Ph.D. Preliminary Exam
Thursday, January 2, 1992

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.

Please put each problem on a separate sheet of paper.

Problem 1: In the following question about fractional factorial designs there are two types of factors which we have indicated by using upper and lower case.
Suppose a 2^5 factorial design is constructed in factors **A, B, C, D** and **e**.

a) A sixth factor **f** is introduced by setting $\mathbf{f} = \mathbf{ABCD}$. What is the generator of this design?

b) Define *resolution* of a fractional factorial design. What is the resolution of this design?

c) Is this the highest resolution obtainable with six factors in 32 runs? Justify your answer.

d) Is high resolution desirable? Why or why not?

e) A seventh factor **g** is introduced by setting $\mathbf{g} = \mathbf{ABCDe}$. What is the defining relation of this design?

f) Justify the following statements:

- **i)** This design (in **A, B, C, D, e, f,** and **g**) is Resolution III in the three lower-case letters **e, f, g**.
- **ii)** Every pair of letters from **A, B, C, D** can be estimated unconfounded with any other pair.
- **iii)** Every interaction between an upper-case letter and a lower-case letter can be estimated unconfounded with any other two-factor interaction.

Problem 2: Consider the linear model in which we observe $Y \sim \mathrm{N}_n(\mathbf{X}\beta, \sigma^2\mathbf{I}_n)$, $\sigma^2 > 0$, where Y is a $n\times 1$ random vector, $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p\times 1$ vector of parameters, and $\mathbf{X}$ is a $n \times p$ matrix of constants. Assume that the rank of $\mathbf{X}$ is $p < n$. Let $\mathbf{A}$ be a $(p-k) \times p$ matrix of full rank such that the elements of $\mathbf{A}\beta$ are estimable.

Derive a reasonable test statistic for the hypothesis

$$H_0 : \mathbf{A}\beta = d, \qquad \text{versus} \qquad H_1 : \mathbf{A}\beta \neq d$$

where d is a $(p-k) \times 1$ vector of constants. Identify the non-null distribution of the test statistic.

Problem 3: Let $X_1, X_2, \ldots, X_n$ be i.i.d. Bernoulli random variables with $\Pr\{X_i = 1\} = p$. Assume $n \geq 2$. Define $\hat{p} = \frac{1}{n}\sum_i X_i$. Find the uniformly minimum variance unbiased estimator of $p(1-p)$ and show that it has the stated property.

Problem 4: Let X be a random variable with characteristic function $\phi(t)$.

a) Show that

$$P(X = a) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-ita}\phi(t)\, dt.$$

b) Using (a), or otherwise, show that if Y is an independent copy of X, then

$$P(X = Y) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\phi(t)|^2\, dt.$$

c) Using (b), or otherwise, show that if $|\phi(t)|^2$ is integrable then X has a continuous distribution function.

Problem 5: Answer the following.

a) Let X be a random variable and let $0 < b < \infty$ be a constant. Let $Y = XI(|X| \leq b)$. Show that

$$\frac{E(Y^2)}{2b^2} + \frac{P\{|X| > b\}}{2} \leq E\left(\frac{X^2}{X^2 + b^2}\right).$$

b) Let $\{X_n, n = 1, 2, \ldots\}$ be *independent* random variables and let $b_n \nearrow \infty$. Let $Y_n = X_n I(|X_n| \leq b_n), n \geq 1$. Suppose that

$$\sum_1^\infty E\left(\frac{X_n^2}{X_n^2 + b_n^2}\right) < \infty.$$

Show that

$$\frac{\sum_1^n (X_j - E(Y_j))}{b_n} \to 0$$

with probability 1.

(*Hint. Show that* X_n *and* Y_n *are tail equivalent.*)

Problem 6: Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables distributed uniformly on $[0, 1]$, *i.e.* with $F(x) = x,\ 0 < x < 1$. Let the order statistics of the first n random variables be $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Let a_n be a sequence of integers such that $a_n/n \to 1$ and $n - a_n \to \infty$ as $n \to \infty$. An example of such a sequence is $[n - n^{3/4}]$, where $[\ ]$ represents the largest integer function. Let $Y_n = X_{(a_n)}$. Let $\beta_n = \frac{n}{\sqrt{(n-a_n)}}$. Show that the limiting distribution of $\beta_n(Y_n - \frac{a_n}{n})$ is $N(0, 1)$.

Problem 7: A measurement process is repeated over and over in a certain lab. A concern arises that day-to-day variability in lab conditions might be affecting variability in the measurement process. For each of three recent days, five sample variances are recorded for sets of ten measurements taken under similar conditions. [That is, you are given 15 sample variances s_{ij}^2, $1 \leq i \leq 3$, $1 \leq j \leq 5$, where each is computed from a sample of size 10.]
Describe how you would test the null hypothesis that the measurement variances on each of the three 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 sample variance s^2 for a random sample of size n from a population with variance σ^2 and kurtosis γ_2 has a variance given by

$$\mathrm{Var}(s^2) = \sigma^4\left(\frac{2}{n-1} + \frac{\gamma_2}{n}\right).$$

Problem 8: Let $X_1, X_2, \ldots$ be a sequence of i. i. d. Bernoulli trials, where $\Pr\{X_i = 1\} = p \neq \frac{1}{2}$. (Please note that $p \neq \frac{1}{2}$.)

a) Show that $\sum_{i=1}^{\infty} X_i/2^i$ converges in distribution to a random variable Y.

b) Show that the distribution of Y is continuous.

c) Show that the distribution of Y is singular with respect to Lebesgue measure.

Problem 9: Let α, β, γ be the interior angles of a triangle so that $\alpha + \beta + \gamma = 180$ degrees. Independent measurements A, B, C are obtained for α, β, γ respectively. What is the 'best' estimate for α? In what sense is it best? (Prove your answer. Clearly state the model that you use and any assumptions that you make.)