Test-1 For Statistics 410

Here is the exam. If you have any questions, please e-mail. If you can't get this exam to print properly a PostScript copy of this exam is available from
http://www.math.csi.cuny.edu/ verzani/410/test-1.ps
(except for the figure.) There are 11 questions. Please do any 10. It is due on Tueday the 16th.

1. Answer the questions below based on this stem and leaf diagram of test scores from a Bio 123 class.

stem	leaves
10	
9	1 4 4 7
8	1 1 3 4 5 5 5 7 8
7	0 0 2 3 8 8 9
6	0 4 8
5	2 3 3
4	7
3	
2	8 8 9
1	

   a. Find the 5 number summary for the data.

   b. Make a box plot.

   c. Make a histogram.

   d. Which students deserve a failing grade?

2. Let X and Y be two random variables with joint density

$$f(x,y) = c\, x^2\, y^2,\ 0 \le x,y \le 1.$$

   a. Find c.

   b. Find the marginal density of X, $f_X(x)$.

   c. Are X and Y exchangeable? Why?

   d. Are X and Y independent? Why?

   e. Find E(X), var(X), E(XY).

3. In a certain area 25% of all cars emit excessive pollutants. At the yearly testing suppose the following is

true. The probability that a car emitting too many pollutants will fail is 99% and the probability a good car will falsely fail is 15%.

    a. What is the probability that a car that fails the exam actually emit too much pollution?

    b. Suppose it is known that cars that are 65% of cars 10 years old or older emit excessive pollutants. The other probabilities remain the same. Now if a 12 year old car fails the test, what is the probability it emits too much pollution?

4. Pick a point (X,Y) uniformly from the triangle formed by the three points (-1,0), (1,0), (1,1).

    a. Find the cumulative distribution function

$$F(x,y) = P(X \le x, Y \le y)$$

    b. Find the joint density $f(x,y)$.

    c. Find the marginal density of X.

    d. Show that the conditional density of X given $Y = y$ is uniform.

5. For the picture below, if T is assumed to be uniform on $[-\pi/2, \pi/2]$ show that X has a Cauchy distribution. It is enough to find $F(x) = P(X \le x)$.
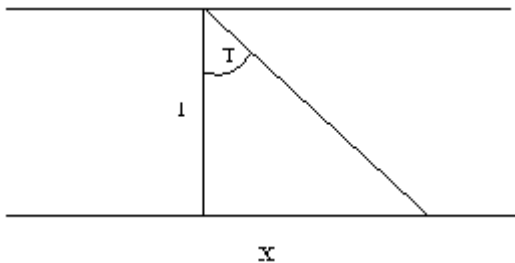
6.

    a. Verify the formula

$$\mathrm{var}(X_1 + \ldots + X_n) = \mathrm{var}(X_1) + \ldots + \mathrm{var}(X_n) + \sum_i \sum_{j \ne i} \mathrm{cov}(X_i, X_j)$$

    in the case $n = 2$.

    b. In $n = 2$ what happens if X_1 and X_2 are independent. Why?

    c. Compute $\mathrm{var}(X_1 - X_2)$ when X_1 and X_2 are independent.

7. Suppose X_i $1 \le i \le n$ are normal with mean μ_1 and standard deviation σ_1 and Y_j $1 \le j \le m$ are normal with mean μ_2 and standard deviation σ_2. Furthermore all n+m random variables are assumed to be independent. Set

$$Xbar = \bar{X} = \frac{X_1 + \ldots X_n}{n}, \quad Ybar = \frac{\bar{Y} = Y_1 + \ldots Y_m}{m}.$$

    a. Compute E(Xbar) and var(Xbar).

    b. Compute E(Xbar - Ybar), var(Xbar - Ybar),

8. In section 4.8 the multinomial distribution is discussed which generalizes the binomial distribution. The key formula is on page 150. Use this to answer the following.

   Let X_1, X_2, ... X_n be i.i.d. random variables which are $U(0,1)$ (uniform on $[0,1]$). What is the probability the kth largest is in the interval $[y,y+dy)$, k-1 are in the interval $[0,y)$ and n-k are in the interval $[y+dy,1]$?

9. Suppose n people from Staten Island were called at random via sampling with replacement and asked if they supported keeping the dump open beyond the current deadline. Only 22% said yes.

   a. In the notation from class, what is $[\hat{p}]$? (phat), and what does p represent?

   b. If n was 1000, what is an interval around $[\hat{p}]$ for which we are 95% certain p lies in.

   c. If we want to know 95% of the time that p is in the interval $[\hat{p}]$ -.01, $[\hat{p}]$ + .01) how large should n be?

   (*This question relies on class notes from Tuesday March 2. The basic formula that we derived was*

   $$P(|\hat{p} - p| < \frac{z}{z\sqrt{n}} \geq \Phi(z) - \Phi(-z)$$

   where Φ(z) is the c.d.f. for a normal(0,1) random variable.)

10. Suppose 25% of the population is a certain minority. However, in 1000 arrests by the police 33% were this minority type. How likely is this under an assumption that each arrest is independent, and each person has an equal chance of being arrested? Use the Normal approximation to give a numeric answer.

11. In trying to estimate population sizes a method of recapture is used. Let N be the size of a population. (For example spotted owls in a region of Oregon.) Researchers capture n_1 of the population and band them. These are released to the wild. After some time n_2 of the population are captured. Suppose X of the n_2 are banded. Suppose $n_1 = 3$ and $n_2 = 4$.

    a. What is the likelihood function for X given N, $L(N) = f(x \mid N)$.

    b. If X = 1 what value of N maximizes the likelihood function?

    (*This type of question is from section 8.2 which was in the lecture on March 4.*)