Ph. D. Qualifying Exam (Part II of Written Exam)
Friday, January 5, 1996

You have four hours. Do as many problems as you can. No one is expected to answer all the problems correctly. Partial credit will be given. All problems are worth an equal amount of credit.

Put your solution to each problem on a separate sheet of paper.

Problem 1. Let $X_1, X_2, \ldots, X_n$ be i.i.d. from the density $f(x) = \lambda e^{-\lambda(x-\mu)}$ for $x \geq \mu$ (and $f(x) = 0$ otherwise).

(a) Find a sufficient statistic for (λ, μ).

(b) Show that your statistic is complete.

(c) Let $\overline{X}$ and S^2 denote the sample mean and variance respectively. To be precise, we define $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$. Let $X_{(1)} = \min X_i$. Suppose you are given that

$$E(\overline{X} - X_{(1)})^2 = \frac{n-1}{n\lambda^2}\,.$$

Use this fact and the results of parts (a) and (b) to compute the value of

$$E\left(\frac{S^2}{(\overline{X} - X_{(1)})^2}\right)\,.$$

Justify each step of your calculation. (If you could not prove the completeness in (b), simply assume it now.)

Problem 2. The problem of bias in judging Olympic diving and ice skating seemed to be particularly noticeable during the Cold War, and there is usually a *perception* of bias indicated by some noisy members of the audience at every Olympics.

Suppose you have seven judges and ten ice skaters in the Olympics; each judge observes a skater's performance and assigns a score to him/her, for 70 scores in all.

(a) State a model which allows for systematic differences among judges and among skaters. Identify symbols; give the ANOVA table.

Suppose that Skater 1 is Nancy Kerrigan from the USA and Skater 10 is Oksana Baiul from the former USSR; Judge 1 is from the USA and Judge 7 is from the former USSR.

(b) How would you graphically analyze the data in order to see whether (say) Judge 1 favors Skater 1 and/or downrates Skater 10, and/or Judge 7 downrates Skater 1 and/or favors Skater 10?

(c) We know from the work of Cuthbert Daniel that an outlier in cell i, j of a two-way layout affects the predictions for other cells in Row i and Column j in the obvious direction. How does this knowledge affect your opinion about the sensitivity of the analysis in b)?

(d) Modify the model of a) to allow for any or all of the biases mentioned in b).

(e) Describe how you would test the null hypothesis that Judge 1 is not biased in favor of Nancy Kerrigan, using the model of d). (Identify the things you need to calculate and give the test statistic and the df. You may assume, if you need to, that you have a standard regression program available.)

Problem 3.

(a) Show that if a random variable X is independent of itself, then X is degenerate.

(b) Let X and Y be two random variables with $X > 0, Y > 0$ and such that $E(X|Y) = Y$ and $E(Y|X) = X$. Show that $X = Y$ with probability 1.
Hint: use the fact that $\frac{x}{y} + \frac{y}{x} \geq 2$ if $x > 0, y > 0$.

Problem 4. Consider the linear regression model: $Y = X\beta + \xi$ where $Y = (y_1, \ldots, y_n)'$, $\xi = (\xi_1, \ldots, \xi_n)'$, $\beta = (\beta_1, \ldots, \beta_p)'$ and X is an $n \times p$ full-rank matrix. The process $\{\xi_i\}$ is generated by the moving-average model: $\xi_i = \epsilon_i - 0.70\epsilon_{i-1}$, where $\{\epsilon_i, i = 0, 1, \ldots, n\}$ are iid $N(0, \sigma^2)$ variables.

(a) Calculate the covariance matrix $\Gamma = \mathrm{Cov}(\xi)$

(b) Let $\hat{\beta}$ be the least squares estimate of β, $\hat{Y} = X\hat{\beta}$ and $\hat{\xi} = Y - \hat{Y}$. Are the following two statements valid? Why or why not?

i) $\hat{\beta}$ is an unbiased estimate of β; ii) $\hat{Y}$ and $\hat{\xi}$ are independent.

