Problem Sessions

Chapter 1:

1. Textbook Exercise 36
2. Textbook Exercise 74
3. Use the appropriate measure of spread to identify if there are any outliers in the following dataset: 4,7,8,8,10,11,15
4. Calculate the standard deviation of the following dataset by hand: 3,6,9,10

Chapter 2:

1. Textbook Exercise 58
2. Set Theory – Show that if $A_1, A_2, \dots, A_n$ are equally likely independent events with probability p, then find $P(\cap_{i=1}^{n} A_i | A_2 \cup A_3)$
3. Counting Techniques – Consider the following scenario. Austin is hosting party this weekend and recently went to Big Y to pick up some things to put out for everyone. While he was there, he sat in awe as he looked at the copious number of types of chips there are. Being the math nerd he is, he decided to count all of them and say there are 31 different types of chips to purchase from Big Y. However, he noticed that a lot are the same flavor, just different brands. There were 7 Lime, 9 Salt, 4 BBQ chips, and 11 different types of pretzels.
    a. A decent party has 5 types of chips. What is the total number of ways to choose 5 chips randomly regardless of the flavor.
    b. What is the probability that all 5 randomly selected bags are Lime?
    c. No one *actually* likes pretzels, so Austin definitely doesn't want those. Instead he tries to be trendy and get 2 bags of Lime, 2 bags of BBQ, and one Salt. How many ways can he choose this out of all 31 types of chips?
    d. What is the probability of randomly selecting 2 bags of Lime, 2 bags of BBQ, and one Salt?
    e. What is the probability that when randomly selecting 4 bags, all 4 bags are different flavors?
    f. What is the probability that when randomly selecting 3 bags, all 3 bags are of the same flavor?
4. Bayes' Rule – Polygraph tests (lie detector tests) are often routinely administered to employees or prospective employees in sensitive positions (or every time on Criminal Minds…). According to studies of polygraph reliability (Gastwirth, J., 1987. The Statistical precision of medical screening procedures, Statistical Science, 3, 213-222) if a person is lying, the probability that this is detected by the polygraph is .88, whereas if the person is telling the truth, the polygraph is indicating that he/she is telling the truth with probability .86. Now, suppose that on a particular question the vast majority of subjects have no reason to lie, so that 99% tell the truth.
    a. Write down the 3 pieces of given information in terms of the probability of events A (the polygraph reading is positive – meaning the machine thinks the person is lying) and B (the subject is telling the truth).
    b. Draw a probability tree for this problem.
    c. What is the probability that the polygraph produces a positive reading?

d. Given that a subject produces a positive reading on the polygraph, what is the probability that the polygraph is incorrect and that he/she is telling the truth?

Chapter 3:

0. *Review of Poisson* – Proofs of Valid pdf, Expectation, & Variance
1. *Discrete RV's & Expectation* - Consider the following probability distribution functions for the RV's X and Y:

X	-1	0	1	2	3
Y	0	2	4	6	8
P(X)	.42	.2	.24	.04	.1

   a. Is this a valid pdf for the RV X? How about Y?
   b. Find the $E[X^3 + 5X - 1]$
   c. Using the same logic from part (b), how do you think you would find $E[XY + 3Y]$?
2. *Known RV Distribution #1* - Consider the situation where we have 20 marbles in an urn (7 yellow, 6 red, 4 white, and 3 pink) and we are taking a sample of size 4 without replacement.
   a. Does the RV X, the number of yellow marbles in your sample, follow a distribution that we studied? If so, properly label the distribution.
   b. Prove that the distribution is valid.
   c. What is the expected number of yellow marbles from your sample?
   d. Suppose we instead sample with replacement until we get the first pink marble. Derive the expectation for this new distribution. What is the expected number of trials until the first pink marble? What is the expected number of trials until the 3rd pink marble (no derivation necessary)?
3. *Known RV Distribution #2* - Consider the situation where we are looking at statistics PhD programs. The year that I applied for PhD programs, Harvard accepted 6 students with funding. When I got my rejection letter, it said that close to 1000 people applied for the statistics PhD program at Harvard that year. Harvard's funding is quite uniform, in that they accept the same number of statistics PhD students each year. So, it's safe to say that the average number of accepted students is 6.
   a. Does the RV Y, the number of accepted students in the Harvard statistics PhD program, follow a distribution that we studied? If so, properly label the distribution.
   b. Prove that the distribution is valid.
   c. What is $E[Y(Y-1)(Y-2)]$?
   d. Suppose my friend Jenna applies this year. What is the probability that she gets accepted? Think of this as a distribution we covered. Derive the expectation of this distribution.

