Math 494. Final Exam. May 10, 2004

Instructions. Work each of Problems 1-8 and exactly one of Problem 9 or 10.

1. In HW 10 we considered a normal homoscedastic bilinear regression model

\[ Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \epsilon_i, \]

and had the following data:

\[
\begin{bmatrix}
  x_1 & 0 & 0 & 0 & 1 & 1 & 1 \\
  x_2 & 0 & 1 & -1 & 0 & 1 & -1 \\
  y & 1 & 1.5 & .5 & 2 & 5 & 2
\end{bmatrix}
\]

With the notations of class, we found that

\[ (Z'Z)^{-1} = \begin{bmatrix}
  1/3 & -1/3 & 0 \\
  -1/3 & 2/3 & 0 \\
  0 & 0 & 1/4
\end{bmatrix}, \quad \hat{y} = (1, 2, 0, 3, 4, 2), \quad S^2 = |\bar{y} - \hat{y}|^2 = 2.5, \]

and that the observed values of the mle's were \( \hat{\beta}_0 = 1, \hat{\beta}_1 = 2, \hat{\beta}_2 = 1. \)

(a) Find \( \text{Var}(\hat{\beta}_1), \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \) and \( \text{Var}(\hat{\beta}_0 - \hat{\beta}_1). \)

(b) Find a 95\% confidence interval for \( \beta_2. \)

(c) Test the null hypothesis \( \beta_0 = \beta_1 = \beta_2 = 0 \) against the alternative that at least one \( \beta_i \) is not zero, at significance level .05.

(d) There is an \( F \)-test which can be used to test \( H_0 : \beta_1 = \beta_2 = 0 \) against the alternative that \( H_0 \) is false. The numerator of the \( F \)-statistic involves a term \( |\hat{y} - \bar{y}|^2. \)

Find \( \hat{\epsilon} \). You may use the fact that the design matrix \( Z \) maps vectors \((t,0,0)\) in \( \mathbb{R}^3 \) to vectors \((t,t,t,t,t,t)\) in \( \mathbb{R}^6 \), where \( t \) denotes a real number.
2. In a simple linear regression model $EY_i = \beta_0 + \beta_1 x_i$ we let $\mathbf{Z} = A\mathbf{Y}$ where $A$ is an $n \times n$ orthogonal matrix with first row $a(1,1,...,1)$, second row $b(x_1 - \bar{x},...x_n - \bar{x})$, and $a$, $b$ are certain real numbers. Let also, as usual, $S^2 = |\mathbf{Y} - \hat{\mathbf{Y}}|^2$. Write down a formula for $S^2$ in terms of the $Z_i$, and tell me what $a$ and $b$ are.

3. Suppose that $X_1, ..., X_n$ form a random sample from a normal distribution with mean $\mu$ and variance $\sigma^2$. Let $S^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.

(a) For a certain $c$, $cS^2$ has a $xx$ distribution. Tell me what $c$ and $xx$ are. Be sure to fully specify the parameters in $xx$.

(b) State a simple formula for $E(S^2)$.

(c) Suppose that $n = 6$ and that $\sum_{i=1}^{6} X_i = 30$, $\sum_{i=1}^{6} X_i^2 = 187$. Find $S^2$.

(d) Suppose that $n \geq 6$. Let

$$W = \frac{\sum_{i=1}^{3}(X_i - \mu)^2}{\sum_{i=4}^{6}(X_i - \frac{1}{3}(X_4 + X_5 + X_6))^2}.$$ 

Find the .95 quantile of $W$.

4. Suppose that a single observation is taken from a distribution for which the pdf is either $f_0$ or $f_1$, where $f_0(x) = e^{-x}$ and $f_1(x) = 2e^{-2x}$ for $x > 0$, $f_0(x) = f_1(x) = 0$ for $x \leq 0$. The following hypotheses are to be tested:

$$H_0 : f = f_0, \quad H_1 : f = f_1.$$ 

Let $\delta$ denote the test whose critical region consists of the outcomes with $X < 1/4$. Find $\alpha(\delta)$ and $\beta(\delta)$, the errors of type 1 and of type 2, respectively.
5. Suppose that $X_1, \ldots, X_4$ form a random sample of size 4 from a normal distribution for which the mean $\mu$ is unknown and the variance is known to be 25. The following hypotheses are to be tested:

$$H_0 : \mu = 1, \quad H_1 : \mu = 7.$$ 

(a) Among all test procedures for which the size $\alpha(\delta) \leq .01$, describe a procedure for which $\beta(\delta)$ is a minimum.

