Ph.D. Qualifying Exam
Friday–Saturday, January 3–4, 2014

Put your solution to each problem on a separate sheet of paper.

Problem 1. (5166) Assume that two random samples $\{x_i\}$ and $\{y_i\}$ are independently drawn from distributions with means μ_1 and μ_2 and the same variance σ^2.

Observation ID	1	2	3	4	5	6	7	8	9	10
x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
y	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}

(a) Describe how to test hypotheses on the means of the two distributions based on the two samples. State any additional assumptions on the distributions you need for the test. Give your procedure and formulas.

(b) Describe how to test the two distributions using a nonparametric method such as the Wilcoxon Rank-Sum test. Give your procedure and formulas.

(c) Describe how to test the mean difference $(\mu_1 - \mu_2)$ using a randomization distribution method. Give your procedure and compare this method with the other two methods in (a) and (b).

Problem 2. (5166) In a two-way factorial experiment with replicates, consider the following linear model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}, \qquad i = 1, \ldots, a; \;\; j = 1, \ldots, b; \;\; k = 1, \ldots, r,$$

where μ is the overall mean, α_i is the effect for the ith level of factor A, and β_j is the effect for the jth level of factor B. The random errors $\{\epsilon_{ijk}\}$ are assumed to be independent and normally distributed with mean zero and variance $\mathrm{Var}(\epsilon_{ijk}) = \sigma_j^2$.

(a) What are the distributions of $\bar{Y}_{i\cdot\cdot}$ and $\bar{Y}_{\cdot\cdot\cdot}$?

(b) Let $S_A = br \sum_{i=1}^{a} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$, $S_B = ar \sum_{j=1}^{b} (\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$, and

$$S_R = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (Y_{ijk} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j \cdot} + \bar{Y}_{\cdot\cdot\cdot})^2.$$

Find the mean values of S_A and S_R.

(c) Show that the decomposition of variation of this model is:

$$S_D = S_A + S_B + S_R$$

where

$$S_D = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{r} (Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2.$$

Problem 3. (5167) Consider the linear regression model:

$$Y = X\boldsymbol{\beta} + \boldsymbol{\xi},$$

where $Y = (y_1, \ldots, y_n)'$, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and X is an $n \times p$ full-rank matrix. The process $\{\xi_i\}$ is generated by the moving-average model:

$$\xi_i = \epsilon_i - \theta_1 \epsilon_{i-1},$$

where $\{\epsilon_i, i = 0, 1, \ldots, n\}$ are iid $N(0, \sigma^2)$ variables.

(a) Calculate the variance-covariance matrix of $\boldsymbol{\xi}$. Describe how to use the least squares method, the weighted least squares method, and the maximum likelihood method to estimate the coefficients $\boldsymbol{\beta}$ in the model. Give details of the three procedures.

(b) Let $\hat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$. Define $\hat{Y} = X\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\xi}} = Y - \hat{Y}$. Is $\hat{\boldsymbol{\beta}}$ an unbiased estimate of $\boldsymbol{\beta}$? What is the variance matrix of $\hat{\boldsymbol{\beta}}$? Are $\hat{Y}$ and $\hat{\boldsymbol{\xi}}$ independent? Show your reasons.

Problem 4. (5167) Consider a multiple linear regression model with the form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \cdots, n,$$

where $\{\epsilon_i, i = 1, \ldots, n\}$ are independent random variables with mean zero and variance σ_i^2. Define $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$.

(a) Assume that all the observations $\{y_i; i = 1, \cdots, n\}$ are positive values. Describe how to choose a transformation of the response y to make the distribution of y_i approximately normal with constant variance. Give a procedure and reasons.

(b) Describe how to select variables $\{x_1, \cdots, x_p\}$ in this linear model. Give at least two procedures for model selection and compare the procedures.

(c) Give a procedure to evaluate the influence of each case $\{y_i, x_{i1}, x_{i2}, \cdots, x_{ip}\}$ and give your justifications for the procedure.

Problem 5. (5106) Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be a sequence of i.i.d. paired samples. Conditioned on X_i, the value Y_i follows a Poisson distribution with parameter $\lambda_{X_i} = e^{\beta X_i}$. That is,

$$Y_i | X_i, \beta \sim \text{Poisson}(\lambda_{X_i})$$

Our goal is to find the maximum likelihood estimate (MLE) of β with the observations $\{(X_i, Y_i)\}_{i=1}^n$.

(a) Derive an expression for the log-likelihood function

$$l(\beta) = \sum_{i=1}^{n} \log(f(Y_i | X_i, \beta)),$$

such that the MLE is given by

$$\hat{\beta} = \text{argmax}_{\beta}\, l(\beta).$$

(b) Find the expressions for $\dot{l}(\beta)$ and $\ddot{l}(\beta)$, the first and second derivatives of l with respect to β. Verify that $\ddot{l}(\beta) < 0$.

(c) Write out the Newton-Raphson algorithm to find the root of $\dot{l}(\beta)$.

