Ph. D. Qualifying Exam (Part II of Written Exam)
Friday, January 5, 2001

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 $X_1, X_2, \ldots$ be i.i.d. with common distribution $N(1,1)$. Let $V_n = \frac{1}{n}\sum_1^n X_i$ and $M_n =$ median of $\{X_1, \ldots, X_n\}$ be the mean and median based on the first n random variables.

(a) Write down the limiting distributions of $\sqrt{n}(V_n - 1)$ and $\sqrt{n}(M_n - 1)$.

(b) Let

$$T_n = \begin{cases} V_n & \text{if } |V_n| \le n^{-\frac{1}{4}} \\ M_n & \text{if } |V_n| > n^{-\frac{1}{4}} \end{cases}$$

in other words,

$$T_n = V_n I(|V_n| \le n^{-\frac{1}{4}}) + M_n I(|V_n| > n^{-\frac{1}{4}}).$$

Obtain the limiting distribution of $\sqrt{n}(T_n - 1)$. *Hint: First compute the limit of* $E(I(|V_n| \le n^{-\frac{1}{4}})) = P(|V_n| \le n^{-\frac{1}{4}})$.

Problem 2. Suppose that an experiment involves 6 factors each with two levels. Consider the following two fractional factorial designs:

(i). 2^{6-2} with $5 = 1234$ and $6 = 134$

(ii.) 2^{6-2} with $5 = 123$ and $6 = 124$

(a) What is the resolution for each of the fractional factorial designs? Which design do you prefer? Justify your answers.

(b) For the design in (ii), if we further know that all two-factor interactions involving factor 6 (i.e., 16, 26, 36, 46, 56) are negligible, which two-factor interactions in the design are estimable under the assumption that three-factor and higher interactions are negligible?

Problem 3. 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:

$$\xi_i = \epsilon_i + \theta\epsilon_{i-1},$$

where $\{\epsilon_i, i = 0, 1, \ldots, n\}$ are iid $N(0, \sigma^2)$ variables. Let $\hat{\beta}$ be the least squares estimate of β, $\hat{Y} = X\hat{\beta}$, and $\hat{\xi} = Y - \hat{Y}$.

(a) What are means and covariance matrices of $\hat{Y}$ and $\hat{\xi}$? What are the distributions of $\hat{Y}$ and $\hat{\xi}$?

(b) Are $\hat{Y}$ and $\hat{\xi}$ independent?

Problem 4. A social scientist was studying students' admiration for world leaders. She selected a random sample of 10 seniors from each of three majors at FSU. Each senior was asked to name eight world leaders. She then scored the results by counting, for each of $30 \times 29/2 = 435$ pairs of students, the number of world leaders they agreed on (X = 0, 1, ..., 8). Her hypothesis is that students from the same major will name more leaders in common than students from different majors.

How would you test this hypothesis, given data like this?

Problem 5. Suppose one assumes the multiple linear regression model

$$\mathbf{Y} = \mathbf{X_1}\boldsymbol{\beta_1} + \mathbf{X_2}\boldsymbol{\beta_2} + \epsilon$$

when in fact $\boldsymbol{\beta_2} = \mathbf{0}$, in other words, the model fitted by the experimenter is an overfitted model and the true model is $\mathbf{Y} = \mathbf{X_1}\boldsymbol{\beta_1} + \epsilon$, where $\mathbf{Y}$ is a column vector of dimension $n \times 1$, $\mathbf{X_1}$ and $\mathbf{X_2}$ are design matrices of dimensions $n \times p$ and $n \times (m - p)$, respectively, $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are $p \times 1$ and $(m - p) \times 1$ parameter vectors, respectively, and $\epsilon \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\sigma^2}\mathbf{I})$. Let s_m^2 denote the residual mean square based on the overfitted model, i.e., $s_m^2 = \mathbf{Y^t}[\mathbf{I} - \mathbf{X}(\mathbf{X^t}\mathbf{X})^{-\mathbf{1}}\mathbf{X^t}]\mathbf{Y}/(n - m)$, where $\mathbf{X} = [\mathbf{X_1} \vdots \mathbf{X_2}]$. Show that the residual mean square s_m^2 is unbiased for σ^2 even if one overfits.

Problem 6. Let $\{X_n\}$ be a sequence of identically distributed random variables with finite mean, and write $M_n = \max_{1 \leq i \leq n} |X_i|$.

(a) Show that $E(M_n)/n \to 0$.

(b) Under the additional assumption that the X_n are independent show that $M_n/n \to 0$ almost surely.

Problem 7. Let P be a probability measure on the real line having a density f with respect to Lebesgue measure λ.

(a) Show that $\lambda(A_n) \to 0$ implies $P(A_n) \to 0$ for every sequence of Lebesgue measurable sets A_n.

(b) Show that

$$P(A) = \lambda_2\{(x, y) \colon x \in A, 0 \le y \le f(x)\}$$

for any Lebesgue measurable set A, where λ_2 is two-dimensional Lebesgue measure.

Problem 8. Let X be a random variable taking values in $\{-2, -1, 0, 1, 2\}$ with mass function given by $P(X = 0) = (1 - \theta)^2$, $P(X = 1) = P(X = -1) = \theta(1 - \theta)$, and $P(X = 2) = P(X = -2) = \theta^2/2$ where θ is unknown and $0 < \theta < 1$. Suppose you observe data $X_1, X_2, \ldots, X_n$ which are i.i.d. with the same distribution as X. The following two parts are unrelated and can be worked independently of each other.

(a) Find a complete sufficient statistic for θ (and show that your statistic has these properties).

(b) Find the Cramér-Rao lower bound for the variance of an unbiased estimator of $\theta^{\mathbf{3}}$.

Problem 9.

(a) Prove the following fact:

$$\text{If } \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right),$$

$$\text{then } X_1 \mid X_2 \sim N(\mu^*, \Sigma^*)$$

$$\text{with } \mu^* = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2) \text{ and } \Sigma^* = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\,.$$

(Assume the matrices and vectors above are conformable and that Σ_{22}^{-1} exists.)

(**Hint:** Show that $X_1 - \mu^*$ and X_2 are independent.)

(b) Let X be a known $n \times p$ matrix. Consider the Bayesian linear model where $Y \sim N(X\beta, I)$ and β has the prior distribution $\beta \sim N(\xi, I)$. Find the posterior distribution of β.

(**Hint:** Use the result of part (a). You do **not** have to write the answer in the simplest possible form; just leave it the way you get it upon application of part (a).)