Ph.D. Qualifying Exam
Monday–Tuesday, January 4–5, 2016

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

Problem 1. (5106) Find the maximum likelihood estimate of θ where θ is a parameter in the multinomial distribution:

$$(x_1, x_2, x_3, x_4) \sim M(n; 0.4(1+\theta), 0.3\theta, 0.6(1-2\theta), 0.5\theta)$$

(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 the parameter θ. Prove that

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

Problem 2. (5106) Let $x = (x_1, \cdots, x_n)$ be a given binary, Markovian sequence. In particular,

$$P(x_1 = 0) = 0.5,\ P(x_1 = 1) = 0.5,$$

and

$$P(x_i = x_{i-1}) = p,\ P(x_i = 1 - x_{i-1}) = 1 - p,$$

$i = 2, \cdots, n$. Let $y = (y_1, \cdots, y_n)$ be a noisy observation of x. That is,

$$y_i = x_i + e_i,$$

with $e_i \sim N(0, \sigma^2)$, $i = 1, \cdots, n$.

Assuming p and σ^2 known, we can use the Maximum A Posteriori method to reconstruct x from y in the following form:

$$\hat{x} = \text{argmax}_{\{x_i\}} \sum_{i=1}^{n} \left(-\frac{(y_i - x_i)^2}{2\sigma^2}\right) + \sum_{i=2}^{n} \log(1_{x_i = x_{i-1}} p + 1_{x_i \neq x_{i-1}} (1-p))$$

Write out a pseudocode for a Dynamic Programming procedure to compute $\hat{x}$ in the computational order of n.

Problem 3. (5166) Let $\{X_1, X_2, \ldots, X_n\}$ be a sample generated from the model:

$$X_i = \epsilon_i + \theta\epsilon_{i-1},$$

where the ϵ_i's are independent and identically distributed random variables with mean zero and variance σ^2. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(a) Find the variance of $\bar{X}$.

(b) Let $s^2 = \dfrac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$. Find $\mathrm{E}(s^2)$.

(c) Compare $\mathrm{Var}(\bar{X})$ with $\mathrm{E}(s^2/n)$ for $\theta = 0$, $\theta > 0$, and $\theta < 0$.

Problem 4. (5166)

Consider a 2^{5-2} fractional factorial design with generators **4=12** and **5=13**. After analyzing the results from the design the experimenter decided to perform a second 2^{5-2} design exactly the same as the first one but switching the signs of column **1** from "+" to"−" (a fold-over design).

(a) What are the advantages and disadvantages of a fractional factorial design compared with a full factorial design?

(b) What are the defining relation and design resolution of the first design? List the confounding patterns of all the two-factor interactions in the first design.

(c) What are the defining relation and design resolution of the combined design? List the confounding patterns of all the two-factor interactions in the combined design.

Problem 5. (5167) In a one-way layout random-effects experiment, consider the following linear model:

$$Y_{ti} = \mu + \tau_t + \epsilon_{ti}, \quad t = 1, \ldots, k; \;\; i = 1, \ldots, n_t,$$

where μ is the overall mean, $\{\tau_t, t = 1, \ldots, k\}$ are iid $N(0, \sigma_\tau^2)$, $\{\epsilon_{ti}, t = 1, \ldots, k; i = 1, \ldots, n_t\}$ are iid $N(0, \sigma_\epsilon^2)$, and the τ_t's and ϵ_{ti}'s are independent.

(a) Let $N = \sum_{t=1}^k n_t$, $\bar{Y}_{t\cdot} = \frac{1}{n_t}\sum_{i=1}^{n_t} Y_{ti}$, and $\bar{Y} = \frac{1}{N}\sum_{t=1}^k \sum_{i=1}^{n_t} Y_{ti}$. Calculate the expected values and variances of $\bar{Y}_{t\cdot}$ and $\bar{Y}$.

(b) Find $\mathrm{Cov}(\bar{Y}_{t\cdot}, \bar{Y})$.

(c) Let $S_T = \sum_{t=1}^k n_t(\bar{Y}_{t\cdot} - \bar{Y})^2$, $S_R = \sum_{t=1}^k \sum_{i=1}^{n_t}(Y_{ti} - \bar{Y}_{t\cdot})^2$, and $S_D = \sum_{t=1}^k \sum_{i=1}^{n_t}(Y_{ti} - \bar{Y})^2$. Show that $S_D = S_T + S_R$. Find the expected values of S_T and S_R.

Problem 6. (5167) Consider the linear regression model with p predictors:

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

where $Y = (y_1, \ldots, y_n)'$, $\boldsymbol{e} = (e_1, \ldots, e_n)'$ with $e_1, e_2, \ldots, e_n$ iid $N(0, \sigma^2)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$, and X is an $n \times (p+1)$ full-rank matrix. Let $\hat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$.