Chapter 4:

1. *Continuous Random Variables & Probability Plots* – Create a Normal probability plot for the following data: 0.70, 1.23, 1.98, 2.87, 3.84, 4.13, 5.89, 8.42
   What are your conclusions?
2. *Continuous Random Variables* – Find the 83rd percentile of a chi-square distribution with 2 degrees of freedom.
3. *Continuous Random Variables* – Find the $E(X^7)$ if $X \sim Gamma(\alpha, \beta)$.
4. *Continuous Random Variables* – Find the k such that the following is a valid pdf:
   $$f(x) = \begin{cases} \frac{k}{\sqrt{\pi}\sqrt{x}} e^{-k^2 x} & 0 \leq x \\ 0 & otherwise \end{cases}$$
5. *Continuous Random Variables* – Prove that for a continuous random variable X, the $P(X \leq x) = P(X < x)$.
6. *Continuous Random Variables* – Derive the expectation of the generalized uniform distribution (generalized means that we do not know the actual numerical values of the parameters).
7. *Discrete Random Variables* – Consider the scenario where we are looking at school acceptance rates for a particular high school with 406 seniors. The UConn acceptance rate last cycle was 53.2%. What is the *approximate* probability that at least 250 of the high school seniors get into UConn? How would my calculation change if instead we only looked at 15 of the seniors?

Chapter 6:

0. Maximum Likelihood Estimation
   a. Likelihood function $\left(L(\vec{\theta})\right)$
   b. If $X \sim Bin(n, \theta)$ with n known, find $\hat{\theta}_{MLE}$.
   c. Suppose $X_1, \dots, X_n$ are iid $Bin(n, \theta)$ with n known, find $\hat{\theta}_{MLE}$
   d. The Invariance Principle for MLE's $\left(\widehat{g(\theta)_{MLE}} = g(\hat{\theta}_{MLE})\right)$ (the function doesn't even have to be one-to-one!)
      i. Find the MLE for $\theta^2 - 5\theta + 2$ from part (c)
1. *Point Estimation General Concepts* – If $X_1, X_2, \dots, X_N$ are i.i.d. with common distribution $X \sim Gamma(\alpha, \beta)$ (α known), determine whether the estimator $\hat{\beta} = \frac{1}{\alpha N} \sum_{i=1}^{N} x_i$ is unbiased. What is the precision of this estimator? What is the MSE?
2. *Point Estimation General Concepts* – Consider $X_1, X_2, \dots, X_N$ i.i.d. with common distribution $X \sim U(-2\theta, \theta + 1)$
   a. Is $\hat{\theta} = X_{(n)} - X_{(1)}$ an unbiased estimator for θ?
   b. Can we find a better estimator for θ?
   c. Consider the estimator $\hat{\theta} = \left(X_{(3)} - 4\right)^2$. Is this estimator unbiased? If not, quantify the bias. How will the MSE of this estimator compare to the MSE of your estimator from part (b)?

d. Consider a new situation $X_1, X_2, \ldots, X_N$ i.i.d. with common distribution $X \sim U(3, \theta)$. Do you think intuitively that the estimator $\hat{\theta} = X_{(3)}$ would be any good when N = 56?

3. *Method of Moments*
   a. Prove the moment estimators for $Gamma(\alpha, \beta)$ are $\hat{\alpha} = \frac{\bar{X}^2}{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(X_i - \bar{X})^2}$ and $\hat{\beta} = \frac{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(X_i - \bar{X})^2}{\bar{X}}$
   b. Suppose that $X_1, \ldots, X_n$ are iid with common pdf $f(x) = \begin{cases} (\theta+1)x^{\theta}, & 0 < x < 1 \\ 0 & , o.w. \end{cases}$ where $\theta (> 0)$ is the unknown parameter. Derive an estimator for θ by the method of moments.
4. *Maximum Likelihood Estimation* - Suppose $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$
   a. Show that $\hat{\mu}_{MLE} = \bar{x}$ and $\hat{\sigma}^2{}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
   b. Find the MLE for σ

