Ph.D. Qualifying Exam
Monday–Tuesday, January 5–6, 2015

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

Problem 1. (5106) Let H be an $n \times n$ householder matrix given by

$$H = I_n - 2\frac{vv^T}{v^T v},$$

for any non-zero n-length column vector v . Prove that H is a symmetric, orthogonal, and reflection matrix. That is, H satisfies

$$\text{i) } H = H^T, \quad \text{ii) } HH^T = I_n, \quad \text{and iii) } \det(H) = -1.$$

Problem 2. (5106) Let Y be a continuous random variable with probability density function:

$$Y \sim \alpha_1 f_1(y; \mu_1, \sigma_1^2) + \alpha_2 f_2(y; \mu_2, \sigma_2^2),$$

where f_1 and f_2 are two Gaussian density functions with means μ_1, μ_2 and variances σ_1^2, σ_2^2, respectively. Also, $0 \leq \alpha_1, \alpha_2 \leq 1$, such that $\alpha_1 + \alpha_2 = 1$. Given n *i.i.d.* observations $\{Y_i\}_{i=1}^n$, our goal is to find the maximum likelihood estimate of

$$\theta = (\alpha_1, \mu_1, \sigma_1, \alpha_2, \mu_2, \sigma_2).$$

(a) Use the EM algorithm for iteratively estimating θ. Let $\theta^{(m)}$ be the current values 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

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

Problem 3. (5166) Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample (iid) from a $N(\mu, \sigma^2)$ distribution. Please show the following.

(a) $\bar{X} \sim N(\mu, \sigma^2/n)$.

(b) $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{n-1}^2$.

(c) $\bar{X}$ and s^2 are independent.

Problem 4. (5166) Consider the following linear model involving two factors A and B in an experiment:

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

where μ, $\{\alpha_i, i = 1, \ldots, a\}$, and $\{\beta_j, j = 1, \ldots, b\}$ are unknown constants. The random errors $\{e_{ik}\}$ are iid $N(0, \sigma_e^2)$, $\{\epsilon_{ijk}\}$ are iid $N(0, \sigma_\epsilon^2)$, and the two groups $\{e_{ik}\}$ and $\{\epsilon_{ijk}\}$ are independent.

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

(b) Find $\mathrm{Cov}(Y_{ijk}, Y_{lmn})$ and $\mathrm{Cov}(\bar{Y}_{i\cdot\cdot}, \bar{Y}_{\cdot j\cdot})$.

(c) Let $S_A = br \sum_{i=1}^{a} (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2$. Find $\mathrm{E}(S_A)$.

Problem 5. (5167) Consider the following simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \ldots, n,$$

where $\{e_i, i = 1, \ldots, n\}$ are i.i.d. $N(0, \sigma^2)$ variables. Let $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\}$ denote the least-squares estimates of the unknown parameters $\{\beta_0, \beta_1, \sigma^2\}$.

(a) In the model, suppose that the values of the predictor x_i are replaced by $100x_i$ for $i = 1, \ldots, n$. How are $\{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\}$ affected by this change? How is the t-test of $H_0 : \beta_1 = 0$ affected by this change?

(b) Let $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ be the fitted values and $\hat{e}_i = y_i - \hat{y}_i$ be the residuals for $i = 1, \ldots, n$. What is the sample correlation between the $\hat{e}_i$'s and the y_i's? What is the sample correlation between the $\hat{e}_i$'s and the $\hat{y}_i$'s?

Problem 6. (5167) Consider a multiple linear regression model with the form:

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

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

(a) Let $\hat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$ be the weighted least squares estimate of $\boldsymbol{\beta}$, respectively. Give the expressions for $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$.

(b) Discuss the properties of $\tilde{\boldsymbol{\beta}}$, such as mean, covariance matrix, and distribution of $\tilde{\boldsymbol{\beta}}$.

(c) Compare the two estimates $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$.

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

Problem 7. (5326) Let X be a random variable with density (pdf) given by

$$f(x) = \begin{cases} b\,e^x & \text{for } -1 < x < 2 \\ 0 & \text{otherwise} \end{cases} \qquad \text{where } b = \frac{1}{e^2 - e^{-1}}\,.$$

(a) Find the density (pdf) of the random variable $Y = X^2$.

(b) Find the moment generating function (mgf) of X.

(c) Find the moment generating function of $\frac{1}{2}(X_1 - X_2)$ where X_1, X_2 are iid with the density $f(x)$ given above.
(If you could not answer part (b), then in your answer to part (c) just let $M(t)$ denote the answer to part (b).)

Problem 8. (5326) A type of flowering plant comes in two varieties, red and blue. Plants of the red variety produce only red flowers; plants of the blue variety produce only blue flowers. A large barrel of seeds contains a mixture of the two varieties, with proportion p of the seeds being the red variety. From this barrel, k seeds are selected at random and planted. Each seed grows into a plant which produces a random number of flowers having a Poisson distribution with mean λ. The plants are independent of each other. Let X be the number "red" seeds planted, and let Y be the total number red flowers produced.

(a) Find the following quantities: EY, $\text{Var}(Y)$, and $\text{Cov}(X, Y)$.

(b) Find $P(Y = 0)$ and $P(Y = 1)$. (Simplify your answers if you can.)

Problem 9. (5327) Let $X_1, X_2, \ldots, X_n$ be iid $N(\theta, \theta^2)$ (i.e., normal with mean θ and variance θ^2) with $\theta > 0$ and $T = (\bar{x}, s^2)$ where $\bar{x}$ and s^2 are the sample mean and sample variance of $X_1, \ldots, X_n$. Answer the following. Justify your answers.

(a) Is T minimal sufficient for θ?

(b) Is T complete?

(c) Is $\bar{x}$ an asymptotically efficient estimate of θ?

Problem 10. (5327) Let $X_1, X_2, \ldots, X_n$ be iid Uniform$(0, \theta)$.

(a) Find $\hat{\theta}$, the maximum likelihood estimator (MLE) for θ.

(b) Calculate the mean squared error (MSE) of $\hat{\theta}$.

(c) Find the best unbiased estimator of $\boldsymbol{\theta^2}$. Justify your answer. (You may use results from lecture without proof.)

Problem 11. (6346) Suppose

- $X_1, X_2, \ldots$ are independent and identically distributed uniform$(0, 1)$ random variables
- $X_{(n)}$ is the maximum of the $X_i, i = 1, 2, \ldots, n$
- Y is normal$(0, 1)$
- $Y_i = -Y$ for $i = 1, 2, \ldots$

(a) Does $X_{(n)} \xrightarrow{P} 1$? Prove your answer.

(b) Find the limiting distribution of $n(1 - X_{(n)})$. That is, give the distribution of X where $n(1 - X_{(n)}) \xrightarrow{D} X$.

(c) Does $Y_n \xrightarrow{P} Y$? Prove your answer.

(d) Show $Y_n \xrightarrow{D} Y$.

(e) In general, what property would Y need to have in order for $Y_n \xrightarrow{D} Y \Rightarrow Y_n \xrightarrow{P} Y$?

Problem 12. (6346) Let $(\Omega, \mathcal{F})$ be a measurable space. For $A \in \mathcal{F}$, define δ_x by

$$\delta_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

(a) Show δ_x is a measure for all $x \in \Omega$.

(b) Prove or give a counterexample: $\delta_x \ll \mu$, where μ is the counting measure.

(c) Prove or give a counterexample: $\delta_x \ll \mu$, where μ is Lebesgue measure.

(d) Find $\int f(\omega) d\delta_x(\omega)$ for any real-valued non-negative measurable function f.