STA 35C – Homework 1

Submission due: Tue, April 15 at 11:59 PM PT

Instructor: Dogyoon Song

Instructions: Upload a PDF file, named with your UC Davis email ID and homework number (e.g., `dgsong_hw1.pdf`) to Gradescope (accessible through Canvas). Please make sure to include "STA 35C," your name, and the last four digits of your student ID on the front page. For more details on submission requirements and the late submission policy, see the syllabus.

Problem 1 (25 points in total).

(a) (9 points) Let X take values in $\{1, 2, 3\}$ with probabilities proportional to

$$P(X = k) \propto \theta^k, \quad k = 1, 2, 3,$$

for some constant $\theta > 0$.

(i) Find the normalizing constant 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 $F_X(k)$ and sketch its graph.

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

(b) (9 points) Let $\alpha > 0$, and suppose that Y is a continuous random variable with density

$$f_Y(y) = c\, e^{-\alpha y}, \quad y \geq 0,$$

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 $F_y(y)$ from $f_Y(y)$ and sketch its graph.

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

(c) (7 points) Define $Z = 2X + Y$.

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

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

(iii) Repeat parts (i) and (ii) under the assumption that the correlation coefficient between X and Y is 0.4. Discuss any differences in the resulting distribution and the values of expectation/variance of Z.

Problem 2 (25 points in total).

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:

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

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

(a) **(8 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.

(b) **(7 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) **(10 points)** Now consider a *stream of n flips* from a coin whose *true* probability could be $p_{\text{true}} = 0.3$ or $p_{\text{true}} = 0.5$ (or any other value between 0 and 1). Implement the following steps in an R script, and produce plots for each p_{true}. Discuss any differences you observe when $p_{\text{true}} = 0.3$ versus $p_{\text{true}} = 0.5$.

(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 and plot how Posterior(θ) changes over time for each θ.

Problem 3 (30 points in total + 5 bonus points).

Suppose a model $Y = \beta_0 + \beta_1 X + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \sigma^2)$.

(a) (7 points) Prove that the minimizer of

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

(b) (8 points) Given a dataset

$$(x, y) \in \{(1, 0), (2, 3), (3, 5), (4, 6), (5, 8)\},$$

(i) Compute $\hat{\beta}_1$, $\text{SE}(\hat{\beta}_1)$, and a 95% confidence interval for β_1.

(ii) Compute the t-statistic for testing the null hypothesis $H_0 : \beta_1 = 0$ and decide whether to reject. How would you interpret the result?

(c) (15 points) Implement a short R script that:

(i) Generates $n = 100$ points via

$$Y = 2 + 3X + \varepsilon, \quad X \sim \text{Uniform}(0, 10), \quad \varepsilon \sim \mathcal{N}(0, 2^2).$$

(ii) Fits the linear model and reports $\hat{\beta}_0, \hat{\beta}_1$.

(iii) Computes the residual standard error (RSE) and compares it to the true $\sigma = 2$.

(iv) Constructs confidence intervals for β_0 and β_1.

(v) Performs a hypothesis test for $H : \beta_1 = 0$.

(vi) Computes R^2 and interprets it.

Report the results you obtain. Repeat for $n = 1000$ and discuss any differences you observe.

(d*) (5 bonus points) Repeat (**(c)**) but simulate $Y = 2 + 3X^2 + \varepsilon$. Fit a linear model anyway and compare results with the true quadratic form. Comment on the obtained results ($\hat{\beta}_0, \hat{\beta}_1$, R^2, etc.).

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

We collect $n = 100$ observations (x_i, y_i) and fit two regression models:

- A linear regression: $Y = \beta_0 + \beta_1 X + \varepsilon$.
- A cubic regression: $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \varepsilon$.

(a) **(5 points)** Suppose the *true* relationship between X and Y is actually linear. Compare the training RSS for the linear regression and for the cubic regression. Would we expect one to be lower than the other, or would they be about the same, or is there insufficient information to tell? Justify your answer.

(b) **(5 points)** Answer (**(a)**) again, but using the *test* RSS instead of the training RSS.

(c) **(10 points)** Implement a short R script to do the following:

(i) Generate data from $Y = 5 + 2X + \varepsilon$, where $X \sim \text{Uniform}(0, 10)$ and $\varepsilon \sim \mathcal{N}(0, 2^2)$. Fit both the linear and cubic models, and compare their training and test RSS.

(ii) Generate data from $Y = 0.5X^2 + 2\sin X + \varepsilon$, where $X \sim \text{Uniform}(0, 10)$ and $\varepsilon \sim \mathcal{N}(0, 2^2)$. Fit both models, compare the results, and discuss.

Finally, discuss any differences that emerge if you reduce the training sample size to $n = 10$.

(d*) **(5 bonus points)** Relate your findings from **(c)** to your answers in **(a)** and **(b)**. If the experimental results match your earlier reasoning, discuss why. If they do not match, speculate on possible reasons.