Ph. D. Qualifying Exam (Part II of Written Exam)
Saturday, January 15, 1994

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. Consider the following two simple linear models

$$\begin{aligned} y_{1j} &= \beta_0 + \beta_1 x_{1j} + \epsilon_{1j}, \quad j = 1, 2, \ldots, n_1 \\ y_{2j} &= \gamma_0 + \gamma_1 x_{2j} + \epsilon_{2j}, \quad j = 1, 2, \ldots, n_2 \end{aligned}$$

where the x_{1j}'s and x_{2j}'s are fixed values, $\{\epsilon_{ij}, j = 1, 2, \ldots, n_i; i = 1, 2\}$ are i.i.d N(0, σ^2) random variables, and $\beta_1 \neq \gamma_1$. The two straight lines $\mathrm{E}(y_1) = \beta_0 + \beta_1 x$ and $\mathrm{E}(y_2) = \gamma_0 + \gamma_1 x$ cross at the (unknown) value $x_0 = \frac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1}$. Let

$$S_x^{(i)} = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \quad S_y^{(i)} = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2, \quad S_{xy}^{(i)} = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i).$$

Denote the least squares estimates of the parameters by $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$.

(a) Based on the observations $\{(x_{ij}, y_{ij}), j = 1, 2, \ldots, n_i; i = 1, 2\}$, find an estimate, say $\hat{\sigma}^2$, for the variance σ^2. Express $\hat{\sigma}^2$ in terms of $S_x^{(i)}$, $S_y^{(i)}$ and $S_{xy}^{(i)}$. What is the distribution of $\hat{\sigma}^2$?

(b) Let

$$\xi = \left(\hat{\beta}_0 + \hat{\beta}_1 x_0\right) - \left(\hat{\gamma}_0 + \hat{\gamma}_1 x_0\right).$$

Find the mean value and variance of ξ. What is the distribution of ξ?

(c) Based on the results in **(a)** and **(b)**, construct a 100(1-α)% confidence interval for x_0 at a given significance level $\alpha > 0$.

Problem 2.

(a) Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with moment generating function (mgf) $\psi(t) = E(e^{tY_1})$. Show that

$$P\left(\sum_{i=1}^{k} Y_i > 0\right) \leq \{\psi(t)\}^k$$

for all $t > 0$.

(b) Let $X_1, X_2, \ldots$ be zero mean i.i.d. random variables having finite mgf in some neighborhood of the origin. Let $S_n = \sum_{i=1}^n X_i$ and set $\epsilon > 0$. Show that there exists $0 < \tau < 1$ and a constant c such that

$$P(|\frac{S_k}{k}| > \epsilon \quad \text{for some } k \geq n) \leq c\tau^n.$$

Hint: Apply (a) to $Y_i = X_i - \epsilon$ and $Y_i = -X_i - \epsilon$, and note that the mgfs are differentiable at $t = 0$.

(c) Deduce from (b) that $S_n/n \to 0$ almost surely.

Problem 3. Let μ be an arbitrary measure and let f_n, $n = 1, \ldots, \infty$ be integrable functions such that $f_n \to f_\infty$ almost everywhere $[\mu]$.

(a) Show that if each f_n is a probability density function (pdf) with respect to μ, then

$$\int |f_n - f_\infty| \, d\mu \to 0.$$

Hint: Split $f_n - f_\infty$ into its positive and negative parts.

(b) Does the conclusion to (a) hold without the pdf assumption? If so, give a proof, otherwise give a counterexample.

Problem 4. A local movie rental shop owns a single tape of the movie *The Wizard of Oz.* End-of-the-year records indicate that this movie has been checked out for 40 of the last 52 Sundays. Since this means that some business is probably being lost, the manager wishes to consider buying more tapes of the movie. Find an approximate lower 90% confidence interval (i.e. a right infinite interval) for the total number of tapes the shop should own if it is desired that the probability is only 20% that "business is lost" on a given Sunday night. ("Business is lost" when the number of customers who wish to rent the movie exceeds the number of tapes owned.)

