Ph. D. Qualifying Exam
Saturday, January 5, 2013

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

Problem 1. (5166) 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 σ^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 θ for which s^2 over-estimates σ^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 2. (5167) Consider the linear regression model:

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

where $Y = (y_1, \ldots, y_n)'$, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, and X is an $n \times p$ full-rank matrix. $\{\epsilon_i, i = 1, \ldots, n\}$ are assumed to be iid $N(0, \sigma^2)$ variables. Let $\hat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$.

(a) Find the distributions of $\hat{\boldsymbol{\beta}}$ and $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(X'X)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

(b) Define $\hat{Y} = X\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\epsilon}} = Y - \hat{Y}$. Find the distribution of $\hat{\boldsymbol{\epsilon}}$.

(c) Show that the statistic

$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(X'X)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/p}{\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}/(n-p)}$$

has an F-distribution.

Problem 3. (5326) A particular parallel system has k independent components, where k is a given positive integer, $k \geq 2$, the i-th component having a lifetime with an exponential distribution of mean λ, for all $i = 1, \ldots, k$. The lifetime of this parallel system is the maximum of the individual lifetimes of its components. What is the probability that the lifetime of this parallel system will be at least λ? Fully justify your answer.

Problem 4. (5327) Consider the parameter space $\mathbf{\Theta} = \mathbb{R}^p \times Sym_+(p, \mathbb{R})$, and the observation space $\boldsymbol{\mathcal{X}} = (\mathbb{R}^p)^n$, where $\mathbf{X} = (X_1, \ldots, X_n) \in \boldsymbol{\mathcal{X}}$ are independent identically distributed random vectors, X_1 having a multivariate $\mathcal{N}_p(\mu, \Sigma)$ distribution. Given the action space $\mathcal{A} = \mathbb{R}^p$ and the loss function $L(\theta, a) = \|\mu - a\|^2$, where $\theta = (\mu, \Sigma)$, compute the risk function $R(\theta, d)$ for the decision rule d, given by $d(\mathbf{X}) = \bar{\mathbf{X}}$. Fully justify your answer.

Problem 5. (6346) Let X be a Poisson random variable with mean $\lambda > 0$.

(a) Find the characteristic function of X.

(b) Show X is infinitely divisible.

(c) A random variable Y is of the Poisson type if $Y = aX + b$ where X is Poisson and $a \neq 0$. Show Y is infinitely divisible.

Problem 6. (5106) Let $X_1, X_2, \cdots, X_n$ be a sequence of *i.i.d.* observations from a logistic distribution with the probability density function

$$f(x|\theta) = \frac{\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}.$$

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

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

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

such that the MLE is given by

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

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

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

Problem 7. (5166) Consider the following unbalanced random one-way model:

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad i = 1, \ldots, k; \; j = 1, \ldots, n_i,$$

where $\{\alpha_i, i = 1, \ldots, k\}$ are iid $N(0, \sigma_\alpha^2)$, $\{\epsilon_{ij}, i = 1, \ldots, k; j = 1, \ldots, n_i\}$ are iid $N(0, \sigma_\epsilon^2)$, and the α_i's and ϵ_{ij}'s are independent. Define

$$\text{SSA} = \sum_{i=1}^k n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^k \frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{\sum_{i=1}^k n_i},$$

$$\text{SSE} = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2,$$

where

$$y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}, \quad \bar{y}_{i\cdot} = \frac{y_{i\cdot}}{n_i}; \qquad y_{\cdot\cdot} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}, \quad \bar{y}_{\cdot\cdot} = \frac{y_{\cdot\cdot}}{\sum_{i=1}^k n_i}.$$

(a) Find the expectations and variances of $\bar{y}_{i\cdot}$ and $\bar{y}_{\cdot\cdot}$. What are the distributions of $\bar{y}_{i\cdot}$ and $\bar{y}_{\cdot\cdot}$?

(b) Show that SSA and SSE are independent.

(c) Find the expectations of SSA and SSE.

Problem 8. (5167) In 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.

(a) Show that the simple regression model can be expressed in the form:

$$y_i = \alpha + \beta_1 (x_i - \bar{x}) + e_i, \quad i = 1, \ldots, n,$$

where $\alpha = \beta_0 + \beta_1 \bar{x}$.

(b) Let $\{\hat{\alpha}, \hat{\beta}_1, \hat{\sigma}^2\}$ denote the least-squares estimates of the unknown parameters $\{\alpha, \beta_1, \sigma^2\}$. Give the expressions of $\{\hat{\alpha}, \hat{\beta}_1, \hat{\sigma}^2\}$

(c) Find the distributions of $\hat{\alpha}$ and $\hat{\beta}_1$. Find the covariance between $\hat{\alpha}$ and $\hat{\beta}_1$.

Problem 9. (5326) Assume $\alpha \in (0, 1)$ is given, and $x_1, x_2, \ldots, x_n$ is a random sample from a probability distribution on $\mathbb{R}^m$ with finite mean vector μ, and positive definite covariance matrix. Derive a large sample $(1-\alpha)100\%$ confidence region $\mathcal{C}_\alpha(x_1, \ldots, x_n)$ for μ. Fully justify your answer.

Problem 10. (5327) Consider the parameter space $\Theta = [0,1] = \mathcal{A}$, $\boldsymbol{\mathcal{X}} = \{0,1\}^n$, $P_\theta(\{\mathbf{x}\}) = \theta^r(1-\theta)^{n-r}$ for $\mathbf{x} \in \boldsymbol{\mathcal{X}}$ where $r = \sum_1^n x_i$, and the loss function $L : \Theta \times \mathcal{A} \to \mathbb{R}$, $L(\theta, a) = c(\theta - a)^2$ $(c > 0)$. Find a Bayes estimator of θ for the prior on Θ having a beta distribution $\mathcal{B}e(\alpha, \beta)$. Fully justify your answer.

Problem 11. (6346) $(\Omega, \mathcal{F}, \mu)$ is a probability space and $(\mathcal{O}, \mathcal{G})$ is a measurable space (you may take them to be the reals and the Borel sets). Define $X : \Omega \to \mathcal{O}$. Let $B \in \mathcal{F}$ be a fixed set. Let $\mu_0(A) = \mu(\{X \in A\} \cap B), A \in \mathcal{G}$.

(a) Show μ_0 is a measure.

(b) Show $\mu_0 \ll \mu_X$, where μ_X is the measure induced by X.

(c) Show there exists a function $g : \mathcal{O} \to \mathbb{R}$ such that

$$\mu(\{X \in A\} \cap B) = \int_A g(x) d\mu_X(x), \quad A \in \mathcal{G}\,.$$

(d) Suppose $\mathcal{O} = \{0,1\}$, $\mathcal{G} = 2^{\mathcal{O}}$ and $X = 0, 1$. Let $\mu_X(1) = p$ and $\mu_X(0) = 1-p$. Let B be fixed. Give an explicit form for g. Writing P in place of μ, show this is a conditional probability.

Problem 12. (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

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

Use an EM algorithm for this estimation. Let $\theta^{(m)}$ be the current values of the unknown. Derive the mathematical formula to update for $\theta^{(m+1)}$.