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

(b) Let $\mathbf{1} = (1, 1, \cdots, 1)'$, a column of ones with length n . Define

$$SS_{reg} = (\hat{Y} - \bar{Y}\mathbf{1})'(\hat{Y} - \bar{Y}\mathbf{1}).$$

Find the distribution of SS_{reg} under $H_0 : \boldsymbol{\beta} = 0$.

(c) Show that the statistic

$$F = \frac{SS_{reg}/p}{\hat{\boldsymbol{e}}'\hat{\boldsymbol{e}}/(n-p-1)}$$

has an F-distribution under $H_0 : \boldsymbol{\beta} = 0$.

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

Problem 7. (5326) An urn contains R red balls and G green balls. Ed draws out K balls one by one, randomly and with**OUT** replacement. Whenever Ed draws two red balls in a row he wins a dollar. To be precise, Ed receives a dollar after the i^{th} draw if he draws red balls on draws $i-1$ and i. Note that under these rules, if Ed draw 3 red balls in a row he receives 2 dollars; if he draws 4 red balls in a row he receives 3 dollars, etc. Let X be the amount of Ed's total winnings.

(a) Find EX.

(b) Find $\text{Var}(X)$.

Problem 8. (5326) Consider the Gambler's Ruin problem: a gambler is betting on the tosses of a **fair** coin, winning \$1 when it comes up heads and losing \$1 when it comes up tails. The gambler has an initial fortune of z dollars and a goal of g dollars, i.e., starting with z dollars, he continues playing until he reaches g dollars or goes broke (loses all his money).

(a) Let $\psi(z)$ denote the probability of reaching the goal g as a function of the initial fortune z. Use the Law of Total Probability to find an equation that $\psi(z)$ must satisfy. Then find $\psi(z)$ by obtaining a solution of this equation which satisfies $\psi(0) = 0$ and $\psi(g) = 1$.

(b) **Given that the gambler reached the goal**, what is the probability that his first six tosses had a **total** of four heads and two tails? (Use your answer to part (a). To avoid complications, assume $6 \leq z \leq g-6$ so that the gambler makes at least six tosses.)

Problem 9. (5327) Let X_i be independently distributed as Poisson(βN_i) for $i = 1, 2, \ldots, n$ where β is an unknown positive number and $N_1, N_2, \ldots, N_n$ are known positive constants.

(a) Find the maximum likelihood estimate (MLE) of β. Call it T_1.

(b) Show that T_1 is unbiased for β.

(c) Observe that $T_2 = \dfrac{1}{n}\sum_{i=1}^{n}\dfrac{X_i}{N_i}$ is also an unbiased estimator of β. Which one of T_1 and T_2 has smaller variance? Prove your claim.

Problem 10. (5327) Let $X_1, X_2, \ldots, X_n$ be i.i.d from a distribution with pdf given by

$$f(x \mid \theta) = \begin{cases} \theta^{-c} c x^{c-1} e^{-(x/\theta)^c}, \text{ if } x > 0 \\ 0, \text{ otherwise,} \end{cases}$$

where $c > 0$ is a known constant and $\theta > 0$ is the unknown parameter.

(a) Find a complete sufficient statistic for θ.

(b) Find the minimum variance unbiased estimator (MVUE) for θ. (Hint: Find an unbiased statistic by taking a suitable power of the complete sufficient statistic.)

(c) Find the most powerful (MP) test of size α for testing

$$H_0 : \theta = \theta_0, \quad \text{vs} \quad H_1 : \theta = \theta_1,$$

where $0 < \theta_0 < \theta_1$ are known constants.

Problem 11. (6346) Suppose $(\Omega, \mathcal{F}_0, \mu)$ is a probability space, $\mathcal{F}$ is a σ-field with $\mathcal{F} \subset \mathcal{F}_0$, $X : \Omega \to \mathbb{R}$ is measurable with respect to $\mathcal{F}_0$, and $\mathbb{E}|X| < \infty$.

(a) Define conditional expectation, $\mathbb{E}(X|\mathcal{F})$.

(b) Let $Y = I_A$ and $X = I_B$ for some $A, B \in \mathcal{F}_0$ with $0 < \mu(A) < 1$. Let $\mathcal{G} = \sigma(Y)$. Find $\mathbb{E}(X|\mathcal{G})$.

(c) What happens if A and B are the same set?

Problem 12. (6346) Let $B(t)$ be standard Brownian motion.

(a) Show that the joint density of $B(t_1) = X_1, B(t_2) = X_2, \ldots, B(t_n) = X_n$ where $0 < t_1 < \cdots < t_n$ is a product of univariate normal densities.

(b) Find $E[B(t)|B(s)]$, $0 \le s \le t$.

(c) Define $X(t) = B(1/t), t > 0$, and $X(0) = 0$. Describe the process $X(t)$ in terms of its mean and covariance. Is $X(t)$ Brownian motion? Is it a Gaussian process?