Ph. D. Preliminary Exam, Part II (Advanced Exam)
Monday, January 4, 1993

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 Let $X_1, X_2, \ldots$ be i.i.d. nonnegative random variables with a finite second moment. Show that

1 $n^{-\frac{1}{2}} \max_{1 \le i \le n} X_i \to 0$ almost surely.

2 $n^{-\frac{3}{2}} \sum_{i=1}^{n} X_i^3 \to 0$ almost surely.

Problem 2 Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with the $\Gamma(2,1)$ distribution, i.e. the X_i's have density

$$f(x) = x \exp(-x) \quad \text{for } x > 0$$

Show that

$$\limsup_{n \to \infty} \frac{X_n}{\log n} = 1 \quad \text{a.s.}$$

Hint: write down the expression for $P(X_i > a)$.

Problem 3 Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with the exponential distribution with unknown parameter λ. Suppose that λ has as prior distribution the Gamma distribution with parameters α and β. Recall that this distribution has density

$$f(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda) \quad \text{for } \lambda > 0.$$

Consider the censoring model in which the X_i's are (possibly) censored by constants c_i. That is, we do not necessarily observe the X_i's; rather, our data consists of pairs (Z_i, δ_i); $i = 1, \ldots, n$, where

$$Z_i = \min(X_i, c_i) \qquad \text{and} \qquad \delta_i = I(X_i \le c_i).$$

Find the posterior distribution of λ.

Let t_0 be a fixed constant. Find a 95% probability interval for $\exp(-\lambda t_0)$.

Problem 4 Let X be a Poisson random variable with mean θ and let F_θ be the conditional distribution of X given that $X > 0$. Let $Z_1, Z_2, Z_3, \ldots$ be i.i.d. from F_θ.

1 Suppose you observe $Z_1 = Z_2 = 1$, $Z_3 = 2$. What is the likelihood equation for θ? Use the likelihood equation to calculate the MLE of θ to the nearest tenth.

2 Suppose you observe $Z_1 = Z_2 = Z_3 = 1$. What can you say about the MLE of θ?

3 Does there exist an unbiased estimator of θ based on a single observation Z_1? If so, what is this estimator?

4 What is the asymptotic variance of the MLE of θ based on a sample $Z_1, Z_2, \ldots, Z_n$ as $n \to \infty$? (Be explicit.)

Problem 5 Let $X_1, X_2, \ldots$ be i.i.d. random variables with common normal distribution $N(\mu, \sigma^2)$ with $\sigma^2 > 0$. Obtain the asymptotic distribution of $T_n = \frac{s_n^2}{\bar{X}_n}$ where $\bar{X}_n = \frac{1}{n}\sum_1^n X_i$ and $s_n^2 = \frac{1}{n}\sum_1^n X_i^2 - \bar{X}_n^2$.

(Hint: Consider separately the cases $\mu = 0$ and $\mu \neq 0$).

Problem 6 Suppose that $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are i.i.d. with density $\frac{1}{\sigma} f(\frac{\epsilon}{\sigma})$, $-\infty < \epsilon < \infty$, and $\text{Var}(\epsilon_i) = \sigma^2 < \infty$. Suppose also that the vector of observations $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n)$ satisfies the linear model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\epsilon},$$

where $\boldsymbol{X}$, $\boldsymbol{Z}$ are known $n \times p$ and $n \times q$ full-rank matrices and $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are unknown parameters. Let $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ be the least squares estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Define

$$\hat{\boldsymbol{\epsilon}} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}\hat{\boldsymbol{\gamma}}, \quad \hat{\sigma}^2 = \frac{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}}{n - p - q}$$

and

$$\boldsymbol{P}_X = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.$$

1 Show that the distribution of $\boldsymbol{R} = (I - \boldsymbol{P}_X)\boldsymbol{Y}$ does not depend on $\boldsymbol{\beta}$.

2 If the ϵ_i's have normal distribution with mean equal to 0, show that the estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent.

3 If the ϵ_i's have normal distribution with mean equal to 0, show that the likelihood for $(\boldsymbol{\gamma}, \sigma^2)$ based on $\boldsymbol{R}$ is identical (up to a multiplicative constant) to the conditional likelihood of $(\boldsymbol{\gamma}, \sigma^2)$ given $\boldsymbol{P}_X\boldsymbol{Y}$.

Problem 7 A researcher locates four nests of birds of the same species, each of which contains four eggs. For each nest, the four eggs it contains are transferred to the other nests so that each nest ends up with one egg each from nests A, B, C, and D. Thus, each nest contains one egg which was originally laid in it. Suppose that all the eggs hatch, and the response Y = weight gain after 14 days is observed. The origin of the chicks is known. The researcher wants to test the hypothesis that chicks that are hatched and grown in their own nests gain more than chicks from eggs which have been transferred from another nest.

Describe a test of this hypothesis. The researcher feels that it is reasonable to assume that the weight gain of chicks does not depend on which nest they are taken from. However, it is preferable to not assume that different nests produce the same mean weight gain.

Problem 8 An investigator collects data (Y_i, x_i); $i = 1, \ldots, 10$, where the Y_i's are response variables and the x_i's are the observed values of some univariate covariate. The least squares linear regression of Y_i on x_i has residual sum of squares equal to 27. Later, the investigator collects 20 more observations (Y_i', x_i'); $i = 1, \ldots, 20$. The least squares linear fit now has residual sum of squares equal to 64. When all 30 points are fitted by a single linear regression the residual sum of squares equals 99. Test the hypothesis that the two sets of data follow the same linear regression.