Problem 6. (5106)

Find the maximum likelihood estimate of θ where θ is a parameter in the multinomial distribution:

$$(x_1, x_2, x_3, x_4) \sim M(n; 0.1(1-\theta), 0.4(2-3\theta), 0.1(1+3\theta), \theta)$$

(a) Choose a variable for the missing data and use the EM algorithm for iteratively estimating θ. Let $\theta^{(m)}$ be the current value of the unknown. Derive the mathematical formula to update for $\theta^{(m+1)}$.

(b) Let $L(\theta)$ denote the likelihood with parameter θ. Prove that

$$L(\theta^{(m+1)}) \geq L(\theta^{(m)}).$$

Put your solution to each problem on a separate sheet of paper.

Problem 7. (5326) Let $X_1, X_2, X_3, \ldots$ be iid Uniform$(0,1)$.

(a) Find both $P(X_1^2 + X_2^2 < 1)$ and $P(X_2 - 2X_1 > 0)$.

(b) Find the density of $Y = \dfrac{X_2}{X_1}$.

(c) Find the moment generating function of $Z = 2(X_1 + X_2 + \cdots + X_{20}) - 5$.

(d) What is the approximate distribution of $Z = 2(X_1 + X_2 + \cdots + X_{20}) - 5$? (Give the name of the distribution and the values of any parameters.)

Problem 8. (5326) A fair dodecahedron (a 12-sided solid) has its sides labeled 1, 2, $\ldots$, 12. The dodecahedron is rolled five times. Let $(X_1, X_2, X_3, X_4, X_5)$ denote the observed sequence of values.

(a) Find $P(X_1 < X_2 < X_3 < X_4 < X_5)$.

(b) Find $P(X_1 \le X_2 \le X_3 \le X_4 \le X_5)$.

(c) Find the probability that the sequence $(X_1, X_2, X_3, X_4, X_5)$ contains three consecutive consecutive integers. (In other words, $(X_i, X_{i+1}, X_{i+2}) = (y, y+1, y+2)$ for some values i and y. Some possible sequences $(X_1, \ldots, X_5)$ which satisfy this are $(8, \mathbf{6}, \mathbf{7}, \mathbf{8}, 3)$, $(2, \mathbf{8}, \mathbf{9}, \mathbf{10}, \mathbf{11})$, and $(\mathbf{4}, \mathbf{5}, \mathbf{6}, \mathbf{7}, \mathbf{8})$.)

Problem 9. (5327) The following parts use the family of densities

$$f(x|\theta) = \frac{\theta^2 \log x}{x^{\theta+1}} \quad \text{for} \quad 1 \le x \le \infty\,, \quad \theta > 0\,.$$

(a) Suppose $X_1, \ldots, X_n$ are iid $f(x|\theta)$. Find the MLE for θ.

(b) Suppose that we have two independent random samples: $X_1, \ldots, X_n$ are iid $f(x|\theta)$, and $Y_1, \ldots, Y_m$ are iid $f(x|\beta)$. Find the likelihood ratio test (LRT) of

$$H_0 : \theta = \beta \quad \text{versus} \quad H_1 : \theta \ne \beta\,.$$

(c) Show that the test in part (b) can be based on the statistic $U = \dfrac{\sum_{i=1}^n \log X_i}{\sum_{i=1}^n \log X_i + \sum_{i=1}^m \log Y_i}$.

Problem 10. (5327) Let $X_1, X_2, \ldots, X_n$ be a random sample from the density

$$f(x|\theta) = \theta(\theta+1)\, x(1-x)^{\theta-1} \quad \text{for} \quad 0 < x < 1\,, \quad \theta > 0\,.$$

Answer the following. Justify your answers.

(a) Find a complete sufficient statistic for θ.

(b) Is $\sum X_i$ sufficient for θ?

(c) Find $I(\theta)$, the Fisher information in a single observation from $f(x|\theta)$.

Problem 11. (6346) Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and $(\mathcal{O}, \mathcal{G})$ be a measurable space.

(a) If $f : \Omega \to [0, \infty)$ is a non-negative simple function with $\int f d\mu < \infty$ and B is a set with $\mu(B) = 0$, show

$$\int_B f d\mu = 0.$$

(b) Extend part (a) to include non-negative measurable functions f.

(c) Suppose a non-negative random variable $Y : \Omega \to [0, \infty)$ has $EY < \infty$, and $X : \Omega \to \mathcal{O}$ is a measurable function. Show that there exists a function g such that

$$\int_{X \in A} Y(\omega) d\mu(\omega) = \int_A g(\omega) d\mu_X(\omega) \qquad \text{for any } A \in \mathcal{G}.$$

Problem 12. (6346) For the parts below, you may assume that X_n and X are random variables with all moments finite.

(a) State Markov's inequality.

(b) Define convergence in probability.

(c) Define convergence in L_p.

(d) Show $X_n \xrightarrow{L_p} X \Rightarrow X_n \xrightarrow{P} X$ for finite p.

(e) Prove Lyapunov's inequality: For $0 < q < p$,

$$\left[E\left(|X|^q\right)\right]^{1/q} \leq \left[E\left(|X|^p\right)\right]^{1/p}$$

(f) Show that finite second moment implies finite variance.