

STA 35C – Homework 1

Submission due: Thu, October 1 at 11:59 PM PT

Instructor: Dogyoon Song

Instructions: Upload a single PDF file to Canvas/Gradescope (“Homework 1” under “Assignments”). For Problem 4, include your R code, plots, and written discussion in the same PDF. 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).

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

(b) (7 points) Consider the experiment of tossing a fair coin twice, independently. 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

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

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

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:

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

(c) (9 points) Continuing on the setup of (b), another student suggests using the alternative sample space

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

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

- (i) Write down the probability law on $\tilde{\Omega}$ induced by two fair coin tosses.
- (ii) Can the event $A \cup B = \{\text{at least one Head}\}$ be represented as a subset of $\tilde{\Omega}$? If yes, write that subset.
- (iii) Can the event $A = \{HH, HT\}$ be represented as a subset of $\tilde{\Omega}$? 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

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

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

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

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

- (i) If X and Y are *independent*, compute $\mathbb{E}[X + 2Y]$ and $\text{Var}(X + 2Y)$.
 - (ii) Now assume $\text{corr}(X, Y) = 0.5$. Recompute $\mathbb{E}[X + 2Y]$ and $\text{Var}(X + 2Y)$.
 - (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 $\Pr(S) = 0.05$, $\Pr(F | S) = 0.90$, and $\Pr(F^c | S^c) = 0.95$. If an email is flagged, what is $\Pr(S | F)$, the conditional probability that the flagged email is actually spam?
 - (ii) Briefly explain why this conditional probability $\Pr(S | F)$ can be much smaller than $\Pr(F | S)$.
(*Hint:* consider the roles of the base rate $\Pr(S)$ and the false-positive rate $\Pr(F | S^c)$)

Problem 3 (25 points in total + 5 bonus points).

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

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

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

- (i) Find the normalizing constant $C = C(\theta)$ 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 $F_X(x) = \Pr(X \leq x)$ for all $x \in \mathbb{R}$ and sketch its graph, labeling the jump heights.
- (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) (5 points + 5 bonus points) For this part, first assume that X and Y are independent, and define $Z = 2X + Y$.

- (i) Compute $\mathbb{E}[Z]$ and $\text{Var}(Z)$ (still assuming X and Y are independent).
- (ii*) (5 bonus points) Now drop the independence assumption. Express $\text{Var}(Z)$ in terms of $\text{Var}(X)$, $\text{Var}(Y)$, and $\text{Cov}(X, Y)$. Briefly explain whether or not knowing $\text{corr}(X, Y)$ together with the marginal distributions determines the full distribution of Z , beyond $\mathbb{E}[Z]$ and $\text{Var}(Z)$.

Problem 4 (25 points in total + 5 bonus points).

Suppose a coin has unknown head probability $p_{\text{true}} = \Pr(\text{Head}) \in [0, 1]$. We model it using candidate values:

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

Under this model, $P(\text{Head} \mid \theta) = \theta$. Assign the *uniform prior*

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

This reflects no prior preference, or lack of any prior information, among the three values (As such, we subjectively believe that θ can be each of these three values equally likely before collecting data).

- (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) **(8 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:

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

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) **(12 points)** Now consider a *stream of $n = 100$ coin flips* generated from $\text{Bernoulli}(p_{\text{true}})$, 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 R for $p_{\text{true}} = 0.3$ and $p_{\text{true}} = 0.5$. Use one simulated sequence per case and set a random seed for reproducibility. In particular, compare the cases $p_{\text{true}} = 0.3$ and $p_{\text{true}} = 0.5$. Discuss any similarities and differences you observe between the two plots in 2–4 sentences.

Below is a suggested workflow:

- (i) Generate n independent coin flips from $\text{Bernoulli}(p_{\text{true}})$.
- (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) Store the posterior probabilities after each flip, and for each value of p_{true} , produce a single plot showing the three trajectories $\text{Posterior}(\theta = 0.2)$, $\text{Posterior}(\theta = 0.5)$, and $\text{Posterior}(\theta = 0.8)$ versus the flip number.

- (d*) **(5 bonus points)** Use $p_{\text{true}} = 0.3$ and the same simulated flips as in part (c).

- (i) Replace the uniform prior with something different, e.g., $(0, 1/2, 1/2)$ or $(1/10000, 4999/10000, 1/2)$.
- (ii) Try a different set of possible model parameters, for example $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ or $\{0, 0.1, 0.2, \dots, 1\}$.

Compare the plots in a few sentences: does the most probable final candidate change, and how quickly does the posterior appear to concentrate? For (i), also explain the effect of assigning prior probability 0 to a candidate.