(c) Find the maximum likelihood estimate $\tilde{\beta}$ of β. Define $\tilde{Y} = X\tilde{\beta}$ and $\tilde{\xi} = Y - \tilde{Y}$. Are the two statements in (b) valid for $\tilde{\beta}, \tilde{Y}$ and $\tilde{\xi}$?

Problem 5. An engineer who studies traffic congestion comes to you with the problem of finding the expectation of a waiting time $T \geq 0$. The tail probabilities of T are given by the mean of a certain nonnegative stochastic process $\{X(t), t \geq 0\}$:

$$P(T > t) = E[X(t)] \quad \text{for all } t \geq 0.$$

(a) Let L be a mean-1 exponential random variable which is independent of the process $\{X(t), t \geq 0\}$. Show that

$$Y = \int_0^L e^t X(t)\, dt$$

is an unbiased estimator of $E(T)$.

(b) Express the condition of independence that is used in a) in terms of independence between events.

(c) State any extra assumptions (*if* any) that are needed for your solution to a).

Problem 6.

(a) Let X be a binomial random variable with parameters n and p, and set $\hat{p} = X/n$. Show that

$$2\sqrt{n}\left(\sin^{-1}\sqrt{\hat{p}} - \sin^{-1}\sqrt{p}\right) \xrightarrow{\mathcal{D}} N(0,1).$$

(b) Researchers in a school of medicine plan to conduct a clinical experiment comparing the infection rates of minor cut and injury when the fresh wound is cleansed either with saline solution or with tap water. Let p_1 and p_2 be the infection rates using saline solution and tap water, respectively.

The researchers wish to know how large the sample sizes n_1 and n_2 need to be to test the hypothesis

$$H_0\colon p_1 = p_2 \quad \text{vs} \quad H_1\colon p_1 < p_2$$

at the 5% level with 80% power. (Preliminary results indicate that $p_1 \approx .03$ and $p_2 \approx .08$.) Using a) or otherwise, determine suitable n_1 and n_2. Express your answer in a form that could be numerically evaluated using Splus. Make all necessary assumptions.

Problem 7. Let $X_1, X_2, \ldots, X_n$ be i.i.d. with uniform distribution on $(0, \theta), \theta > 0$.

(a) Show that $T = \max_i X_i$ is sufficient for θ.

(b) Show that T is complete for θ.

(c) Find c such that cT is unbiased for θ and deduce that cT is the uniformly minimum variance unbiased (UMVU) estimator for θ.

(d) Consider all estimators of the form cT. Show that the UMVU estimator is inadmissible with respect to mean square error.

Problem 8. Let X_n, Y_n, $n \geq 1$, be random variables such that the conditional distribution of Y_n given X_n is Bernoulli with parameter $p_n(X_n)$, where $p_n : \mathbb{R} \to [0,1]$ is non-random.

(a) Suppose that $X_1, X_2, \ldots$ are i.i.d. Bernoulli with parameter $0 < p < 1$ and $p_n(0) = n^{-2} = 1 - p_n(1)$. Show that

$$n^{-1/2} \sum_{i=1}^{n} (Y_i - p)$$

converges in distribution as $n \to \infty$.

(b) Suppose that X_n converges in distribution to a continuous random variable X, and $p_n = F$, where F is a fixed distribution function. Show that Y_n converges in distribution as $n \to \infty$.

(c) Identify the limits established in a) and b).

Problem 9. Let $X_1, ..., X_n$ be i.i.d. with pdf $f(x, \theta) = \theta x^{-(\theta+1)} I_{(1,\infty)}(x)$ and suppose that $\theta > 2$. Let $\hat{\theta}_n$ be the method of moments estimator of θ.

(a) Determine which of the following is true (you need to show your work):

1. $\hat{\theta}_n$ is an unbiased estimator of θ;
2. $\hat{\theta}_n$ has an upward bias (i.e., $E_\theta(\hat{\theta}_n) > \theta$);
3. $\hat{\theta}_n$ has a downward bias.

(b) Find the asymptotic distribution of $\hat{\theta}_n$.

(c) Calculate the asymptotic relative efficiency of $\hat{\theta}_n$ when $\theta = 3$.