考一

2006.9.8

數理統計95年9月

1. (25 pts) Let X be a discrete random variable with sample space N = {1,2,3,...}, the nature number, and cdf $F(x)$.
   (a) Find the distribution of the new random variable $F(X)$, specify the sample space and the corresponding probability.
   (b) Show that EX = $\sum_{x=1}^{\infty}(1-F(x))$.

2. (25 pts) Let (X_{0j}, X_{1j}, X_{2j}), $j = 1, 2, \ldots, n$, be a random sample from a multivariate normal with unknown mean and covariance, (μ_0, μ_1, μ_2) and $\sigma_i^2\begin{bmatrix}\sigma_0^2 & \sigma_{01} & \sigma_{02}\\ \sigma_{01} & \sigma_1^2 & 0\\ \sigma_{02} & 0 & \sigma_2^2\end{bmatrix}$, respectively. Consider regression models $E(X_{ij}\,|X_{0j}=x_{0j}) = \alpha_i + \beta_i x_{0j}$ for $i = 1, 2$. Find a testing procedure to test the equality of β_1 and β_2.

3. (25pts) Let X_i, $i = 1, \ldots, n$ be a random sample from a normal N(μ, 1). Define $g(\mu)=\begin{cases}0 & \mu\neq 0\\ 1 & \mu=0\end{cases}$. Find the consistent estimator of $g(\mu)$. Justify your answer.

4. (25%) Let X_i, $i = 1, \ldots,$ n be a random sample from a population with Beta(λ.,1).
   (a) Find the most powerful level α test of H_0: $\lambda = 1$ vs. H_1: $\lambda = 3$.
   (b) Is there a UMP test of H_0: $\lambda \leq 1$ vs. H_1: $\lambda > 1$? Justify your answer.

Mathematical Statistics: Ph.D. Qualifier Exam September 8, 2006
(100 minutes)

1. The following three problems are related.

   (a) (10pts). Let X_1, X_2, X_3 be i.i.d. Unif([0, 1]) random variables. After observing $X_1 = x_1, X_2 = x_2, X_3 = x_3$, define Y_1, Y_2 to be the results of drawing two numbers at random without replacement from the set $\{x_1, x_2, x_3\}$. Please prove that Y_1 and Y_2 are i.i.d. Unif([0, 1]).

   (b) (10pts). Let $X_1, \ldots, X_n$ be i.i.d. with some probability density function $f_X(x)$ on real line R. Let the conditional distribution of $Y_1, \ldots, Y_k$ (for $k < n$) given $X_1 = x_1, \ldots, X_n = x_n$ be that of k draws without replacement from the set $\{x_1, \ldots, x_n\}$. Please prove that the joint distribution of $Y_1, \ldots, Y_k$ is i.i.d. with probability density function $f_X(\cdot)$.

   (c) (5pts). Below is a generalization of problem (b). Please complete the statement. You do not need to prove it.
   Suppose that $X_1, \ldots, X_n$ are exchangeable on the space $(R, \mathcal{B})$, where $\mathcal{B}$ is the Borel set. Let the conditional distribution of $Y_1, \ldots, Y_k$ (for $k < n$) given $X_1 = x_1, \ldots, X_n = x_n$ be that of k draws without replacement from the set $\{x_1, \ldots, x_n\}$. Then ______________________________.

2. Let $X \sim \text{Exp}(\theta)$ given $\theta > 0$. Namely, the density function of X is $f_\theta(x) = \theta e^{-\theta x} \mathbf{1}_{0 \leq x}$. Consider the two hypotheses H_1: $\theta \leq 1$ versus A_1: $\theta > 1$ and H_2: $\theta \leq 1$ or $\theta \geq 2$ versus A_2: $1 < \theta < 2$. Fix $0 < \alpha < 1$.

   (a) (10pts). Please find the uniformly most powerful level α test for H_1 versus A_1.

   (b) (10pts). Please give the uniformly most powerful level α test for H_2 versus A_2. (You do not need to prove it).

   (c) (5pts). Is it possible to find x value such that the UMP level α test of H_2 rejects H_2 but the UMP level α test of H_1 does not reject H_1?

3. Suppose that $\{(X_i, Y_i)\}_{i=1}^{\infty}$ given θ are i.i.d. with distribution uniform on the disk of radius θ centered at $(0,0)$, that is, the density of (X_i, Y_i) is

$$f(x_i, y_i|\theta) = \frac{1}{\pi\theta^2}\mathbf{1}_{0\le\sqrt{x_i^2+y_i^2}\le\theta}.$$

Let $Z_n = \{(X_i, Y_i)\}_{i=1}^{n}$. We are interested in estimating θ given Z_n. Below are facts you may apply directly:

(i) We say $W \sim A(\alpha,\beta)$ for $\alpha > 0$ and $\beta > 0$ if its density is $f(w) = \alpha\frac{1}{\beta^\alpha}w^{\alpha-1}\mathbf{1}_{0\le w\le\beta}$. Moreover, E$(W)$= $\frac{\alpha}{\alpha+1}\beta$, E$(W^2)$= $\frac{\alpha}{\alpha+2}\beta^2$, and Var$(W)$= $\frac{\alpha}{(\alpha+1)^2(\alpha+2)}\beta^2$.

