set $\{0.2, 0.5, 0.8\}$. Implement the following steps in an R script, and produce plots for each p_{true}. In particular, compare the cases $p_{\text{true}} = 0.3$ and $p_{\text{true}} = 0.5$. Discuss any differences you observe between these two cases.

(i) Generate n independent coin flips from Bernoulli(p_{true}).

(ii) Initialize your prior to (1/3, 1/3, 1/3). Update it *sequentially* after each flip, using

$$\text{Posterior}(\theta) \ \propto \ \begin{cases} \theta \times \text{Prior}(\theta), & X = 1, \\ (1 - \theta) \times \text{Prior}(\theta), & X = 0, \end{cases}$$

then normalize so the each set of posterior probabilities sum to 1. Use the posterior after the i-th flip as the prior for the $(i + 1)$-th flip.

(iii) Let $n = 100$. Store the posterior probabilities after each flip, and for each value of p_{true}, produce a single plot showing the three trajectories Posterior($\theta = 0.2$), Posterior($\theta = 0.5$), and Posterior($\theta = 0.8$) versus the flip number.

(d*) **(5 bonus points)** Repeat (**(c)**) but with the following changes:

(i) Try different priors, e.g., (0, 1/2, 1/2) or (1/10000, 4999/10000, 1/2).

(ii) Try a different set of possible model parameters instead of $\{0.2, 0.5, 0.8\}$, for example $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ or $\{0, 0.1, 0.2, \ldots, 1\}$.

Discuss your observations, including (1) whether your final estimate (the value of θ with the highest posterior probability) changes, (2) how the speed of convergence is affected, and (3) what happens when a prior assigns probability 0 to a candidate value of θ.