Preparation

On the exam, you may be asked to do the following:

- Rephrase a key definition and/or theorem in your own words.
- Determine whether a given statement is true or false.
- Interpret or explain a statistics concept in everyday language.
- Sketch the proof of an important result discussed in class.
- Perform calculations using relevant techniques from the course.
- Provide a short, rigorous proof of a novel statement or result.
- Create and analyze a statistical model for a particular phenomenon.
- Use R to simulate a random phenomenon, plot densities, or visualize data.

There are some practice problems below. These are by no means exhaustive, nor are they necessarily representative of the expected length/difficulty of the midterm.

Practice Problems

1. Let $X_1, \ldots, X_n | \theta \overset{\text{iid}}{\sim} f(x|\theta)$ where for $x \in \mathbb{R}$,

$$f(x|\theta) = \frac{1}{2\theta} e^{-|x|/\theta}$$

   where θ is an unknown parameter with $\theta > 0$.

   (a) Find a formula for the MLE of θ.

   (b) Use R to simulate the sampling distribution for the MLE estimator and for the Method of Moments estimator when $\theta = 2$ and $n = 10$. (Note, you found a MoM estimator for θ on homework 4). Use your simulation to estimate the mean and variance of each estimator.

2. The method of *randomized response* is sometimes used to conduct surveys on sensitive topics. A simple version of the method can be described as follows:

   A random sample of n people are drawn from a large population. For each person in the sample, there is probability 1/2 that the person will be asked a standard question and probability 1/2 that the person will be asked a sensitive question. Furthermore, this selection of the standard or sensitive question is made independently from person to person. If a person is asked the standard question, then there is probability 1/2 that the person will give a positive response; however, if the person is asked the sensitive question, then there is an unknown probability p that they will give a positive response. The statistician can observe only the total number X of positive responses that were given by the n persons in the sample, but cannot observe which of these persons were asked the sensitive question or how many person in the sample were asked the sensitive question.

Determine the MLE of p based on the observation X.

3. Suppose that the random variable X has a binomial distribution with unknown value n and a known value of p for $0 < p < 1$. Determine the MLE of n based on the observation X. *Hint:* consider the ratio
$$\frac{f(x|n+1,p)}{f(x|n,p)}.$$

4. Suppose $X_1, \ldots, X_n|\boldsymbol{\theta} \overset{\text{iid}}{\sim} \text{Binom}(m,p)$ where $\boldsymbol{\theta} = (m,p)$ are unknown. Find a method of moments estimator for $\boldsymbol{\theta}$.

5. The *Pareto* distribution with shape θ and minimum value 1 has PDF:
$$f(x|\theta) = \begin{cases} \frac{\theta}{x^{\theta+1}} & x > 1 \\ 0 & o.w. \end{cases}$$

   (a) Show that the family of Gamma distributions Gamma(α, β) is a conjugate family of prior distributions for θ, when samples are taken from a Pareto distribution with shape θ and minimum value 1.

   (b) Suppose $X_1, \ldots, X_n|\theta \overset{\text{iid}}{\sim}$ Pareto with shape θ and minimum value 1, and let $\theta \sim \text{Gamma}(\alpha, \beta)$. What is the Bayes estimate for θ under squared loss?

6. The heights of individuals in a population have a normal distribution with unknown mean θ and standard deviation of 2 (variance of 4). Assume $\theta \sim N(68, 1)$. Suppose we sample 10 people from this population at random. Their average height is found to be 69.5 inches.

   (a) Find the Bayes estimate of θ under squared loss.

   (b) Find the Bayes estimate of θ under absolute loss.

7. Suppose $X|\theta \sim \text{Binom}(n, \theta)$ where the success probability $\theta \in [0, 1]$ is unknown. Suppose we are interested in estimating $g(\theta) = \Pr(X = 1)$, the probability of a single success in this Binomial observation. I propose the estimator $\delta(X) = \mathbf{1}_{\{X=1\}}$.

   (a) Is my estimator $\delta(X)$ an unbiased estimator for the quantity of interest?

   (b) Find $MSE_{g(\theta)}(\delta(X))$.

8. Suppose you are assessing the rate of radioactive decay for three different isotopes of an unknown element. The amount of time X_1 between particle emissions for the **first** isotope is exponential with mean θ, the amount of time X_2 between particle emissions for the **second** isotope is exponential with mean 2λ, and the amount of time

X_3 between particle emissions for the **third** isotope is exponential with mean $3\alpha\lambda$. Assume that $\alpha > 1, \lambda > 0$, and that given the values of α and λ, the amount of time between particle decays are independent for the three isotopes. Let $\boldsymbol{\theta} = (\lambda, \alpha)$ be the unknown parameters.

(a) Give an explicit description of the statistical model by identifying the joint probability distribution of the data and by identifying the parameters and the parameter space.

(b) Calculate the MLE of $\boldsymbol{\theta}$ based on the observations $X_1 = x_1$, $X_2 = x_2$ and $X_3 = x_3$. *You technically should justify that the critical points of the likelihood function correspond to a local maximum, but in this problem the Hessian is not beautiful so don't worry about it.*

9. Section 7.6 Problem 9

10. Section 8.8 Problem 19

11. Section 8.9 Problem 19