Department of Statistics
University of Wisconsin, Madison
PhD Qualifying Exam Part II
August 29, 2013
1:00-4:00pm, Room 133 SMI

- There are a total of FOUR (4) problems in this exam. Please do a total of TWO (2) problems.
- Each problem must be done in a separate exam book.
- Please turn in TWO (2) exam books.
- Please write your code name and **NOT** your real name on each exam book.

1. Let X and Y be integrable random variables and $\mathcal{F}$ and $\mathcal{G}$ be two σ-fields on a probability space with probability P. For a random variable Z, define $\sigma(Z)$ be the σ-field generated by Z. For parts (a), (b), and (c), suppose that

$$P(B|\mathcal{F}) = P(B|\mathcal{G}) \quad \text{a.s. for any } B \in \sigma(X). \tag{1}$$

   (a) Show that

   $$E[E(Y|X)|\mathcal{F}] = E[E(Y|X)|\mathcal{G}] \quad \text{a.s.}$$

   (b) Under condition (1) with $\mathcal{F} = \sigma(X)$, show that

   $$\sigma(X) \subset \mathcal{G}$$

   and

   $$E(X|\mathcal{F}) = E(X|\mathcal{G}) \quad \text{a.s.}$$

   (c) Under condition (1) with $\mathcal{F} \subset \mathcal{G}$, show that, for any $A \in \mathcal{G}$ and $B \in \sigma(X)$,

   $$P(A \cap B|\mathcal{F}) = P(A|\mathcal{F})P(B|\mathcal{F}) \quad \text{a.s.}$$

   (d) For any event B, show that, for any $A \in \sigma(Y)$,

   $$\int_A P(B|Y)dP = P(A \cap B).$$

   Without further assumptions, can we conclude that $P(B|Y) = I_B$ a.s.? If not, then determine the conditions needed so that this is true. Prove your claims.

2. Fix $s > 0$. Let $\{Z_n\}_{n\geq 1}$ be a collection of independent random variables with

$$P(Z_n = -1) = P(Z_n = 1) = \frac{n^{-s}}{2}, \quad \text{and} \quad P(Z_n = 0) = 1 - n^{-s}.$$

Set $Y_0 = 0$ and for $n \geq 1$ define

$$Y_n = n^s Y_{n-1}|Z_n| + Z_n I\left(Y_{n-1} = 0\right).$$

where $I(A)$ denotes the indicator function for the event A.

(a) Prove that $\{Y_n\}$ satisfies, for every n,

i. $\mathsf{E}\left(|Y_n|\right) < \infty$, and

ii. $\mathsf{E}\left(Y_{n+1} \middle| Y_1, Y_2, \ldots, Y_n\right) = Y_n$.

(b) Determine the values of $s > 0$, if any, for which $Y_n \xrightarrow{P} 0$. Prove both when convergence holds and when it does not.

(c) Determine the values of $s > 0$, if any, for which $Y_n \xrightarrow{a.s.} 0$. Prove both when convergence holds and when it does not.

(d) Determine the values of $s > 0$, if any, for which $Y_n \xrightarrow{L_1} 0$. Prove both when convergence holds and when it does not.

(e) Prove that for any $x > 0$,

$$P\left(\max_{1\leq k\leq n} Y_k \geq x\right) \leq \frac{1}{2x}\left(1 + \sum_{k=1}^{n-1}(k+1)^{-s}(1-k^{-s})\right).$$

3. Suppose we have n pairs of observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ from the model $Y_i \sim N(\beta x_i, \sigma^2)$ where we treat the $x_1, x_2, \ldots, x_n$ values as fixed.

   (a) Find the maximum likelihood estimate of β and of σ^2.

   (b) State the distribution of the maximum likelihood estimate $\hat{\beta}$.

   (c) Suppose σ^2 is known to be equal to 5. We seek a test of the null hypothesis $H_0 : \beta = 0$ versus the alternative $H_A : \beta = 2$. Which of the following designs will result in more power if in fact H_A is true? Explain your choice.

      (i) $x_1 = x_2 = \cdots = x_{10} = 0$; $x_{11} = x_{12} = \cdots = x_{20} = 10$; or

      (ii) $x_1 = x_2 = \cdots = x_5 = 0$; $x_6 = x_7 = \cdots = x_{15} = 12$?

   (d) Suppose instead that the true model is $Y_i \sim N(\gamma_0 + \gamma_1 x_i, \sigma^2)$. Repeat part (c) for $H_0 : \gamma_1 = 0$ versus the alternative $H_A : \gamma_1 = 2$.

