CSE 312: Foundations of Computing II
Practice Final

Time Limit: 3.5 hours

#	Topic	Possible	Earned
1	Random True/False, Short Answer	30	
2	Continuous Random Variables [+10 EC]	20	
3	Maximum Likelihood Estimation	40	
4	Counting	25	
5	Approximations	20	
6	Discrete Random Variables	15	
7	Conditional Probability	25	
8	The Poisson Distribution	35	
9	Expectation Lemma [+10 EC]	0	
10	Random Random Variables [+10 EC]	0	
	Total	210	

You do not have to do any of the Extra Credit problems 2c, 9, and 10 in order to have a good practice final.

1. [30 points] True/False, Short Answer. Provide a short justification for your answer.

a) [2 points] True or False. For any μ and σ^2, if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma^2} \sim N(0,1)$.

b) [2 points] True or False. For any μ and σ^2, if $X \sim N(\mu, \sigma^2)$ has pdf $f_X(x)$, then $\int_0^\infty f_X(x)dx = 1/2$.

c) [2 points] Let X have a continuous uniform distribution, $X \sim Unif(a, b)$. What is $E[X^3]$, the skewness of X? Do not simplify further than evaluating the integral.

d) [2 points] True or False. Suppose X has a continuous uniform distribution, $Unif(0,1)$. If $Y = -\ln X$, then $Y \sim Exp(1)$.

e) [2 points] True or False. For any random variable X and any $\alpha > 0$, $P(X \geq \alpha) \leq \frac{E[X]}{\alpha}$.

f) [2 points] True or False. If $Z \sim N(0,1)$, then $P(Z \geq -z) = \Phi(z)$.

g) [2 points] True or False. For any random variable X, $P(X \leq x) = P(X < x)$.

h) [2 points] True or False. If $X_1, X_2, \ldots, X_n$ are iid, each with standard deviation σ, then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ has standard deviation $\sigma_{\bar{X}} = \sigma/n$.

i) [2 points] Suppose $Z \sim N(0,1)$. Find $Var(Var(Z^3))$.

j) [2 points] Suppose $X \sim N(\mu_1, \sigma_1^2)$. If we want to apply a linear transformation f to X to get $X' = f(X) \sim N(\mu_2, \sigma_2^2)$, what should $f(X)$ be?

k) [2 points] True or False. Suppose X is a continuous random variable with pdf $f_X(x)$. Then, $0 \leq f_X(x) \leq 1 \; \forall x \in \mathbb{R}$.

l) [2 points] True or False. Suppose X is a discrete random variable with pmf $p_X(x)$. Then, $0 \leq p_X(x) \leq 1 \; \forall x \in \mathbb{R}$.

m) [2 points] True or False. Suppose X is a continuous random variable with cdf $F_X(x)$. Then, $0 \leq F_X(x) \leq 1 \; \forall x \in \mathbb{R}$.

n) [2 points] True or False. Suppose X is a continuous random variable with cdf $F_X(x)$. Then, $F_X(x)$ is a strictly increasing function. (That is, $\forall a, b \in \mathbb{R}, a < b \rightarrow F_X(b) > F_X(a)$).

o) [2 points] True or False. Let $A_1, A_2, \ldots, A_n$ be events such that $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$ and $A_1 \cap A_2 \cap \ldots \cap A_n = \emptyset$, and let B be any event. Then,

$$P(B) = P(A_1 \cap B) + \cdots + P(A_n \cap B)$$

2. [20 points + 10 points Extra Credit] Alex decided he wanted to create a "new" type of distribution that will be famous, but he needs some help. He knows he wants it to be continuous and have uniform density, but he needs help working out some of the details. We'll denote a random variable X having the "Uniform-2" distribution as $X \sim Unif2(a, b, c, d)$, where $a < b < c < d$. We want the density to be non-zero in $[a, b]$ and $[c, d]$, and zero everywhere else. Anywhere the density is non-zero, it must be equal to the same constant.

a) [5 points] Find the probability density function, $f_X(x)$. Be sure to specify the values it takes on for every point in $(-\infty, \infty)$. (Hint: use a piece-wise definition).

b) [15 points] Find the cumulative distribution function, $F_X(x)$. Be sure to specify the values it takes on for every point in $(-\infty, \infty)$. (Hint: use a piece-wise definition).

c) [10 points **Extra Credit**] Suppose $x_1, x_2, \ldots, x_n$ are iid from $Unif2(a, b, c, d)$. For simplicity, you may assume $n \geq 4$. Find the maximum likelihood estimators for a, b, c, and d. (Hint: Do not take any derivatives: think about the likelihood function and how you would maximize it).

