STT 872, 867-868 Spring Preliminary Examination
Wednesday, Jan 3, 2018
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.

1. Let X be Poisson(λ) and Y independent of X be Poisson(2λ).

   (a) Find the minimal sufficient statistic for λ based on (X, Y). Is it complete. Justify your answer [5]

   (b) Find the MVUE of $P(X = 0) = e^{-\lambda}$ [8]

   (c) Does it attain the Cramer -Rao bound? [7]

2. $X_1, X_2, \ldots, X_n$ are n samples from a population with density

$$f(x) = \begin{cases} (\theta+1)x^{\theta} \quad 0 < x < 1; \qquad \theta > 0 \\ 0 \text{ otherwise} \end{cases}$$

(a) Find $\hat{\theta}^0$, solution to the likelihood equation [5]

(b) Suppose n=2, and $X_1 = .125, X_2 = .375$, what is your estimated value of θ based on $\hat{\theta}^0$? (log(.125) = -2.08. log(.375) = -.98) [5]

(c) find mle and show that it is consistent [10]

(d) find asymptotic variance of the MLE [8]

3. A random variable T has density, (k is a fixed and known)

$$f_\theta(t) = \frac{kt^{k-1}}{\theta^k} I_{(0,\theta)}(t), \qquad \theta > 0$$

(a) Find the likelihood ratio test, based on one observation, to test $H_0 : \theta \leq 3$ against the alternative $H_1 : \theta > 3$. What is the type 1 error and power. [8]

(b) If $T_1, T_2, \ldots T_n$ are i.id. observations of T, what can you say about the limit as $n \to \infty$ of the power function? [7]

4. Let $X_1, X_2, \ldots$ be i.i.d with mean μ and variance σ^2. Suppose h is a function on $\mathbb{R}$ which is twice differentiable and h'' is continuous at μ and $h'(\mu) = 0$.

   (a) Show that $\sqrt{n}\left[h(\bar{X} - h(\mu)\right]$ converges to 0 in probability [8]

   (b) Show that $n\left[h(\bar{X} - h(\mu)\right]$ is asymptotically distributed as $\frac{1}{2}h''(\mu)\sigma^2 V$, [8] where V has a χ^2_1 distribution

   (c) When $\mu = 1/2$ find the asymptotic distribution of $n\left[\bar{X}(1-\bar{X}) - \mu(1-\mu)\right]$ [8]

5. Suppose that the parameter space in a decision problem is $\Theta = \Theta_1 \times \Theta_2$.

   (a) Show that, if δ_0 is a decision rule such that, for each fixed $\theta_1 \in \Theta_1$, δ_0 is admissible for the decision problem with parameter space $\{(\theta_1, \theta_2) : \theta_2 \in \Theta_2\}$, then δ_0 is admissible when the parameter space is $\Theta_1 \times \Theta_2$ [6]

   (b) $\Theta = \Theta_1 \times \Theta_2 = (0,1) \times (0,1)$. Let X be binomial(n, θ_1) and Y independent of X be binomial(n, θ_2). Consider the problem of estimating (θ_1, θ_2) with loss function $(\theta_1 - a_1)^2 + (\theta_2 - a_2)^2$. Consider the decision rule $\delta_0(x, y) = (1/2, 1/2)$. Use the fact that X/n is admissible to show that that converse of the last item need not hold. [8]

6. Consider the linear regression model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ where X is an $n \times p$ full rank design matrix. Suppose we partition the model as $\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\varepsilon}$ where X_1 is $n \times p_1$ and X_2 is $n \times p_2$. Moreover, $\underline{\varepsilon}$ is normally distributed with mean 0 and variance $\sigma^2 I_n$.

   (a) Show that the least squares estimate of $\underline{\beta}_2$ can be written as $\hat{\underline{\beta}}_2 = C'\underline{y}$ where $C = X_1C_{12} + X_2C_{22}$, C_{12} and C_{22} are matrices of orders $p_1 \times p_2$ and $p_2 \times p_2$, respectively, that depend on X_1 and X_2, and C_{22} is positive definite. [8pts]

   (b) Is the least squares estimate of $\underline{\beta}_2$ obtained in part (a) the best linear unbiased estimate of $\underline{\beta}_2$? Why or why not? [9pts]

   (c) Show that $[I_n - X_1(X_1'X_1)^{-1}X_1']C$ is of rank p_2, where C is defined in part (a). [8pts]

   (d) Let SS_{E_1} be the residual sum of squares for the regression of $\underline{y}$ on X_1 alone. Show that SS_{E_1} and $\hat{\beta}_2$ are not independent. [8pts]

   (e) Find an F-statistic for testing $H_0 : \underline{\beta}_2 = 0$ and construct an α level rejection region. [8pts]

   (f) Assume $\underline{\beta}_2 = (\beta_{21}, \cdots, \beta_{2p_2})^T$. Using Scheffe's method to find $1 - \alpha$ simultaneous confidence intervals for β_{2j} for $j = 1, \cdots, p_2$. [8pts]

   In the following questions, we assume $\underline{\beta}_2$ is a random effect vector with mean zero and variance $\sigma_2 I_{p_2}$, and $\underline{\beta}_2$ and $\underline{\varepsilon}$ are independent.

   (g) Find the ANOVA table corresponding to the above linear model. Please decompose the the total variation into three parts corresponding to β_1, β_2 and the error part, respectively. [8pts]

   (h) Find the expectations of mean squares corresponding to β_1, β_2 and random error $\underline{\varepsilon}$. [8pts]

7. Suppose a study was done to compare two plant genotypes (genotype 1 and genotype 2). Assume that 10 seeds (5 of genotype 1 and 5 of genotype 2) were planted in a total of 4 pots. Suppose 3 genotype 1 seeds were planted in one plot, and the other 2 genotype 1 seeds were planted in another pot. Suppose the same planting strategy was used when planting genotype 2 seeds in the other two pots. The seeds germinated and emerged from the soil as seedlings. After a five-week growing period, each seedling was dried and weighted. Let y_{ijk} denote the weight for genotype i, pot j, seedling k.

   Consider the following linear model

$$y_{ijk} = \mu + \gamma_i + p_{ij} + e_{ijk},$$

   where $p_{11}, p_{12}, p_{21}, p_{22} \sim N(0, \sigma_p^2)$ is independent of the e_{ijk} terms, which are assumed to be iid $N(0, \sigma_e^2)$. This model can be written in the form

$$y = X\beta + Zu + e,$$

   where y is the observation on the dependent variable, X is the fixed-effect design matrix, β is the vector of fixed-effect coefficients, Z is the random-effect design matrix, and u is the vector of random effect coefficients, and e is the vector of residual errors.

   (a) Determine y, X, β, Z, and u. [10]

   (b) Determine $\text{var}(u)$ and $\text{var}(e)$. [10]

   (c) Show that $\text{var}(y)$ is a block diagonal matrix and obtain the first block.
   Identify the covariance of any two observations from the same pot.
   Identify the covariance of any two observations from different pot.
   [10]