M. S. Comprehensive Exam (Part I of Written Exam)
Thursday, January 13, 1994

You have three 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. An urn contains balls numbered 1 through N. Suppose that n balls ($n \leq N$) are randomly selected without replacement. Let Y denote the largest number selected. Find $P(Y = j), n \leq j \leq N$.

Problem 2. A poker hand consists of five standard playing cards. Many poker games assume that certain cards are "wild," that is that they may be taken to have *any* value, not simply the face value. Suppose we are playing poker with deuces (twos) wild. Assume all possible five card hands are equally likely.

A hand is said to be a five-of-a-kind if all cards have the same rank. For example, the hand K♠, K♡, K♢, K♣, 2♠ is a five-of-a-kind. What is the probability of a five-of-a-kind?

Problem 3. Let $X_1, \ldots, X_n$ be i.i.d. with pdf $\lambda e^{-\lambda x}, x > 0$. Let $Y_1, \ldots, Y_n$ be i.i.d with pdf $\mu e^{-\mu y}, y > 0$ and be independent of the X's. Let $Z_i = \min(X_i, Y_i), t_i = I(X_i \leq Y_i), i = 1, \ldots, n$, where $I(\cdot)$ is the usual indicator function. Obtain the joint distribution of $(Z_1, \ldots, Z_n, t_1, \ldots, t_n)$ and write down the sufficient satistics for (λ, μ).

Problem 4. Two models are fit by least-squares to a data set which consists of a response variable Y and three potential predictor variables X_1, X_2, and X_3.

Model 1: $EY = \beta_0 + \beta_1 X_1 + \beta_2 X_2$, $R^2 = 81\%$.

Model 2: $EY = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$, $R^2 = 87\%$.

(a) Do you prefer Model 1 or Model 2?

(b) In order to carry out a formal statistical test of whether Model 1 or Model 2 is preferred, what additional information would you need? Explain fully. (It is not enough to give a one-word answer.)

Problem 5. Let $(X_1, Y_1), ..., (X_n, Y_n)$ form a random sample from some continuous bivariate distribution $F(x, y) = P(X \le x, Y \le y)$. For each point (x, y), the value of the empirical distribution function $F_n(x, y)$ is defined to be the proportion of sample values (X_i, Y_i) that satisfy both $X_i \le x$ and $Y_i \le y$. Thus, if exactly k of the sample values satisfy both $X_i \le x$ and $Y_i \le y$, then $F_n(x, y) = k/n$.

(a) Find the mean and variance of $F_n(x, y)$.

(b) Let (u, v) and (x, y) be points such that $u < x$ and $v < y$.
Find $\text{Cov}(F_n(u, v), F_n(x, y))$.

Problem 6.

(a) State the model for a randomized block design with k treatments, b blocks, and kb observations.

(b) Write an ANOVA table listing the sources of variation, their degrees of freedom, and their expected mean squares.

(c) Why is the word "randomized" used in describing this design?

(d) Give an example of an experiment in which the data are taken in a treatment by blocks pattern but randomization is not possible. What effect does this have on the strength of your conclusions?

(e) Suppose that treatment and blocks interact. Use your answer from part (b) to discuss what happens to your inference. What can you do to check whether this interaction is occurring in your experiment?

Problem 7. Suppose you have a random sample $x_1, x_2, \ldots, x_n$ from a distribution with density $f(x) = (x - \theta)e^{-(x-\theta)}$ for $x \ge \theta$ and $f(x) = 0$ for $x < \theta$.

(a) Find an unbiased estimator of θ and show that it is unbiased. (There are many unbiased estimators. You are *not* being asked to find a good one, a simple one will do.)

(b) The maximum likelihood estimate (MLE) of θ can be written as the solution of a certain equation. What is this equation? (Simplify the equation as much as possible, but do not attempt to obtain a general closed form solution.) Show that this equation has a unique "reasonable" solution.