(ii) We say $W \sim B(\alpha,\beta)$ for $\alpha > 0$ and $\beta > 0$ if its density is $f(w) = \alpha\beta^\alpha\frac{1}{w^{\alpha+1}}\mathbf{1}_{\beta\le w}$. Moreover, E$(W)$= $\frac{\alpha}{\alpha-1}\beta$ for $\alpha > 1$, E(W^2)= $\frac{\alpha}{\alpha-2}\beta^2$ for $\alpha > 2$, and Var(W)= $\frac{\alpha}{(\alpha-1)^2(\alpha-2)}\beta^2$ for $\alpha > 2$.

Please answer the following questions.

(a) (5pts). Please find the maximum likelihood estimate of θ. Denote it by $T(Z_n)$.

(b) (5pts). Please find the distribution of $T(Z_n)$.

(c) (10pts). Please find the asymptotic joint distribution of $T(Z_n)$. That is , find a_n and b_n so that $a_n(T(Z_n) - b_n)$ converges in distribution to a nondegenerate distribution. Please also give the nondegenerate distribution.

(d) (5pts). Please find a complete sufficient statistic.

(e) (5pts). Please find the UMVUE of θ. Denote it by $S(Z_n)$.

(f) (5pts). Please compute mse of $S(Z_n)$. Namely, find the value of $\mathrm{E}_{Z_n}(S(Z_n) - \theta)^2)$.

(g) (10pts). If θ has a prior density $B(a, b)$, that is, θ has a density $f(\theta) = ab^a\frac{1}{\theta^{a+1}}\mathbf{1}_{b\le\theta}$, please find the posterior mean of θ given Z_n. Denote it by $U(Z_n, a, b)$.

(h) (5pts). Consider the case of $b = 0$. It is computed for you that the mse of $U(Z_n, a, 0)$ is $g(a) \equiv \mathrm{E}_{Z_n}(U(Z_n, a, 0)-\theta)^2) = \frac{(2n+a)^2n+(a-1)^2(n+1)}{(2n+1)^2(n+1)(2n+a-1)^2}\theta^2$ and $\frac{\partial g}{\partial a} = \frac{4n(a-2)}{(2n+1)(n+1)(2n+a-1)^3}$. Is the UMVUE the optimal estimate under the mse criterion among the decision family of $\{U(Z_n, a, 0)\}_{a>0}$.

参考(三) 2006.9.8

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a uniform $(0,\theta)$ population where $\theta > 0$ is unknown and let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the corresponding order statistics.
   (a) Show that $X_{(n)}$ is a complete sufficient statistic. (20%)
   (b) Is this family a member of the exponential family? Justify your answer.
   (c) Find d such that $E(dX_{(1)}) = \theta$ and compute $E(dX_{(1)} \mid X_{(n)})$ by first finding the conditional density of $X_{(1)}$ given $X_{(n)}$.
   (d) Show that $\frac{X_{(1)}}{X_{(n)}}$ is an ancillary statistic and $E\left(\frac{X_{(1)}}{X_{(n)}}\right) = \frac{E(X_{(1)})}{E(X_{(n)})} = \frac{1}{n}$.

2. Let $X_1, X_2, \ldots, X_n$ be n independent identically distributed Bernoulli random variables with $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$ where $0 < p < 1$ and $n \geq 2$.
   Define $\hat{p}_2 = \frac{X_1 + X_2}{2}$ and $\hat{p}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$. (25%)
   (a) Find an unbiased estimator that is a function of $\hat{p}_2$ of $p(1-p)$.
   (b) Compute $P(X_i = 1 \mid \sum_1^n X_i = k)$ and $P(X_i + X_2 = 1 \mid \sum_1^n X_i = k)$
   (c) Compute $E(T \mid \sum_1^n X_i)$ where T is denoted as the answer of (a).
   (d) Find the UMVU estimator in terms of $\hat{p}_n$ of $p(1-p)$.

3. Suppose we have the usual simple linear regression model $y_i = \beta_0 + \beta_1 x_i + e_i$, where e_i are iid $N(0, \sigma^2)$, and x_i are fixed constants. Instead of the usual least squares estimate, suppose we estimate the slope parameter β_1 with $\tilde{\beta}_1 = \frac{\sum_{i=1}^n (z_i - \bar{z}) y_i}{\sum_{i=1}^n (z_i - \bar{z}) x_i}$, where z_i are also fixed constants. (35%)
   (a) Compute the bias in the estimator $\tilde{\beta}_1$.
   (b) Find the variance of the estimator $\tilde{\beta}_1$.
   (c) The usual least squares estimator $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$ is the BLUE for the slope parameter β_1. What does each letter in "BLUE" mean?
   (d) Under what circumstances, other than $z_i = x_i$, will $\tilde{\beta}_1$ also be BLUE?
   (e) If z_i are random variables which are independent of both y_i and x_i, will $\tilde{\beta}_1$ be a good or poor estimator of β_1? Why?