# Ph. D. Qualifying Exam (Part II of Written Exam)
# Tuesday, January 6, 1998

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 $N$ be a positive integer-valued random variable with a finite third moment. Let $\{X_n\}$ be a sequence of random variables with uniformly bounded second moments.

**(a)** Show that $\sum_{n=1}^{\infty} n^2 P(N \geq n) < \infty$.

**(b)** Using (a), show that $\sum_{n=1}^{\infty} \sqrt{P(N \geq n)} < \infty$.

**(c)** Show that $\sum_{n=1}^{N} X_n$ has a finite first moment. HINT: Apply the Cauchy–Schwarz inequality.

**Problem 2.**

**(a)** State the Lévy continuity theorem involving a pointwise limit of characteristic functions.

**(b)** Show that if $\varphi(t)$ is a characteristic function, then so is $|\varphi(t)|^2$.

**(c)** Suppose that the distribution function $F$ is infinitely divisible. That is, for each $n \geq 1$ there is a distribution function $F_n$ such that $F$ coincides with the $n$-fold convolution $F_n^{*n}$. Show that the characteristic function of $F$ never vanishes. HINT: Apply (b) to the characteristic function of $F_n$ and take limits.

**Problem 3.** Let $X_1, X_2, \ldots X_m$ be i.i.d. random variables with distribution $N(0, \sigma^2)$. Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. $U[-2\pi, 2\pi]$ and be independent of the $X_i$'s.

Consider the kernel $\phi(x_1; y_1) = x_1 \sin(y_1)$ and the two sample $U$-statistic $U_N = \frac{1}{mn} \sum_{1 \leq i \leq m, 1 \leq j \leq n} \phi(X_i; Y_j)$ where $m + n = N$ and $\frac{m}{N} \to \lambda, 0 < \lambda < 1$. Find the asymptotic distribution of $U_N$. (*You may not find the U-statistics two-sample theory very useful for this problem, so proceed directly*)

**Problem 4.** Let $X_1, X_2, \ldots, X_n$ be i.i.d. with density

$$f(x) = \frac{\exp(\alpha x - \beta|x|)}{C(\alpha, \beta)}, \quad -\infty < x < \infty,$$

where $\beta > 0$, $|\alpha| < \beta$, and the normalizing constant is

$$C(\alpha, \beta) = \frac{1}{\beta - \alpha} + \frac{1}{\beta + \alpha}.$$

**(a)** Give a complete sufficient statistic for $\theta = (\alpha, \beta)$. (Prove your answer.)

**(b)** Suppose we know $\beta = 1$. What is the Cramer-Rao lower bound for the variance of an unbiased estimator of $\alpha$?

**(c)** Does there exist an unbiased estimator which achieves the lower bound you found in the previous part?

**Problem 5.** Consider the linear model $Y = X\beta + \epsilon$ where $\epsilon \sim N(0, \sigma^2 I)$, $\beta = (\beta_1, \beta_2, \ldots, \beta_6)'$, and $X$ is the $12 \times 6$ matrix given below:

$$\begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & -1 & 1 & 0 & 0 \\
1 & -1 & 1 & 1 & 0 & 0 \\
1 & -1 & -1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & -1 & 0 & 1 & 0 \\
1 & -1 & 1 & 0 & 1 & 0 \\
1 & -1 & -1 & 0 & 1 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & -1 & 0 & 0 & 1 \\
1 & -1 & 1 & 0 & 0 & 1 \\
1 & -1 & -1 & 0 & 0 & 1
\end{pmatrix}$$

**(a)** Is the parameter $\beta_2$ estimable? How about $\beta_6$? (Prove your answers.)

**(b)** Suppose the data vector is $Y = (8, 3, 4, 1, 9, 7, 2, 2, 7, 8, 6, 3)'$. Compute the vector of fitted values.

**Problem 6.** Let $Y_i$ be observed from a distribution in the exponential family taking the form

$$f(y_i; \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right\}, \qquad i = 1, \ldots, n$$

for some specific functions $a_i(\cdot)$, $b(\cdot)$, and $c(\cdot)$.