3. [40 points] Let's say a random variable $X \sim Exp2(\mu), \mu > 0$ if it has the following density:

$$f_X(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu}, & x \geq 0 \\ 0, & otherwise \end{cases}$$

You are given that $E[X] = \mu$ and $Var(X) = \mu^2$.

a) [10 points] Suppose we have iid samples, $x_1, \ldots, x_n \sim Exp2(\mu)$. Show that the maximum likelihood estimator for μ is $\widehat{\mu_n} = \frac{1}{n}\sum_{i=1}^{n} x_i,$. You need not verify that your estimator is a maximum.

b) [6 points] Let's consider our estimator from part a). Is your maximum likelihood estimator unbiased? Prove it.

c) [5 points] MLE's have a lot of nice properties, out of the scope of this class. Unfortunately, unbiasedness is not one of them. Give an example of a distribution and a parameter for that distribution where the maximum likelihood estimator is biased. (Hint: You've seen at least two examples in class or on homework).

d) [6 points] Find $Var(\widehat{\mu_n})$.

e) [3 points] The mean squared error of an estimator is defined as $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2(\hat{\theta})$. It measures how much an estimator deviates from the true parameter, whereas the variance of the estimator just measures how much the estimator deviates from its own expectation. The bias of an estimator is determined by $Bias(\hat{\theta}) = E[\hat{\theta}] - \theta$. Find $MSE(\widehat{\mu_n})$.

f) [10 points] We say an estimator $\widehat{\theta_n}$ is consistent for θ if it converges in probability to θ. That is,

$$\forall \varepsilon > 0, \lim_{n \to \infty} P\left(\left|\widehat{\theta_n} - \theta\right| \geq \varepsilon\right) = 0$$

In English, this means that our estimator approaches the true value θ as we take more and more samples (the estimator is accurate). Although MLE's may be biased, MLE's are guaranteed to be consistent (which implies asymptotic unbiasedness, or "unbiased in the limit"). Use Chebyshev's Inequality to show that the MLE for $Exp2(\mu)$, $\widehat{\mu_n}$, is consistent for μ. We have started the proof below.

<u>Proof:</u>

Fix $\varepsilon > 0$. By Chebyshev's inequality, $0 \leq P(|\widehat{\mu_n} - \mu| \geq \varepsilon) \leq$ ______________ = _________.

As ___ $\to \infty$, _________ $\to 0$, so we have ______ $\leq P(|\widehat{\mu_n} - \mu| \geq \varepsilon) \leq$ _______.

Therefore, by properties of limits, $\lim_{n \to \infty} P(|\widehat{\mu_n} - \mu| \geq \varepsilon) =$ _________.

So our estimator $\widehat{\mu_n}$ is consistent.

Q.E.D.

4. [25 points]

a) [21 points] Suppose we have the inequality, $a_1 + a_2 + \cdots + a_n \leq k$, where $k \geq n$. What is the number of solutions to this inequality, with the constraint that each a_i must be a positive integer and that k is also an integer? (Hint 1: Modify the inequality to get an equivalent one in which each a_i is constrained to be nonnegative instead of positive). (Hint 2: Introduce a new term a_{n+1} which will take the "remainder").

b) [4 points] Suppose we have the same setup as in part a), but we flipped the direction of our constraint inequality to be $k < n$. How many solutions are there now with this reversed constraint?

