# Ph. D. Qualifying Exam (Part II of Written Exam)
# Tuesday, January 4, 2000

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.**

**(a)** State a result giving a classification of probability distributions on the real line into three types (one of which is the class of degenerate distributions) based on properties of the characteristic function.

**(b)** Suppose a random variable $X$ has the same distribution as $Y = aX + b$, for some real numbers $a$ and $b$ with $|a| \neq 1$. Show that $X$ is a degenerate random variable.

---

**Problem 2.**

**(a)** State Fubini's theorem.

**(b)** Let $X$ and $Y$ be independent non-negative random variables. Using Fubini's theorem, and carefully stating any other results you use, show that $E(XY) = E(X)E(Y)$.

---

**Problem 3.** The general model for a two-factor experiment ($n$ observations per cell) where factor $B$ is nested in factor $A$ can be written as

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

where factor $A$ is fixed with effects $\alpha_i$'s and factor $B$ is random. The $\beta_{j(i)}$'s and $\epsilon_{ijk}$'s are assumed to be normally distributed independent random effects with means zero and variances $\sigma_1^2$ and $\sigma_2^2$, respectively.

**(a)** The sums of squares are $SSA = bn\sum_{i=1}^{a}(\bar{Y}_{i..} - \bar{Y}_{...})^2$, $SSB = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij.} - \bar{Y}_{i..})^2$, and $SSE = \sum_{k=1}^{n}\sum_{i=1}^{a}\sum_{j=1}^{b}(Y_{ijk} - \bar{Y}_{ij.})^2$. Calculate the expected mean squares $\mathrm{E}(MSA)$, $\mathrm{E}(MSB)$, and $\mathrm{E}(MSE)$.

**(b)** In this experiment, what hypotheses do you want to test about factors $A$ and $B$ and how do you test these hypotheses? Give your explanations.

**Problem 4.** Consider the following regression model:

$$Y_i = \beta_0 + \beta_0 e^{\beta_1} X_i + \epsilon_i,$$

where the $\epsilon_i$'s are independent and identically distributed normal random variables with mean zero and variance $\sigma^2$.

**(a)** Find the maximum likelihood estimates of $\beta_0$, $\beta_1$, and $\sigma^2$. Denote these estimates by $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}^2$.

**(b)** Find the joint density function for $\hat{\beta}_0$ and $\hat{\beta}_1$. Are $\hat{\beta}_0$ and $\hat{\beta}_1$ unbiased estimates of $\beta_0$ and $\beta_1$?

---

**Problem 5.** Many biologists believe that, within one species of bird, the wing length $y$ is related to body length $x$ as $y = \alpha x^\beta$

**(a)** Formulate a statistical model, including an additive error term $\epsilon$, to be used on a sample of wing and body measurements of birds. In your model, if you assume $\epsilon$ has mean zero and constant variance $\sigma^2$, what difficulty do you see in getting least squares estimates? (Will these estimates be available in closed form?)

**(b)** Reformulate the model so that the standard deviation of the errors is proportional to the mean. A simple way to do this is to use $\epsilon$ in a multiplicative way instead of an additive way. After an appropriate transformation of the data, give the least squares estimates of two parameters. Is the condition of homoskedasticity satisfied by the transformed data?

**(c)** If you assume the multiplicative errors $\epsilon$ in part (b) are roughly normal, then after a transformation the error terms will have a different distribution. Under what conditions will the error terms still be roughly normal? (You may use a series expansion.) Discuss whether your condition is reasonable in scientific work.

---

**Problem 6.** Let $X_1, \cdots, X_n$ be independent and identically distributed with distribution function $F(x)$ and density function $f(x)$. Let $F_n(x)$ denote the empirical distribution function based on the observed sample. Define the naive kernel density estimator of $f(x)$ by

$$f_n(x) = \frac{F_n(x + h_n) - F_n(x - h_n)}{2h_n}$$

Suppose that $f(x)$ is continuous. Show that $f_n(x)$ is a consistent estimator for $f(x)$ provided that $h_n \longrightarrow 0$, but $nh_n \longrightarrow \infty$.

---

**Problem 7.** Suppose that, under $P_\theta$, the random variables $X_1, X_2, \ldots, X_n$ are i.i.d. with distribution $N(\theta, 1)$.

**(a)** Write down the asymptotic distribution of $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ under $P_\theta$.

**(b)** Define

$$Y_n = \begin{cases} 1 & \text{if } \ |\bar{X}_n| \le n^{-\frac{1}{4}}, \\ 0 & \text{if } \ |\bar{X}_n| > n^{-\frac{1}{4}}. \end{cases}$$

For each $\theta$, show that

$$Y_n \xrightarrow{p} \begin{cases} 1 & \text{if } \theta = 0 \\ 0 & \text{if } \theta \ne 0 \end{cases}$$

under $P_\theta$, where $\xrightarrow{p}$ stands for convergence in probability.

---

**Problem 8.** Let $X_1, X_2, \ldots, X_n$ be i.i.d. with the density

$$f(x|\beta) = \frac{2}{\beta} e^{-x/\beta} \quad \text{for } 0 \le x \le \beta \ln 2 \quad (f(x|\beta) = 0 \text{ otherwise}).$$

**(a)** Find a sufficient statistic for $\beta$. (Your statistic should also be minimal, but you are not required to prove this.)

**(b)** Find an ancillary statistic.

**(c)** Is the sufficient statistic you found in part (a) complete? Explain.

---

**Problem 9.** Consider the unbalanced two-way design with the cell counts:

| | | |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 1 | 1 |
| 0 | 0 | 1 |

That is, there are 6 observations $Y_{ij}$ in all, with one observation in each of the indicated cells. Three cells are empty.

Assume the data follow the model $Y_{ij} = \mu + \alpha_i + \tau_j + \epsilon_{ij}$ where $i = 1, 2, 3$ and $j = 1, 2, 3$. (The $\alpha$'s are the row effects and the $\tau$'s are the column effects.) The errors $\epsilon_{ij}$ are uncorrelated with mean zero and a common variance $\sigma^2$.

**(a)** Show that $\mu + \alpha_1$ is **not** estimable.

**(b)** Show that $\alpha_3 - \alpha_1$ is estimable.

**(c)** Find the best linear unbiased estimator (BLUE) of $\alpha_3 - \alpha_1$. (You must prove that your answer is BLUE.)