Ph. D. Qualifying Exam
Monday, January 3, 2011

Put your solution to each problem on a separate sheet of paper.

Problem 1. (5166) An experiment was planned to compare three different fertilizers (A, B, C) on watermelon yields. The treatments were randomly assigned according to a Latin square design conducted over a large farm plot which was divided into rows and columns. The watermelon yields (in tons per acre) were recorded below after the growing season.

	Column		
Row	1	2	3
1	6.5 (B)	9.8 (C)	5.0 (A)
2	7.2 (A)	8.1 (B)	8.6 (C)
3	10.5 (C)	6.6 (A)	6.7 (B)

(a) If you were the experimenter, how would you randomize the design? Write an appropriate additive model for the data and state your assumptions on the model.

(b) Conduct an analysis of variance ($\alpha = 0.05$) based on the data and draw your conclusions.

Problem 2. (5166) Consider the following two 2^{7-2} fractional factorial designs:

d_1:	with	6=1234,	7=1235
d_2:	with	6=123,	7=145

(a) What is the defining relation for each of the two designs? What is the resolution of each design? For each design, write out the confounding pattern of the main effects $\{1, 2, 3, 4, 5, 6, 7\}$ with other effects.

(b) Write out the word-length pattern for each design. Which design is better? Justify your choice.

Problem 3. (5106) Let X be a real-valued, continuous random variable. Assuming that you are given a program for simulating U, the uniform random variable on $[0, 1]$:

(a) Describe the inverse transform method for generating random samples of X.

(b) Prove that this method works.

(c) Write a program to simulate the exponential random variable with mean λ.

Problem 4. (5327) Let $X_1, \ldots, X_n$ be a random sample from the density (pdf)

$$f(x|\theta) = \frac{2\theta x e^{\theta x^2}}{e^\theta - 1}, \quad 0 \le x \le 1 .$$

Answer the following. Justify your answers in detail. Show all calculations.

(a) Find a complete sufficient statistic for θ.

(b) Is there a function of θ, say $g(\theta)$, for which there exists an unbiased estimator whose variance attains the Cramér-Rao Lower Bound? If so, find it. If not, show why not.

(c) Find $I(\theta)$, the Fisher information in a **single** observation from $f(x|\theta)$.

Problem 5. (5327) Suppose you observe $X_1, \ldots, X_n$ iid Uniform$(\theta, \theta+1)$. Answer the following. Justify your answers.

(a) Find a minimal sufficient statistic for θ.

(b) Is the statistic you found above complete?

(c) Suppose you are a Bayesian and your prior distribution for θ is exponential with mean 1. Find the posterior density for θ.

Problem 6. (6346) Suppose $X \sim$ Geometric(p) with pmf

$$P(X = x) = (1-p)^{x-1} p \quad \text{for } x = 1, 2, 3, \ldots$$

and $X_1, X_2, X_3, \ldots$ is an arbitrary sequence of random variables.

For each of the functions g below, answer the following question: If $X_n \Rightarrow X$, is it necessarily true that $Eg(X_n) \to Eg(X)$?

If your answer is "yes", state your reason(s). If it is "no", give a counterexample.

[Note: $\Rightarrow$ and $\xrightarrow{\mathcal{L}}$ and $\xrightarrow{d}$ all mean the same thing: convergence in distribution (or weak convergence).]

(a) $g(x) = I_{(-1,5/2)}(x) = I(-1 < x < 5/2)$

(b) $g(x) = I(\sin(\pi x) > 0)$

(c) $g(x) = e^{-|x|}$

(d) $g(x) = \frac{1}{x+1}$

(e) $g(x) = e^{itx}$ where $i = \sqrt{-1}$ and $t \in \mathbb{R}$

Problem 7. (5167) In a complete factorial experiment, consider the following random-effects model:

$$\begin{aligned} Y_{ijk} &= \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \\ & \qquad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, r, \end{aligned}$$

where μ is the overall mean, the random effects α_i and β_j, are assumed to be independent and normally distributed with means of zero and variances σ_α^2 and σ_β^2, respectively, and the random errors $\{\epsilon_{ijk}\}$ are assumed to be independent and normally distributed with mean zero and variance σ^2. Furthermore, $\{\alpha_i\}$, $\{\beta_j\}$, and $\{\epsilon_{ijk}\}$ are assumed to be independent of one another.

