

# CSE 312 – Problem Set 8

Due Wednesday, June 3, 11:59pm

Spring 2026

## Instructions

**Solutions format and late policy.** See PSet 1 for details on expectations for solutions, collaboration policy, late policy, etc. The same requirements and policies still apply. Also follow the typesetting instructions from the prior PSets.

**Solutions submission.** You must submit your solution via Gradescope. In particular:

- Problem 1 will be done on Gradescope.
- Submit the solutions to problems 2-5 under “PSet 8 [Written]” as on previous problem sets. This will be a *single* PDF file containing the solution to problems 2-5 in the homework. Each numbered task should be solved on its own page (or pages). Follow the prompt on Gradescope to link tasks to your pages. Do not write your name on the individual pages – Gradescope will handle that.

## 1 Gradescope Questions (20 points)

Please do these [gradescope questions](#).

## 2 Collisions (15 points)

A total of  $m$  items are to be sequentially inserted into a hash table of size  $n$ , where each item is mapped independently, but the items are not mapped uniformly into the table. Instead, an item is mapped to cell  $j$  of the table with probability  $p_j$ , for  $j = 1, \dots, n$  (where  $\sum_{j=1}^n p_j = 1$ ). We say that a collision occurs whenever an item is mapped into a nonempty cell. Find  $\mathbb{E}[X]$ , where  $X$  is the number of collisions that occur during this process. You can leave your answer as an unsimplified sum.

Hint: Write  $X = X_1 + X_2 + \dots + X_m$  where  $X_i$  is an indicator random variable that is 1 if there is a collision when the  $i$ -th item is put into a cell and 0 otherwise; then use linearity of expectation. To compute  $\mathbb{E}[X_i]$ , use the law of total expectation, conditioning on which cell it is placed into.

## 3 Lazy Grader (12 points)

Prof. Lazy decides to assign final grades in CSE 312 by ignoring all the work the students have done and instead using the following probabilistic method: each student independently will be assigned an A with probability  $\theta$ , a B with probability  $3\theta$ , a C with probability  $\frac{2}{3}$ , and an F with probability

$\frac{1}{3} - 4\theta$ . When the quarter is over, you discover that only 10 students got an A, 35 got a B, 40 got a C, and 15 got an F.

Find the maximum likelihood estimate for the parameter  $\theta$  that Prof. Lazy used. Give an exact answer as a simplified fraction. You do not need to check second order conditions.

## 4 The Place That's Best (12 points)

Let  $y_1, y_2, \dots, y_n$  be i.i.d. samples of a random variable from the family of distributions  $Y(\theta)$  with densities

$$f(y; \theta) = \frac{1}{2\theta} \exp\left(-\frac{|y|}{\theta}\right),$$

where  $\theta > 0$ . Find the MLE for  $\theta$  in terms of the  $|y_i|$  values and  $n$ . You do not need to check second order conditions.

## 5 (Un)biased Estimation (10 points)

- a) Let  $X_1 = x_1, \dots, X_n = x_n$  be independent samples from the Poisson distribution with parameter  $\theta$ . In the section 9 worksheet, we have seen that the MLE estimator for  $\theta$  is  $\hat{\theta}(X_1 = x_1, \dots, X_n = x_n) = \frac{1}{n} \sum_{i=1}^n x_i$ . Thus, viewed as a random variable, the estimator is

$$\hat{\theta}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

Is this estimator unbiased?

- b) Let  $X_1 = x_1, \dots, X_n = x_n$  be independent samples from  $\text{Unif}(0, \theta)$ , the continuous uniform distribution on  $[0, \theta]$ . Consider the estimator

$$\hat{\theta}_{\text{first}}(X_1 = x_1, \dots, X_n = x_n) = 2x_1,$$

i.e., our estimator ignores the samples  $x_2, \dots, x_n$  and just outputs twice the value of the first sample. Is

$$\hat{\theta}_{\text{first}}(X_1, \dots, X_n) = 2X_1$$

unbiased?