Math 30440 — Probability and Statistics

Spring 2010 Final, May 4 2010

Instructors: David Galvin and Daniel Cibotaru

Name: ______________________________

This examination contains 8 problems on 12 pages (including the front cover and three pages of tables at the back). It is closed-book. You may use limited notes. You may use a calculator. **Show all your work** on the paper provided. The honor code is in effect for this examination.

Scores

Question	Score	Out of
1		10
2		10
3		10
4		10
5		10
6		10
7		10
8		10
Total		80

GOOD LUCK !!!

1. (a) My dog Casey suffers from epilepsy, and has a seizure on average twice per month. Assuming that the number of seizures per month is a Poisson random variable, calculate the probability that Casey has at least one seizure in a one month period.

   (b) Casey's vet has prescribed Phenobarbital, a drug which is effective in completely eliminating seizures for 80% of all dogs (for the remaining 20%, the drug has no effect). Casey has been on the drug now for one month, and has not had a seizure in that time. What is the probability that the drug is effective for Casey?

2. Let X be a random variable with the following cumulative distribution function:

$$F(x) = \begin{cases} 0 & \text{if } x < 3.5 \\ x - 3.5 & \text{if } 3.5 \leq x \leq 4.5 \\ 1 & \text{if } x > 4.5. \end{cases}$$

(a) Calculate the expectation of X.

(b) Calculate the variance of X.

(c) X is intended to model the arrival time, somewhere between 3.30pm and 4.30pm, of a daily bus. (So, for example, $X = 4.1$ represents an arrival time of 4.06pm, since .1 of an hour is 6 minutes). Approximate the probability that if the arrival time of the bus is averaged over the course of 50 days (assumed independent), the answer will some time later than 4.03pm.

3. A large array of machines has 3840 generators. The probability that any one of them will fail during the course of a year is $1/1200$.

    (a) Assuming that the generators are independent of each other, write down the expectation and variance of X, the number of generators that fail over the course of a year.

    (b) Write down an expression that calculates exactly the probability that fewer than 3 generators fail over the course of a year

    (c) Use a Poisson approximation to approximate the probability that fewer than 3 generators fail over the course of a year.

4. (a) The Rayleigh probability density function is sometimes used to model wind speed. Its density function is

$$f(x) = \begin{cases} \frac{x}{\theta} e^{-\frac{x^2}{\theta}} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

Independent readings $x_1, \ldots, x_n$ are taken from this density. Find the maximum likelihood estimator for θ.

(b) You draw 4 samples from the population that you suspect has a Rayleigh distribution, but you do not know θ. The four values you get are 1.8, 1.2, 3.3 and 2.7. What is the most likely value for θ?

5. The mean breaking strength $\bar{X}$ of a sample of $n = 32$ steel beams equals 42196 pounds per square inch (psi). From past experience, it is known that the standard deviation of the breaking strength is $\sigma = 500$psi. It is also known that breaking strengths are distributed approximately normally.

   (a) Write down a 90% two-sided confidence interval for μ, the mean breaking strength of the beams.

   (b) The manufacturer's specifications indicate that the steel beams have mean breaking strength at least 42500psi. Does your data provide enough evidence, at 5% significance, to reject this claim?

6. In a study of the amount of calcium in drinking water undertaken as part of water-quality assessment, the same sample was tested in the laboratory six times at random intervals. The six readings (in parts per million) were

$$9.5 \qquad 9.6 \qquad 9.3 \qquad 9.5 \qquad 9.7 \qquad 9.4$$

(a) Give a 90% confidence interval for σ^2, the variance for readings on this sample using this particular test (assuming that readings are normally distributed).

(b) Give a 95% confidence interval for μ, the mean for readings on this sample using this particular test.

7. You want to test a company's claim that two different dyes dry out in the same amount of time. You take samples from both and measure the drying time (in minutes). Assume that drying time for any dye is a normal distributed random variable. The results for the first dye are as follows:

$$12.3 \quad 11.4 \quad 13 \quad 16.5 \quad 14.4 \quad 13.3 \quad 13.9.$$

For the second dye, 8 samples had sample mean 12.875 and sample standard deviation 1.05.

(a) Suppose you know that for both dyes, the standard deviation is 1. Does your data present evidence that suggests that the company's claim is incorrect? Calculate the p-value for this test.

(b) Suppose that you do not know the standard deviations of the two drying times, but are willing to believe that they are equal. Construct a 99% confidence interval for the difference between the two drying times.

8. A study was made of 400 randomly chosen square meters of tropical woodland. The average biomass per square meter was calculated to be $\bar{X}$. Letting μ be the actual average biomass per meter square, we want to use this data to test $H_0 : \mu = 35$ against $H_1 : \mu < 35$, at 5% significance. We assume that the standard deviation of biomass per square meter is 10.

   (a) What is the range of values of $\bar{X}$ that would lead to acceptance of the null hypothesis for this test?

   (b) Suppose that in fact μ is 33 kg per square meter. What is the mean and variance of $\bar{X}$?

   (c) Again suppose that in fact μ is 33 kg per square meter. Compute the probability that $\bar{X}$ falls into the acceptance region computed in part (a).

   (d) What is the probability (for the particular test described in this question) that the null hypothesis will be rejected when in fact it is true?

   (e) What is the probability (for the particular test described in this question) that the null hypothesis will be rejected if the true mean is in fact 33?