Ph. D. Preliminary Exam, Part I (Elementary Exam)
Tuesday, January 5, 1993

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

Suppose that $\{X_1, X_2, \ldots, X_n\}$ is a sample from a population with Weibull density

$$p(x, \lambda) = \lambda c x^{c-1} \exp(-\lambda x^c) \quad x > 0, \ \lambda > 0,$$

where c is a known positive number.

1 Find a sufficient statistic for λ. Is the statistic complete?

2 Find a uniformly minimum variance unbiased estimate for $\frac{1}{\lambda}$. What is the variance of this estimate?

Problem 2 Let $\alpha_1, \alpha_2, \ldots$ be a sequence of independent random variables with

$$\alpha_j \sim \text{Poisson}(z^j/j)$$

where $0 < z < 1$ is a fixed parameter. Let

$$\nu = \sum_{j=1}^{\infty} j\alpha_j$$

Show that $\nu \sim$ Geometric$(1 - z)$.

Hint: For $0 < u < 1$, $\sum_{j=1}^{\infty} u^j/j = -\log(1 - u)$.

Problem 3 Suppose that $Y_1, \ldots, Y_n$ are independent random variables with means $\alpha + \beta x_1, \ldots, \alpha + \beta x_n$, respectively, and common variance $\sigma^2 > 0$, where $x_1, \ldots, x_n$ are known constants, but α, β, σ^2 are unknown. From many previous experiments at a single fixed covariate value, say x_0, the average value of the response variable Y was y_0. The individual data were not saved, but from the previous work it may be assumed that the point (x_0, y_0) lies on the line $y = \alpha + \beta x$. Find the least-squares estimators for α and β.

Problem 4 Suppose X and Y are i.i.d. positive random variables, and $Z = X/(X+Y)$.

1 Find the mean μ_Z of Z.

2 If the variance σ_Z^2 of Z is known, place a lower bound on $E(1/Z)$.
Hint: use Taylor series.

Problem 5

1 Why is it important to have all important predictor variables in a multiple regression analysis? Compare the consequences of including worthless predictors in the model vs. excluding important predictors from the model.

2 Discuss some aspects of how multicollinearity complicates multiple regression analyses.

3 Present one assumption that is commonly made in the multiple regression model. How can you check to see if the data are consistent with this assumption? What remedial measures might be used if the assumption were violated in the data set?

Problem 6 Consider the simple linear regression model with intercept

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n.$$

Assume that the ϵ's have mean 0 and are uncorrelated.

In model diagnosis a common approach to evaluate the adequacy of a model would be to plot the residuals against the predicted values. The model is judged to be adequate if there is no visible apparent pattern in the scatterplot.

1 Give a mathematical justification for this approach.

2 Is this approach still valid in the evaluation of model adequacy if β_0 is known to be 0? Why or why not?

Problem 7 Let X and Y be independent random variables uniformly distributed on $[0, 1]$. Find the distribution of $\min(X, Y)/\max(X, Y)$.

Problem 8 Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. pairs of points from a bivariate normal distribution. Denote the mean of X_1 by μ_X, the mean of Y_1 by μ_Y, the variance of X_1 by σ_X^2, the variance of Y_1 by σ_Y^2, and the correlation coefficient between X_1 and Y_1 by ρ. Suppose that $n \geq 2$ and that there are no ties among the X's or the Y's.

1 Write the formula for $\hat{\rho}$, the Pearson (i.e. sample) correlation coefficient.

2 Prove or disprove the following statement. The distribution of $\hat{\rho}$ is symmetric about ρ.

3 Show that the distribution of $\hat{\rho}$ does not depend on μ_X, μ_Y, σ_X^2, or σ_Y^2.