STATISTICS 200
FINAL
March 2006

Name:

Instructions:

1. Print your name.
2. Each of the problems is worth 10 points, except the first which is worth 20.
3. You must show your work to get full credit.
4. Do not turn the page until the signal is given.
5. Draw a box around your final answers.
6. Good luck!

Points:

Problem 1:

Problem 2:

Problem 3:

Problem 4:

Problem 5:

Total:

1. A thousand individuals sampled were classified according to sex and according to whether or not they were color-blind as follows: 442 were male and normal, 514 were female and normal, 38 were male and color-blind, and 6 were female and color-blind. According to the genetic model, these numbers should have relative frequencies given by: $p/2$, $pq + (p^2/2)$, $q/2$ and $q^2/2$, for some p and $q = 1 - p$.

(i). Does the model allow the two modes of classification to be independent, for any p?

(ii) Assuming the model is correct, compute the maximum likelihood estimator of p.

(iii) Are the data consistent with the model? (If you couldn't do (ii), suggest a reasonable estimator of p and continue as if it were the MLE.)

(iv) Suppose an investigator now collects data based on three attributes: sex (male of female), whether or not the person is color-blind, and handedness (dominant right hand or not). Each of the three attributes has two possibilities. If you wanted to test whether the three attributes are mutually independent, you can do a Chi-squared test. How many degrees of freedom should you use?

2. Suppose an investigator wants to compare a new drug designed to lower LDL cholesterol levels with a placebo. She conducts a well-designed double-blind study using 100 patients (50 receiving treatment and 50 control) in order to test the null hypothesis of no difference. She plans on calculating a p-value based on the Wilcoxon rank sum test at level $\alpha = 0.05$. Independently, the same study is done at a hospital in another part of the country, with another 100 patients recruited (50 receiving treatment and 50 control). Assume the drug is no better than the placebo.

(i) If each test is carried out at level $\alpha = 0.05$, what is the chance that at least one study will find a significant result?

(ii) In order to evaluate results from both studies, you decide you will reject the null hypothesis that there is no difference between the drug and the placebo if there is at least one p-value that is less than c. What should c be so that, if there really is no difference between the drug and placebo, the chance that you reject the null hypothesis of no difference is 5 percent?

3. Toss a coin with unknown probability p of landing heads. Toss it until you observe exactly 6 heads, and let X be the total number of tosses you made.

(i) What is $P\{X = j\}$ as a function of p and j?

(ii) For testing a null hypothesis $H_0 : p = 1/2$ versus the alternative hypothesis $H_1 : p > 1/2$, determine if there exists a uniformly most powerful test at level $\alpha = 1/16$. If it exists, find it. Would you reject if $X = 10$?

4. In n Bernoulli trials, each with success probability p, let X_1 denote the number of successes and X_2 the number of failures. For testing $p = 1/2$, the usual Chi-squared statistic is

$$\chi^2 = \sum_{i=1}^{2} \frac{[X_i - (n/2)]^2}{n/2} \ .$$

Neyman suggested the modified Chi-squared statistic, given by

$$N = \sum_{i=1}^{2} \frac{[X_i - (n/2)]^2}{X_i} \ .$$

(i) Simplify N to be a function of X_1 alone.

(ii) Is it true that $N \geq \chi^2$, or $N \leq \chi^2$, or neither?

(iii) Under the null hypothesis, derive the large sample distribution of N.

5. Suppose $X_1, \ldots, X_n$ are i.i.d. according to the exponential density with parameter λ, i.e. they have common density function $\lambda \exp(-\lambda x)$ for $x > 0$. Independent of the X_i, suppose $Y_1, \ldots, Y_n$ are i.i.d. according the the exponential density with parameter μ. Both λ and μ are unknown and may differ.

(i) What is the maximum likelihood estimator for λ/μ?

(ii) Construct an approximate 95 percent confidence interval for λ/μ and evaluate it in the case $n = 100$ and you observe $\sum_{i=1}^{100} X_i = 100$ and $\sum_{i=1}^{100} Y_i = 200$. *Hint*: You may find it helpful to first construct a confidence interval for $\log(\lambda/\mu)$.