Note: The critical region for this test has the form $A > c$ or $A < c$, where $A$ is some statistic. Your job is to identify $A$, decide whether $<$ or $>$ is correct, and to determine the value of $c$.

(b) Suppose that we perform the test in (a) 100 times under independent conditions. If $H_0$ is true, what is the probability that we will reject $H_0$ at least once?

6. Suppose that a random sample of size $n$ is taken from a Poisson distribution for which the mean $\lambda$ is unknown.

(a) Write down the likelihood functions $f(x|\lambda)$ and $f_n(\overline{x}|\lambda)$.

(b) Suppose that $n = 4$ and that the sample values are 0, 1, 3, 6. Find the maximum likelihood estimate of $\lambda$ under each of the following assumptions on $\lambda$:

(i) $\lambda \in [0, \infty)$
(ii) $\lambda \in [1, 2]$.

7. Suppose that the proportion $\theta$ of defective items in a large manufactured lot is unknown, and that the prior distribution of $\theta$ is a beta distribution with parameters $\alpha = 3$ and $\beta = 5$. When 15 items are randomly selected from the lot, that is, from a Bernoulli distribution with “success probability” $\theta$, it is found that exactly 8 of the items are defective.
(a) The posterior distribution \( \eta \) is a xx distribution with parameters yy and zz. Tell me what xx, yy and zz are, and indicate briefly how one obtains this result.

(b) Find, using the squared error loss function, the Bayes estimate of \( \theta \) and the mean square error of the Bayes estimate.

8. Suppose that \( X_1, \ldots, X_4 \) form a random sample of size 4 from a Bernoulli distribution for which the value of the success parameter \( p \) is unknown, and it is desired to test the following hypotheses:

\[
H_0 : p \leq 0.4, \quad H_1 : p > 0.4.
\]

Let \( \delta \) denote the test whose critical region consists of the outcomes \( Y \geq 3 \), where \( Y \) denotes the number of successes in four trials.

(a) Determine the value of the power function \( \pi(p) \), for each \( p \in [0, 1] \).

(b) Calculate the derivative of \( \pi \). Then draw a rough sketch of the graph of \( \pi \), making sure to show the set(s) where \( \pi \) is increasing and where decreasing.

(c) Determine the size of the test.

In a survey by the Symphony Alliance respondents were asked to rate how well, on a scale of 0-100, they like various composers, among them Bach and Mozart. Do exactly ONE of the following two problems. Take \( X \) to be the Bach rating of an individual and \( Y \) his Mozart rating.

9. In one group with \( n = 100 \) respondents, the summary statistics were as follows:

\[
\bar{x} = 70, \quad \bar{y} = 80, \quad s_x = 30, \quad s_y = 40, \quad r = 0.5.
\]
Assume that \((X, Y)\) has a bivariate normal distribution. Then, for example, the usual estimates of the mean and variance of \(X\) coincide with the maximum likelihood estimates, and are \(\bar{X}_n\) and \(\frac{1}{n}(s_x)^2\), respectively.

(a) The conditional distribution of \(Y\), given that \(X = 76\), is normal. What are the usual estimates for its mean and variance? You may wish to first calculate an estimated regression line.

(b) Using the estimates in (a), find the approximate percentage of people with Bach rating 76 who rated Mozart at least 90.

10. In another group of respondents a question of of interest was whether Bach opinion was related to city of residence. Samples from Atlanta, Boston and Chicago were taken, with the following results:

\[
\begin{bmatrix}
A & B & C \\
10 & 10 & 20 \\
80 & 80 & 50 \\
\end{bmatrix}
\]

The top row indicates the city, the second row the number of observations in that group, and the third row the average Bach rating.

(a) Under appropriate normality and homoscedasticity assumptions, the null hypothesis that all three cities have the same mean Bach opinion can be tested via one-way analysis of variance using a statistic with an \(F(m_1, m_2)\) distribution. What are \(m_1\) and \(m_2\) here?

(b) This statistic contains a quantity \(\bar{Y}_{++}\). What is it here?