CS3130: Probability and Statistics for Engineers **Practice**

Practice Final Exam

Instructions: On the final you will be allowed 1 sheet (front and back) that is hand-written or digitally generated for this purpose. You will turn this in with your exam.

Show all of you work in the space provided for each question. If you run out of space, you may attached another sheet of scratch paper. Attempt to simplify all calculations (seeing the final numbers will be useful for sanity check on answers), but we will try to forgive simple computational mistakes.

1. **(20 pts)** Basic Probability
   A game includes a fair 8-sided die with sides labeled from $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$.
   Three key events occur:
   $A = \{3, 6\}$: a multiple of 3
   $B = \{2, 3, 5, 7\}$: a prime
   $C = \{5, 6, 7, 8\}$: the larger values.

   (a) **(5 pts)** What does $Pr(A \mid B)$ mean in English? What is its value?

   (b) **(5 pts)** Are C and B independent? Explain why or why not.

   (c) **(10 pts)** Given your answer (a) above, and that $Pr(A) = 1/4$ and $Pr(B) = 1/2$, write down Bayes' Rule, and use it to solve for $Pr(B \mid A)$.

2. **(20 pts)** Random Variables

The Utah softball team has an on-base-percentage of 0.4 (the probability each batter reaches base – we assume each at-bat has this rate independently). Consider a point in a game where each of 9 batters in the starting lineup has gotten to bat 3 times (so a total of 27 at-bats).

(a) **(5 pts)** What distribution models the above scenario of the total number of times batter gets "on base"? Write out its name, and the values of its parameters.

(b) **(5 pts)** What is the expected value of the number of times a batter reaches base?

(c) **(5 pts)** Write out the probability (in terms of the above parameters) that the first time through the line-up (the first 9 at bats) that no batter reaches base.

(d) **(5 pts)** What distribution models the number of at bats until the first batter reaches base? Write out its name, and the values of its parameters.

3. **(20 pts)** Expectation and Variance
   Two kids go to the library and choose some books. Xander's number of books is described by a random variable X, with expectation $E[X] = 5$ and variance $Var[X] = 4$. Yolonda's number of books is described by random variable Y, with expectation $E[8]$.

   (a) **(5 pts)** If Xander goes to the library twice, and Yolonda goes to the library three times. What is the expectation of the total number of books they get.

   (b) **(5 pts)** Considering just Xander, what is the variance of the number of books they get on those two trips?

   (c) **(3 pts)** What is the standard deviation of the number of books Xander gets on those two trips?

   (d) **(7 pts)** Calculate the variance of the number of books Yolonda gets on one trip. Use the discrete distributions of books Yolonda gets k is described by the following probability table:

k	7	8	9
Pr(Y=k)	0.25	0.5	0.25

4. **(20 pts)** Joint and Conditional Probabilities
   Consider the joint probability table where random variable X can take values 3 or 4, and random variable Y can take on values 1 or 2.

X \ Y	1	2
3	0.2	0.25
4	0.25	0.3

   (a) **(5 pts)** What is the probability $Pr(X = 3, Y = 2)$?

   (b) **(5 pts)** What is the conditional probability $P(X = 4 \mid Y = 1)$

   (c) **(5 pts)** Write the marginal probability distribution for Y as a table.

   (d) **(5 pts)** Are random variables X and Y independent? Explain why or why not.

5. **(20 pts)** Statistics and Sampling
   Consider n iid samples $X_1, \ldots, X_n \sim Unif(0,1)$. So $E[X_i] = 0.5$ and $Var[X_i] = 1/12$.
   Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

   (a) **(5 pts)** What does <u>iid</u> stand for?

   (b) **(5 pts)** What is the expected value of $\bar{X}_n$?

   (c) **(5 pts)** What is variance of $\bar{X}_n$?

   (d) **(5 pts)** As n gets larger, does the standard deviation of $\bar{X}_n$ get larger or smaller?

6. **(20 pts)** Estimation, Bias, and Variance

Consider an iid random sample $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ and let $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Let $\mu = 0$ and $\sigma^2 = 1$, but we do not assume to know this.

(a) **(5 pts)** Let $\hat{\mu}_1 = \bar{X}_n + 5/n$. Is $\hat{\mu}_1$ an unbiased estimator for μ? Explain why or why not.

(b) **(5 pts)** Let $\hat{\mu}_2 = X_1$. Is $\hat{\mu}_2$ an unbiased estimator for μ? Explain why or why not.

(c) **(5 pts)** The efficiency of an estimator is determined by its variance, the smaller the better. In the limit as n goes to ∞, what is the variance of $\hat{\mu}_1$.

(d) **(5 pts)** In the limit as n goes to ∞, which is a more efficient estimator $\hat{\mu}_1$ or $\hat{\mu}_2$? Explain why.

7. **(20 pts)** Confidence Intervals
   Consider a random sample $X_1, \ldots, X_n$ of $n = 100$ random variables from distribution $f(\mu)$ where $E[X_i] = \mu$ and known variance $\text{Var}[X_i] = 4$.

   (a) **(5 pts)** Write down an unbiased estimator $\hat{\mu}$ of μ that uses all 100 random variables [it should be highly efficient and approach a normal distribution in the limit].

   (b) **(5 pts)** What is the standard deviation of $\hat{\mu}$?

   (c) **(10 pts)** Use a normal approximation of $\hat{\mu}$ and one the following values to write down the 99% confidence interval for $\hat{\mu}$.

   $z_{0.05} = 1.645$ $\qquad$ $z_{0.025} = 1.96$ $\qquad$ $z_{0.01} = 2.32$ $\qquad$ $z_{0.005} = 2.57$

8. **(20 pts)** Hypothesis Testing
   Pro Golfer Tony Finau is worried the average length of his drive is not as long as it used to be (300 yards). Tony asks you to help investigate.

   (a) **(5 pts)** Design a set of hypothesis. What is the null H_0 and alternative H_1 hypothesis.

   (b) **(10 pts)** Tony agrees that tomorrow he will go to the driving range, and hit $n = 100$ golf balls with his driver, and then you measure the distance of each: $X_1, \ldots, X_n$. Design an experiment so if it succeeds, you can be 95% confident his driving distance has decreased. Your answer should come up with a
   (i) test statistic T
   (ii) model it with a distribution [name the distribution]
   (iii) choose a critical value at some parameter α.

   (c) **(5 pts)** Assume that finally you get to go to the range and his average drive length is 301 yards. Do you reject the null hypothesis? Explain why or why not.