Ph. D. Qualifying Exam (Part II of Written Exam)
Saturday, January 28, 1995

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. In a Balanced Incomplete Block (BIB) design, r replications of t treatments are arranged in b blocks of k experimental units with $k < t$. The total number of observations is $n = rt = bk$ with each treatment pair appearing together in the same block $\lambda = r(k-1)/(t-1)$ times someplace in the experiment. The linear model for a BIB design has the form:

$$y_{ij} \;=\; \mu + \tau_i + \beta_j + \epsilon_{ij}, \quad i = 1, \ldots, t; \;\; j = 1, \ldots, b,$$

where the ϵ_{ij}'s are i.i.d. $N(0, \sigma^2)$ and the parameters satisfy the constraints $\sum_{i=1}^{t} \tau_i = 0$ and $\sum_{j=1}^{b} \beta_j = 0$.

(a) Given the observations $\{y_{ij}\}$, find the least squares estimators for the parameters μ, $\{\tau_i\}$ and $\{\beta_j\}$.

(b) Let $\hat{\mu}_i = \hat{\mu} + \hat{\tau}_i$. Show that

$$E\hat{\mu}_i = \mu + \tau_i \quad \text{and} \quad \mathrm{Var}(\hat{\mu}_i) = \frac{\lambda t + kr(t-1)}{r\lambda t^2}\sigma^2 .$$

Problem 2. Let μ be a probability measure on the Borel subsets of the interval $(0, 1]$ that has a density f with respect to Lebesgue measure λ. Let $\mathcal{B}_n$ be the σ-field generated by the dyadic intervals $(k2^{-n}, (k+1)2^{-n}]$, $k = 0, 1, \ldots, 2^n - 1$. Denote the restrictions of μ and λ to $\mathcal{B}_n$ by μ_n and λ_n, respectively.

(a) Show that μ_n is absolutely continuous with respect to λ_n.

(b) Show that $d\mu_n/d\lambda_n$ converges Lebesgue a.e. to f as $n \to \infty$.

Problem 3. Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\theta, \theta^2)$.

(a) What is the Cramér-Rao lower bound for the variance of an unbiased estimator of θ^2

(b) Find an unbiased estimator of θ^2 which is asymptotically efficient.

Problem 4. Let $X_1, X_2, \ldots$ be independent normal random variables such that $EX_i = 0$, $EX_i^2 = \sigma_i^2$, $i = 1, 2, \ldots$. Show that

$$X_i \to 0 \ \text{ a.s.} \quad \text{if and only if} \quad \sum_{i=1}^{\infty} e^{-x^2/\sigma_i^2} < \infty \quad \text{for all } \ x > 0\,.$$

(You may find the following fact useful:

$$1 - \Phi(z) \sim \frac{\phi(z)}{z} \quad \text{as } \ z \to \infty,$$

that is, the ratio of the two sides goes to 1 as $z \to \infty$. Here $\Phi(z)$ and $\phi(z)$ are the c.d.f. and density of the standard normal distribution.)

Problem 5. For fixed $0 < p < 1$, let $\{A_n, n \geq 1\}$ be a sequence of events such that

$$n^{-1/2} \sum_{i=1}^{n} |P(A_i | \mathcal{F}_{i-1}) - p| \xrightarrow{P} 0$$

as $n \to \infty$, where $\mathcal{F}_n$ is the σ-field generated by $A_1, \ldots, A_n$, and $\mathcal{F}_0$ is the trivial σ-field. Let S_n be the number of $A_1, \ldots, A_n$ that occur.

(a) Show that $n^{-1/2}(S_n - np)$ converges in distribution to $N(0, p(1-p))$.

(b) For a fixed positive integer k, let $T = \min\{n \colon S_n \geq k\}$, where $\min \emptyset = \infty$. Deduce from (a) that $T < \infty$ a.s.

Problem 6. Suppose we observe responses $Y_1, Y_2, \ldots, Y_n$ which follow the model $Y_i = x_i^{\beta} U_i$ where β is unknown, the values $U_1, \ldots, U_n$ are unobserved but known to be i.i.d. Uniform(0,1), and the values $x_1, \ldots x_n$ are known fixed (nonrandom) constants with $x_i > 1$ for all i.

(a) Find a consistent, unbiased estimator of β and verify that it is consistent and unbiased. (The estimator does not have to be optimal in any sense.)

(b) What is (are) the sufficient statistic(s) in this problem?

(c) Is your estimator in part (a) admissible under squared error loss?

(d) What is the maximum likelihood estimator of β?

Problem 7.

(a) Show that

$$\lim_{n\to\infty} e^{-nt} \sum_{k\le nT} \frac{(nt)^k}{k!} = \begin{cases} 1 & t < T \\ 0 & t > T \end{cases}$$

[*HINT:* Let $S_n \sim \text{Poisson}(nt)$ and consider $P[S_n \le nT]$.]

(b) Let F be a c.d.f. such that $F(0) = 0$. Define the *Laplace transform* of F by

$$\phi(\lambda) = \int_0^\infty e^{-\lambda t} F(dt) \qquad \lambda \ge 0.$$

For $T \ge 0$ such that F is continuous at T, show that

$$F(T) = \lim_{n\to\infty} \sum_{k\le nT} \frac{(-n)^k}{k!} [D^k \phi](n),$$

where $[D^k\phi](n)$ is the k^{th} derivative of ϕ evaluated at n. [*HINT:* Compute the k^{th} derivative of ϕ and substitute.]

Problem 8. An ecologist samples n leaves; leaf i has area A_i and has on it k_i insects of the same species. The objective is to see whether insects prefer (say) large leaves.

(a) Method 1 is to calculate the correlation between k_i and A_i; if the correlation is positive and significant, the claim is made that the insects prefer large leaves. Briefly criticize this approach. (What null hypothesis are they implicitly making when they use this test?)

(b) Method 2 is to assume that

$$k_i/A_i = \beta_0 A_i^\theta \qquad (\dagger)$$

at least in some average sense. If there is no preference for size of leaf, what should the value of θ be? Why? What did you assume to get your answer?

(c) Using Method 2 described above, researchers wanted to regress $\log k_i$ on $\log A_i$. Why? What should the slope be, if leaf size does not attract insects?

(d) Because some leaves contained no insects researchers used $\log(k_i+1)$ as the dependent variable, instead of $\log k_i$. They found by simulation that this reduced the slope of the line compared to that predicted by ($\dagger$). Give an heuristic argument why the slope should be reduced.

(e) Suppose that the mean number of insects on a leaf of area A is λA^θ. Suggest a reasonable probability model for this situation. Assuming that θ is known, what is the maximum likelihood estimator of λ for your model?