   (e) Suppose that it is uncertain whether the true model is $Y_i \sim N(\beta x_i, \sigma^2)$ or $Y_i \sim N(\gamma_0 + \gamma_1 x_i, \sigma^2)$. Suppose σ^2 is known to be equal to 5. We seek a test of the null hypothesis that the slope, i.e., the coefficient of x_i, is zero, versus the alternative that the slope is 2. Which of the two designs, (i) or (ii), in part (c) would you recommend? Explain your choice.

   (f) We can assess the model $Y_i \sim N(\gamma_0 + \gamma_1 x_i, \sigma^2)$ by using the quantity

   $$R_c^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2},$$

   where $\hat{y}_i$ is the fitted value for i-th observation.

   Consider the data set:

x	0	1	2
y	1	2	1

   For this data set $R_c^2 = 0$.

   Suppose instead we fit the model $Y_i \sim N(\beta x_i, \sigma^2)$. We could assess this model by using the quantity

   $$R_{nc}^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2},$$

   where $\hat{y}_i$ is the fitted value from this model. Some authors argue that this quantity is meaningless as a measure of the success of the regression. By referring to the data set above, explain why.

4. Consider the linear regression model,

$$y_i = x_{i1}\beta_1 + \cdots + x_{iK}\beta_K + \epsilon_i, \quad i = 1, \cdots, n,$$

where $\epsilon_1, \cdots, \epsilon_n$ are independent and identically distributed $\mathcal{N}(0, \sigma^2)$ random variables. Assume that σ^2 is known and, without loss of generality, that $\sigma^2 = 1$. Also, assume that the true K-dimensional vector β is of the form $(\beta_1, \cdots, \beta_{k_0}, 0, \cdots, 0)'$ for some $k_0 \leq K$ and $\beta_j > 0$, $\forall j \leq k_0$. Consider model selection criteria of the form

$$C(k, \lambda) \equiv RSS(k) + \lambda k, \quad k = 0, \cdots, K,$$

where $RSS(k)$ is the residual sum of squares of the least squares fit with the first k covariates only, and $\lambda > 0$ represents penalty for over-fitting. The model dimension is chosen by minimizing this criterion over $k = 0, \cdots, K$.

(a) Fit a linear model with the first k covariates to the data and call the fit $\hat{\mathbf{Y}}_k$. Let $\mathbf{P}_k = \mathbf{X}_k(\mathbf{X}_k'\mathbf{X}_k)^{-1}\mathbf{X}_k'$, where $\mathbf{X}_k$ is the design matrix of a linear model with the first k covariates. Show that for $k = k_0, \cdots, K$, the residual random vector $\mathbf{Y} - \hat{\mathbf{Y}}_k$ is $(\mathbf{I} - \mathbf{P}_k)\epsilon$ for the fitted model, where $\mathbf{Y} = (y_1, \cdots, y_n)'$, $\hat{\mathbf{Y}}_k = (\hat{y}_1, \cdots, \hat{y}_n)'$, $\epsilon = (\epsilon_1, \cdots, \epsilon_n)'$, and $\mathbf{I}$ is n dimensional identity matrix.

(b) Find an expression for $C(k, \lambda)$ that only involves ϵ, $\mathbf{P}_k$, λ, and k.

(c) Consider the sequence of random variables, $Z_i \equiv \epsilon'(\mathbf{P}_i - \mathbf{P}_{i-1})\epsilon$, for $i = 1, \cdots, K$. By convention, define $\mathbf{P}_0 = 0$, the zero matrix. Derive the joint distribution of $Z_1, \cdots, Z_K$.

(d) Define $\delta_k(\lambda) = \epsilon'\mathbf{P}_k\epsilon - \lambda k$. Using the sequence of random variables Z_i derived in (c), show that

$$\delta_k(\lambda) = \sum_{i=1}^{k}(Z_i - \lambda).$$

(e) The results in (c) and (d) show that $\delta_k(\lambda)$ is a sequence of partial sums. Define $\hat{k}_\lambda$ as the minimizer of $C(k, \lambda)$ over $\{k_0 \leq k \leq K\}$. How is $\hat{k}_\lambda$ related to the path defined by the sequence $\{\delta_k(\lambda) : k = k_0, \cdots, K\}$?

(f) How would you simulate the distribution of $\hat{k}_\lambda$ to study statistical properties of the model size selected by the criterion $C(k, \lambda)$?