5. [20 points] Every week, $20{,}000$ students roll a $10{,}000$-sided fair die, numbered 1 to $10{,}000$, to see if they can get their GPA changed to a 4.0. If they roll a 1, they win (they get their GPA changed). You may assume each student's roll is independent. Let X be the number of students who win.

a) [3 points] For any given week, give the appropriate probability distribution (including parameter(s)), and find the expected number of students who win.

b) [4 points] For any given week, find the exact probability that at least 2 students win. Give your answer to 5 decimal places.

c) [5 points] For any given week, estimate the probability that at least 2 students win, using the Poisson approximation. Give your answer to 5 decimal places.

d) [8 points] For any given week, estimate the probability that at least 2 students win, using the Normal approximation. Give your answer to 4 decimal places.

6. [15 points] Suppose X has the following probability mass function:

$$p_X(x) = \begin{cases} c, & x = 0 \\ 2c, & x = \frac{\pi}{2} \\ c, & x = \pi \\ 0, & otherwise \end{cases}$$

Trig Reference

Angle	**Sine**	**Cosine**
0	0	1
$\pi/2$	1	0
π	0	-1

a) [3 points] Suppose $Y_1 = \sin(X)$. Find $E[Y_1^2]$.

b) [3 points] Suppose $Y_2 = \cos(X)$. Find $E[Y_2^2]$.

c) [2 points] Suppose $Y = Y_1^2 + Y_2^2 = \sin^2(X) + \cos^2(X)$. Before any calculation, what do you think $E[Y]$ should be? Find $E[Y]$, and see if your hypothesis was correct. (Recall for any real number x, $\sin^2(x) + \cos^2(x) = 1$).

d) [7 points] Let W be any discrete random variable with probability mass function $p_W(w)$. Then, $E[\sin^2(W) + \cos^2(W)] = 1$. Is this statement always true? If so, prove it. If it is false, give a counterexample by giving a probability mass function for a discrete random variable W for which the statement is false.

7. [25 points] Suppose any given CSE 312 student is three times more likely to come to class regularly than not. A student who doesn’t come to class regularly is twice as likely to have failuritis as a student who does.

a) [21 points] What is the probability a student doesn’t come to class regularly, given that they have failuritis?

b) [4 points] What is the probability a student comes to class regularly, given that they have failuritis?

8. [35 points]

a) [20 points] Let $X \sim Poi(\lambda_1)$ and $Y \sim Poi(\lambda_2)$, where X and Y are independent. Show that their sum $Z = X + Y$ is $Poi(\lambda_1 + \lambda_2)$. (Hint 1: Start with the pmf for Z, $P(Z = n)$). (Hint 2: Use the binomial theorem).

b) [5 points] Suppose $X_1, X_2, \ldots, X_n$ are iid $\sim Poi(\lambda)$. Their sum $X = X_1 + \cdots + X_n$ is Poisson with what rate parameter λ'?

c) [10 points] Let $X \sim Poi(\lambda_1)$ and $Y \sim Poi(\lambda_2)$, where X and Y are independent. In part a), we showed that $Z = X + Y$ is $Poi(\lambda_1 + \lambda_2)$. Prove that $P(X = k|Z = n)$ is the probability mass function for a binomial random variable, and specify its parameters. You may use any results from previous parts. We have started the proof below. (Hint: If $X = k$ and $Z = n$, what does that say about Y?).

$$P(X = k|Z = n) = \frac{P(X = k \cap Z = n)}{P(Z = n)}$$

You can stop here if you like. The remaining problems are Extra Credit.

9. [10 points **Extra Credit**] Recall that for a continuous random variable Y, we have that

$$E[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy$$

Suppose Y is a non-negative continuous random variable. Show that $E[Y] = \int_0^{\infty} (1 - F_Y(y)) dy$, where $F'_Y(y) = f_Y(y)$. (Hint: Use double integrals).

10. [10 points **Extra Credit**] Let P be a continuous uniform random variable distributed on $[a, b]$, where $0 < a < b < 1$. Let $N \sim Poi(\lambda)$. Suppose $X_1, \dots, X_N$ are iid $Ber(P)$. So we have N iid Bernoulli random variables, each with success parameter P, where both N and P are random variables themselves! Let

$$X = \sum_{i=1}^{N} X_i$$

Find $E[X]$ in terms of a, b, and λ. (Hint: Use the Law of Total Expectation).