Note: This problem requires you to make reasonable assumptions. State the assumptions you are making. Explain in detail how to construct the desired confidence interval, and use a hand calculator to obtain an approximate answer. Show your work.

Problem 5. An experiment is conducted to measure a constant θ. Independent unbiased measurements of θ can be made with either of two instruments, both of which measure with normal error. Instrument 1 produces errors that are independent and normally distributed with mean 0 and variance 1, and Instrument 2 behaves identically except that the variance is 2. When a measurement is taken, a record is kept of the instrument that was used. After a series of 10 measurements, the results are $\boldsymbol{Y} = (X_1, I_1), \ldots, (X_{10}, I_{10})$, with I_i being the indicator of the instrument that was used in measurement i. The choice of which instrument is used is beyond the control of the experimenter, and is considered by the experimenter to be Bernoulli(1/2).

(a) Find the sufficient statistic for this problem.

(b) Find the maximum likelihood estimate of θ based on the data $\boldsymbol{Y}$.

(c) Find a 95% confidence interval for θ.

Problem 6. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with a common $N(0, 1)$ distribution. Let $Y_n = \frac{1}{n} \sum_1^n X_i$ and let $Z_n =$ median of $(X_1, X_2, \ldots, X_n)$. With appropriate justifications, obtain the asymptotic distribution of (Y_n, Z_n) and of $(Y_n - Z_n)$.

Problem 7. Consider the linear model

$$\underset{n\times 1}{\boldsymbol{Y}} = \underset{n\times p}{\boldsymbol{X}} \; \underset{p\times 1}{\boldsymbol{\alpha}} + \underset{n\times k}{\boldsymbol{Z}} \; \underset{k\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{e}}$$

where $n > p + k$. Assume that $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\mathrm{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$. Assume further that $\boldsymbol{Z}$ is of full rank but $\boldsymbol{X}$ is not, and that the columns of $\boldsymbol{Z}$ are linearly independent of those of $\boldsymbol{X}$.

(a) Show that the ordinary least squares estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Q}\boldsymbol{Y}$$

where $\boldsymbol{Q}$ is the projection onto the orthogonal complement of the subspace generated by the columns of $\boldsymbol{X}$.

(b) Derive $E(\hat{\boldsymbol{\beta}})$ and $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$.

Problem 8. A measurement process is used often in a certain lab, and concern arises that day-to-day variability in lab conditions might be affecting variability in the measurement process. For each of five recent days, three sample variances are recorded for sets of ten measurements taken under similar conditions.

Describe how you would test the null hypothesis that the measurement variances on each of the five days are equal, and justify your approach.

(a) First assume the individual observations are normally distributed. Describe how to test the null hypothesis using an F-test.

(b) You could use a χ^2 test; describe this approach.

(c) Indicate how your justifications of the procedures in the first two parts would change if the individual measurements were not normally distributed.

Hint: The variance of s^2 is

$$\sigma^4(\frac{2}{n-1} + \frac{\gamma_2}{n}),$$

where σ^2 is the variance of the underlying population distribution and γ_2 is the kurtosis.

Problem 9. Let $X_1, X_2, \ldots$ be i.i.d. Cauchy random variables with pdf

$$f(x) = \frac{1}{\pi(1+x^2)}\,.$$

Let $Y_n = X_n I(|X_n| \leq a_n)$ where $I(\cdot)$ is the indicator function and $\{a_n\}$ is the sequence of constants $a_n = n^p$ where $p < 1$.

(a) Let $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$. Show that $\bar{X}_n$ does not tend to a constant in any notion of convergence.

(b) Show that the series $\sum_{n=1}^\infty (Y_n/n)$ converges with probability 1.

(c) Show that $\bar{Y}_n = n^{-1}\sum_{i=1}^n Y_i$ converges to 0 with probability 1.