**(a)** Show that $E(Y_i) = b'(\theta_i)$ and $\text{Var}(Y_i) = b''(\theta_i)a_i(\phi)$.

**(b)** Write $E(Y_i) = \mu_i$, and assume $g(\mu_i) = \sum_{j=1}^{p} x_{i,j}\beta_j$, $i = 1, \ldots, n$, for some positive integer $p$ and monotonic differentiable function $g$ (i.e. in matrix notation, $g(\mu) = X\beta$, where $g(\mu) = (g(\mu_1), \ldots, g(\mu_n))^T$ is $n \times 1$, $X = (x_{i,j})$ is $n \times p$ and $\beta = (\beta_1, \ldots, \beta_p)^T$ is $p \times 1$)). Let $\ell_i = \log f(y_i; \theta_i, \phi)$. Show that

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{(Y_i - b'(\theta_i))}{\text{Var}(Y_i)} \frac{x_{i,j}}{g'(\mu_i)}, \qquad i = 1, \ldots, n, j = 1, \ldots, p.$$

**Problem 7.**

**(a)** Suppose you have a $2^4$ factorial design in E, F, G, H. Show that you can introduce four other factors a, b, c, and d by confounding such that

- **(i)** the new factors a, b, c, and d are not confounded with two-factor interactions of E, F, G, H.
- **(ii)** an interaction between a new factor and an original factor is confounded with two-factor interactions of the original factors and two-factor interactions of the new factors.
- **(iii)** two-factor interactions of the original factors are not confounded with each other.
- **(iv)** two-factor interactions of the new factors are not confounded with each other.

**(b)** In the design constructed above, if you had four out of eight factors which you thought were most likely to be significant, which letters would you assign them to?

**Problem 8.** Three treatments, A, B, and C, are compared in an experimental design. For each treatment, measurements are recorded successively in time. Suppose that the measurements follow the model:

$$X_i(t) = \mu_i + \epsilon_i(t) + \theta\epsilon_i(t-1),$$

where $\{\epsilon_i(t) : i = 1, 2, 3; t = 0, 1, \ldots, n\}$ are independent and identically distributed random variables with mean zero and variance $\sigma^2$. Let $\bar{X}_i = \sum_{t=1}^{n} X_i(t)/n$ and $\bar{X} = \sum_{i=1}^{3}\sum_{t=1}^{n} X_i(t)/(3n)$.

**(a)** Calculate the means and variances of $\bar{X}_i$ and $\bar{X}$.

**(b)** Let $s^2 = \sum_{i=1}^{3}\sum_{t=1}^{n}(X_i(t) - \bar{X}_i)^2/[3(n-1)]$. Calculate the mean value $\mathrm{E}(s^2)$. Find the range of $\theta$ for which $s^2$ over-estimates $\sigma^2$.

**(c)** How do you test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ in this experiment? What cautions should you make about the conclusions in your analysis?

**Problem 9.** Consider the linear regression model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$ where $\boldsymbol{X}$ is an $n \times p$ full-rank matrix and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 I_n)$. Let $(\hat{Y})$ be the least squares fitted values, $H = X(X'X)^{-1}X'$ the *hat matrix*, and $h_i$ the $i$th diagonal element of $H$. Moreover, let $\hat{\sigma}^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/(n-p)$ and $\hat{\sigma}^2_{(i)}$ be the least squares estimate of $\sigma^2$ in which case $i$ is not used in the modeling. The following three types of residuals may be used for model diagnostics: the ordinary residuals $\hat{\epsilon}_i = Y_i - \hat{Y}_i$, the standardized residuals $r_i = \hat{\epsilon}_i/(\hat{\sigma}\sqrt{1-h_i})$, and the studentized residuals $t_i = \hat{\epsilon}_i/(\hat{\sigma}_{(i)}\sqrt{1-h_i})$.

**(a)** Find the mean values of $\hat{\epsilon}_i$, $r_i$, and $t_i$. What are the (exact or approximate) distributions of $\hat{\epsilon}_i$, $r_i$, and $t_i$?

**(b)** For checking the normality assumption on $\boldsymbol{\epsilon}$, what kind of plots will you use and which type of residuals will you use? Why?

**(c)** Plots of residuals versus fitted values may be used for checking the adequacy of your fitted model. Which type of residuals will you use for this purpose? Why?