Ph. D. Qualifying Exam
Tuesday, August 20, 2002

Please submit solutions to at most **seven** problems. You have four hours. 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 complete factorial experiment, consider the following mixed-effects model:

$$Y_{ij} \;=\; \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \ldots, a; \;\; j = 1, \ldots, b,$$

where μ is the overall mean, α_i is the fixed effect corresponding to the ith level of factor A, and β_j is the random effect due to the jth level of factor B. The β_j's are iid $N(0, \sigma_\beta^2)$ variables and the ϵ_{ij}'s are iid $N(0, \sigma_\epsilon^2)$ variables. Furthermore, $\{\beta_j\}$ and $\{\epsilon_{ij}\}$ are assumed to be independent. Let MSA and MSB denote the mean squares for factors A and B, respectively, and let MSE be the mean square for error.

(a) What are the differences between a random-effect design and a fixed-effect design? What hypotheses do you test in the mixed-effects model?

(b) Calculate the expected values $\mathrm{E}(MSA)$ and $\mathrm{E}(MSB)$.

Problem 2. 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 autoregressive model:

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

where $\{\epsilon_i, i = 0, \pm 1, \pm 2 \ldots,\}$ are iid $N(0, \sigma^2)$ variables.

(a) Show that $\{\xi_i, i = 1, 2, \ldots,\}$ is a second-order stationary process when $|\theta| < 1$. Calculate the autocovariance function $\gamma_k = \mathrm{Cov}(\xi_i, \xi_{i+k})$.

(b) How do you estimate β in this setting? Why do you choose this method? Discuss the properties of your estimate such as mean, covariance matrix, and distribution of $\hat{\beta}$.

Problem 3. A 2^{5-2} design is defined by $4 = 12$ and $5 = 13$.

(a) Find the defining relation and resolution of this design. What are the advantages and disadvantages of this kind of design compared with complete factorial designs?

(b) After the experiment, factor 5 turns out to be inert. It is assumed that all two-factor interactions involving factor 5 and all higher order interactions are negligible. In addition to estimating the four main effects, there are still three degrees of freedom left. What two-factor interactions can be estimated with these three degrees of freedom?

Problem 4. Let $Y_1, \dots, Y_n$ be independent random variables such that $Y_i \sim \text{Binomial}(m, \pi_i)$ where $0 < \pi_i < 1$. Let $Y = \sum_{i=1}^n Y_i$.

(a) Show that, given $\pi_1, \dots, \pi_n$, $\mathrm{E}(Y) = nm\bar{\pi}$ and $\mathrm{Var}(Y) = nm\bar{\pi}(1-\bar{\pi}) - m(n-1)k(\pi)$. Give the expression for $k(\pi)$ in terms of $\pi_1, \dots, \pi_n$.

(b) Assume that $\pi_1, \dots, \pi_n$ are independent random variables with common mean π and common variance $\tau^2\pi(1-\pi)$. Show that, unconditionally, $\mathrm{E}(Y) = nm\pi$ and $\mathrm{Var}(Y) = nm\pi(1-\pi)[1 + (m-1)\tau^2]$.

Problem 5.

(a) Define the notion of product measurability.

(b) Let $f\colon \Omega \to \mathbb{R}$ be a measurable function, where $(\Omega, \mathcal{F})$ is a measurable space. Show that the set

$$\text{Graph}(f) = \{(\omega, f(\omega))\colon \omega \in \Omega\}$$

is product measurable in a suitable sense. Hint: the graph of f is the inverse image under $(\omega, y) \mapsto (f(\omega), y)$ of a certain Borel measurable subset of $\mathbb{R}^2$.

Problem 6.

(a) State Markov's inequality.

(b) State Hölder's inequality and use it to derive Lyapounov's inequality.

(c) What can you say about a random variable X such that $E|X|^n \to 0$ as $n \to \infty$?

(d) Does there exist a random variable X for which $E|X|^n = 1/n$ for all $n \geq 2$?

Problem 7. Let $X_1, \ldots, X_n$ be i.i.d random variables with pdf 1 on [0,1] (i.e. uniform) and let $Y_1, \ldots, Y_m$ be i.i.d. with pdf $2x$ on $[0,1]$, and let the X's be independent of the Y's. Let $\phi(x;y) = I(x \leq y)$ and let $\phi^*(x;y) = \phi(x;y) - \frac{2}{3}$. Let $N = n+m$ and let $\frac{n}{N} \to \lambda$ where $0 < \lambda < 1$. Define

$$U_N = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \phi^*(X_i; Y_j)\,,$$

the two-sample U-statistic based on ϕ^*. State any requisite result on such U-statistics and obtain the asymptotic distribution of U_N.

Problem 8. Let $X_1, X_2, \ldots$ be random variables lying in $[0,1]$. Let $\mathcal{F}_0$ be the σ-field containing the empty set and the whole space, and let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, the smallest σ-field generated by $X_1, \ldots, X_n$. Show that the series

$$\sum_{i=1}^{\infty} X_i$$

and the series

$$\sum_{i=1}^{\infty} E(X_i | \mathcal{F}_{i-1})$$

converge or diverge together, with probability one.

Problem 9. Suppose $Y \sim (X\beta, \sigma^2 I)$ where X has **less than** full rank. Let M denote the orthogonal projection matrix onto the range of X. Prove the following: If $d'Y$ is an unbiased estimate of $\ell'\beta$, then $(Md)'Y$ is the BLUE of $\ell'\beta$.

Problem 10. Let $X_1, X_2, \ldots, X_n$ be iid Uniform$(\theta, 2\theta)$ where $\theta > 0$.

(a) Find a sufficient statistic for θ. (Your statistic should also be minimal, but you are not required to show this.)

(b) Show that the statistic you found in part (a) is **not** complete.

Problem 11. Suppose that the observed failure times of ball bearings $Y_1, \ldots, Y_n$ are a random sample from the lognormal distribution with p.d.f.

$$f(y) = \frac{1}{y\sqrt{2\pi}} \exp\{-(\log y - \mu)^2/2\} \quad \text{for } y > 0$$

where the parameter μ is unknown.

(a) Find the MLE of the mean failure time (i.e., $E(Y)$) of the ball bearings and its asymptotic distribution.

[Note: In your calculations, you may find it useful to recall the mgf of the Normal distribution: if $X \sim N(\mu, \sigma^2)$, then $Ee^{tX} = e^{\mu t + \sigma^2 t^2/2}$.]

(b) Another natural estimator of the mean failure time would be the average of the $Y's$: $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Is this estimator a good estimator relative to the MLE?

(c) Find $E(\bar{Y}_n \,|\, \bar{X}_n)$ where we define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ with $X_i = \log Y_i$.