Chapter 7:

1. *Confidence Intervals – Conceptual*
   a. Every confidence interval for the true value of a parameter is structured as (point estimate ± (critical value)(standard error of the estimate)). Use this information to derive the confidence interval of the true population proportion p when the population is approximately normally distributed.
   b. Find the appropriate value:
      i. The critical value for the two-sided 90% C.I. for the true population mean for a sample of size 12 and unknown σ.
      ii. The confidence level of the one-sided confidence upper bound $\bar{x} + 1.08\frac{\sigma}{\sqrt{n}}$
      iii. The central 95% confidence interval bounds (both upper and lower) on a χ_9^2 distribution.
   c. What is the ideal (ideal in terms of accuracy) margin of error? How do we achieve this?
   d. Visually interpret a 90% confidence interval.
2. *Confidence Interval – An Application* – In an effort to measure the average height of students here at UConn, researchers took a random sample of 632 UConn students and got an average height of 69 inches, give or take 3 inches. How would you estimate this using a point estimate? How about a 90% confidence interval? Interpret the interval, ***in context***, as if you are publishing the result yourself in a journal.
3. *Confidence Interval – An Application* – In an effort to measure average MATH SAT scores at UConn, researchers took a random sample of 124 UConn students and got an average MATH SAT score of 650.
   a. Suppose the distribution of MATH SAT scores is known to come from a Normal distribution with standard deviation 20. Use a 85% confidence interval to

estimate the true average MATH SAT score here at UConn. What sample size would be required in this context for a desired margin of error of 7%?

b. Suppose now a more realistic case where we know the distribution of MATH SAT scores comes from a Normal distribution with unknown variance but sample standard deviation of 18. Estimate the true standard deviation of MATH SAT scores using a 90% C.I.

Chapter 8:

1. *Hypothesis Testing – Conceptual*
    a. Lay out the basic structure of every hypothesis test (the 4 parts)
    b. What is the fundamental assumption made during hypothesis testing?
    c. Why can we never conclude that H_0 is true?
    d. Define what a p-value is.
    e. What is the difference between a one-sided and two-sided alternative hypothesis?
    f. What is the relationship between the power of a test and type one error?
    g. What is one way in which p-values can be "abused."
    h. What are the α and β of an ideal hypothesis test? What is the thought process in choosing the "best" hypothesis test in terms of α and β?
2. *Type I and Type II Error* – Consider a population with the pdf $N(\theta, 1)$ where $\theta \in \mathbb{R}$ is unknown. An experimenter wishes to test $H_0: \theta = 5.5$ vs $H_1: \theta = 8$ by collecting a random sample of $\vec{X} = (X_1, X_2, \dots, X_9)$ and is debating which test to use of the following:
        i. Reject H_0 iff $X_1 > 7$
        ii. Reject H_0 iff $\frac{1}{2}(X_1 + X_2) > 7$
        iii. Reject H_0 iff $\bar{X} > 6$
    a. Calculate alpha and beta for tests 1,2, and 3.
    b. Which test should you use, test 1 or test 3?
    c. Which test should you use, test 1 or test 2?
    d. What is the power of tests 1 and 3? What does this tell you?
3. *Hypothesis Testing – An Application* - Textbook Chapter 8, Exercise 26 (pg 322) – except follow these steps:
    a. What is the appropriate test to use here? Justify your claim by confirming the assumptions are met.
    b. What are the null and alternative hypotheses?
    c. Calculate the test stat and the reject region for alpha significance level of .05. What is your conclusion?
    d. Calculate the p-value. What is your conclusion at the alpha significance level of .05?
    e. What is a Type I error in context in this case? What is a Type II error in context in this case?
    f. How does the test change if instead of 45 specimens they only looked at 21?

4. *Hypothesis Testing – An Application* – Textbook Chapter 8, Exercise 39 (pg 327) – except follow these steps:
    a. What is the appropriate test to use here? Justify your claim by confirming the assumptions are met.
    b. What are the null and alternative hypotheses?
    c. Calculate the test stat and the reject region. What is your conclusion at the alpha significance level of .05?
    d. Calculate the p-value. What is your conclusion at the alpha significance level of .05? How does your answer change if we set the alpha significance level at .01?
    e. What is a Type I error in context in this case? What is a Type II error in context in this case?