(a) What are the distributions of $\bar{Y}_{ij\cdot}$, $\bar{Y}_{i\cdot\cdot}$, and $\bar{Y}_{\cdot j\cdot}$? Find the covariance $\text{Cov}(\bar{Y}_{ij\cdot}, \bar{Y}_{uv\cdot})$.

(b) Let $S_A = br\sum_{i=1}^a(\bar{Y}_{i\cdot\cdot} - \bar{Y})^2$, $S_B = ar\sum_{j=1}^b(\bar{Y}_{\cdot j\cdot} - \bar{Y})^2$, and

$$S_R = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r (Y_{ijk} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y})^2.$$

Find the mean values of S_A, S_B, and S_R.

(c) Find unbiased estimates for the variances σ_α^2 and σ^2 based on the observations $\{Y_{ijk}, \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, , r\}$.

Problem 8. (5167) Consider a multiple linear regression model with the form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \xi_i\,, \quad i = 1, \ldots, n,$$

where $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)'$ has a multivariate normal distribution with mean zero and covariance matrix $V = \sigma^2\Sigma(\boldsymbol{\theta})$, i.e., each element of matrix Σ is a function of the vector parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$. Let $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$.

(a) How do you estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and σ^2 in this model?

(b) How do you define outliers in this linear model?.

(c) Give a procedure to identify and test outliers in this model and give your justifications for the procedure.

Problem 9. (5106) Suppose $\mathbf{x} \in \mathbb{R}^n$ is a vector of random variables and one is interested in finding a smaller vector $\mathbf{z} \in \mathbb{R}^d$ $(d << n)$ that represents $\mathbf{x}$ as closely as possible. The vector $\mathbf{z}$ is constrained to be $\mathbf{z} = U^T\mathbf{x}$, where $U \in \mathbb{R}^{n\times d}$ is a tall, skinny matrix. The criterion for selecting U is such that $E[\|\mathbf{x} - U\mathbf{z}\|^2] = E[\|\mathbf{x} - UU^T\mathbf{x}\|^2]$ is minimized. (Here $\|\cdot\|$ denotes the 2-norm of a vector.) Derive an expression for the optimal U and illustrate your derivation using clear steps.

Problem 10. (5326) Consider a Pareto distribution with density (pdf)

$$f(x) = \begin{cases} \dfrac{3}{x^4} & \text{for } 1 \le x < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

(a) For this density, calculate the mean, median, mode, and variance.

(b) Suppose X and Y are iid with this density. Find the density of their product $Z = XY$. (Make sure to specify the support of the density.)

Problem 11. (5326) A random sequence of men and women enter a random sex-change clinic. The people entering the clinic are independent with probability $2/3$ of being a man and $1/3$ of being a woman. In the clinic, men are transformed into women with probability $1/4$, and women are transformed into men with probability $1/5$, with these decisions being made independently for each man or woman. Assume that one person arrives at the clinic each day, entering in the morning and leaving in the evening.

Give a detailed argument and calculation for each of the following.

(a) A person is seen **leaving** the clinic. What is the probability this person is a male?

(b) If a man is seen **leaving** the clinic, what is the probability that it was a man who **entered** the clinic?

(c) If, over the course of two days, two men are seen **leaving** the clinic, what is the probability that it was two men who **entered** the clinic?

(d) Over the course of two days, two people are observed **leaving** the clinic. If at least one of the people leaving is a man, what is the probability that both are men?

Problem 12. (6346)

(a) State the Radon-Nikodym Theorem, and define the important terms used in this theorem. Be as detailed as you can.

(b) Show that if $\nu \ll \mu \ll m$, then $\dfrac{d\nu}{d\mu} = \dfrac{\frac{d\nu}{dm}}{\frac{d\mu}{dm}}$.