STT 872, 867-868 Fall Preliminary Examination
Wednesday, August 24, 2016
12:30 - 5:30 pm

NOTE: This examination is closed book. Every statement you make must be substantiated. You may do this either by quoting a theorem/result and verifying its applicability or by proving things directly. You may use one part of a problem to solve the other part, even if you are unable to solve the part being used. A complete and clearly written solution of a problem will get a more favorable review than a partial solution.

You must start solution of each problem on the given page. Be sure to put the number assigned to you on the right corner top of every page of your solution.

STT 872, 867-868 Fall Preliminary Examination
Wednesday, August 22, 2018
12:30 - 5:30 pm

1. Let X be distributed according to $P(X = 0) = p^2$, and $P(X = k) = (1-p)(1-p^2)p^{k-1}$, $k = 1, 2, \ldots$ for $p \in (0, 1)$.

(a) Describe all random variables $U(X)$ such that $E_p U = 0$ for all $p \in (0, 1)$.

(b) Describe all estimands $g(p)$ for which UMVUE exist.

(c) Consider $H : p = 0.5$ vs $K : p > 0.5$. Does a UMP test at level $\alpha \in (0, 5/8)$ based on X exist? Explain in as much detail as possible.

(d) Let p have the density $\gamma p^{\gamma-1}, \gamma > 1, p \in (0, 1)$. Find the Bayes estimator of p under the squared loss when $X = 0$.

(e) Let p have the density $\gamma p^{\gamma-1}, \gamma > 1, p \in (0, 1)$. Show that the Bayes estimator of p under the squared loss when $X \geq 1$ is $\left(\frac{2(\gamma+X)+3}{2(\gamma+X)+1}\right)\left(\frac{\gamma+X-1}{\gamma+X+3}\right)$.

2. Let $X_1, ..., X_n$ be i.i.d. from the density $f(x; \theta) = \frac{2x}{\theta^2} I(0 \leq x \leq \theta)$.
(a) Find a complete and sufficient statistics for θ based on $X_1, ..., X_n$.

(b) Let $g(\theta) = \theta^2$. Formulate the two methods of finding UMVUE of a parameter and use one of them to find it for $g(\theta)$.

(c) Find the MRE estimator of θ^2 under the loss function $L(\theta, d) = (1 - d/\theta^2)^2$.

(d) Derive a UMP unbiased test of size $\alpha \in (0, 1)$ for testing $H : \theta = 2$ vs $K : \theta \neq 2$ in as much detail as possible.

(e) Derive the power function of the test in (d).

(f) Propose a consistent estimator of θ.

3. Consider the two-way cross-classified additive model

$$Y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}, k = 1, \cdots, n; i = 1, \cdots, a; j = 1, \cdots, b \tag{1}$$

where Y_{ijk} denotes the k-th replicate in the (i, j)-th cell and all others with usual notations.
(a) If we write the model as $Y = X\beta + \epsilon$, express X and β in terms of design matrix and parameters.

(b) What is the rank of X? How do you estimate β?

(c) Find the test statistic for testing $H_0 : \tau_1 - \tau_2 = 0$ and the corresponding null distribution.

(d) For a modified design matrix X, if the response vector Y follows $N_n(X\beta, \sigma^2 I)$, find the test statistics for testing $H_0 : C\beta = t$ and the corresponding null distribution. Assume C is the $q \times k$ matrix of known constants and k is the dimension of β.

(e) For testing d many contrasts $H_{0j} : a_j^T\beta$, where the vectors $a_1, \cdots, a_d$ are pre-determined, Scheffe's test procedure suggests using the test statistic

$$T = max_a \frac{(a^T\hat{\beta})^T[a^T(X^TX)^{-1}a]^T a^T\hat{\beta}}{s^2},$$

where s^2 is the estimate of σ^2. Find the distribution of T.

(f) In a biological experiment, a scientist is interested in finding important genes for a specific disease. He/she collects 10,000 gene expressions data (continuous measurement) from 100 disease and 100 control subjects. The scientist asked your help to solve the problem using regression based techniques. Write down the model and estimation steps that are appropriate to solve this problem. Be precise in writing your answer.

4. Consider the following model, where Y_{ijk} is the amount of milk produced at farm i on cow j at time t_{ijk}. The cow's weight at time t_{ijk} is denoted by W_{ijk}, and X_{ijk} is a vector of additional covariates. The random effects A, U and V are all normally distributed and are independent of each other.

$$Y_{ijk}|U, V, A \overset{ind}{\sim} N(\mu_{ijk}, \tau^2)$$

$$\mu_{ijk} = X_{ijk}\beta + U_i + V_{ij(1)} + V_{ij(2)}W_{ijk} + A_{ijk}$$

$$\text{cov}(A_{ijk}, A_{lmn}) = \begin{cases} 0 & i \neq l \text{ or } j \neq m \\ \sigma_A^2 \exp(-|t_{ijk} - t_{lmn}|/\phi), & i = l \text{ and } j = m \end{cases}$$

$$\begin{pmatrix} V_{ij1} \\ V_{ij2} \end{pmatrix} \sim MVN(0, G), \qquad U_i \overset{ind}{\sim} N(0, \sigma_U^2)$$

(a) Suppose that we use the maximum likelihood estimation (MLE) to estimate the model parameters. Which model parameters would need to be estimated using a numerical optimizer and which parameters have closed-form expressions for their MLE's if all the other parameters were known?

(b) Derive expressions (and show your work) for the following.

(1) $\text{cov}(Y_{ijk}, A_{ijn}),\ n \neq k$, (2) $\text{cov}(Y_{ijk}, V_{ij1})$,
(3) $\text{cov}(Y_{ijk}, Y_{imn}),\ (m, n) \neq (j, k)$, (4) $\text{cov}(Y_{ijk}, Y_{ijn}),\ n \neq k$,
(5) $\text{var}(Y_{ijk}|A, V)$, (6) $\text{E}[\exp(Y_{ijk})]$.

(c) Suppose you obtained the following parameter estimates: $\hat{\tau}^2 = 2$, $\hat{\sigma}_U^2 = 0.002$, $\hat{\phi} = 4$, $\hat{\sigma}_A^2 = 0.00003$, and

$$\hat{G} = \begin{pmatrix} 1 & 0.05 \\ 0.05 & 0.002 \end{pmatrix}$$

Further, suppose W_{ijk} is the weight measured in pounds, and a typical cow in this study weights between 1,000 and 3,000 pounds. Write down a model which would be more suitable for these data than the model above, where only the random effects, that you feel are important, are included.