Statistics Ph.D. Qualifying Exam: Part II

November 18, 2006

Student Name: ______________________________

1. Answer 8 out of 12 problems. Mark the problems you selected in the following table.

Problem	1	2	3	4	5	6	7	8	9	10	11	12
Selected												
Scores												

2. Write your answer right after each problem selected, attach more pages if necessary.
3. Assemble your work in right order and in the original problem order.

1. Let (X, Y) be a random vector with joint pdf

$$f(x,y) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} x^{a-1} y^{b-1} (1-x-y)^{c-1}, \quad 0 < x < 1, 0 < y < 1, 0 < x+y < 1.$$

where $\Gamma(t)$ is the gamma function. $[\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx.]$

(a) Find the marginal p.d.f. of X and Y.

(b) Find the joint p.d.f. of $W = X + Y$ and $V = X$.

(c) Find the marginal p.d.f. of $W = X + Y$.

2. Let $X_1, \ldots, X_n$ be a sample of size n from $U(0,1)$.

   (a) Find the p.d.f. of $\frac{1}{n}\sum_{i=1}^{n}(-\log(X_i))$.

   (b) Using (a) or other method, find the p.d.f. of $(\prod_{i=1}^{n} X_i)^{\frac{1}{n}}$.

   (c) Using (a) and (b) or other method, show that if n is large, $(\prod_{i=1}^{n} X_i)^{\frac{1}{n}}$ has approximately a $N(\mu, \frac{1}{n}\sigma^2)$ distribution. Find μ and σ^2.

3. Let X_i, $i = 1, 2, \cdots, n$ be iid random variables with $N(\theta, \theta^2)$ distribution, where $\theta > 0$. Let

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

(a) Show that both $\bar{X}$ and cS are unbiased estimator of θ, where $c = \frac{\sqrt{n-1}\,\Gamma((n-1)/2)}{\sqrt{2}\,\Gamma(n/2)}$.

(b) For a constant a, compute $\mathrm{Var}(a\bar{X} + (1-a)(cS))$.

(c) Find the value of a that produces the estimator with minimum variance among the unbiased estimators of the form $\mathrm{Var}(a\bar{X} + (1-a)(cS))$.

(d) Is $T = (\bar{X}, S^2)$ a complete sufficient statistics for θ ? Justify your answer.

4. Let X_i, $i = 1, 2, \cdots, n$ be iid random variables with $N(\theta, 1)$ distribution, where θ is an unknown parameter. Consider testing $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$, where θ_0 is a known fixed constant.

   (a) Derive the maximum likelihood estimator for θ under $H_0 : \theta \leq \theta_0$.

   (b) Show that the likelihood ratio test for $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ is to reject H_0 when
   $$\bar{X} > k,$$
   for some constant k.

   (c) Find the constant k above so that likelihood ratio test is of size α.

   (d) Show that the above likelihood ratio test is a UMP test.

5. For each of the following pdfs, let X_i, $i = 1, 2, \cdots, n$ be a random sample from that distribution. In each case, find the best unbiased estimator of θ^r, where $r < n$.

   (a) $f(x;\theta) = \frac{1}{\theta}, \quad 0 < x < \theta.$

   (b) $f(x;\theta) = \theta^x(1-\theta)^{1-x}, \quad x = 0, 1; 0 < \theta < 1.$

6. Suppose a random sample of size 2, X_1 and X_2, is observed from a distribution whose probability mass function under H_0 and H_1 is given by

x	1	2	3	4	5	6	7
$f(x\|H_0)$	0.01	0.01	0.01	0.01	0.01	0.01	0.94
$f(x\|H_1)$	0.06	0.05	0.04	0.03	0.02	0.01	0.79

According to the Neyman-Pearson Lemma, the most powerful test for H_0 vs. H_1 is to reject H_0 when

$$\lambda = \frac{L_0}{L_1} = \frac{f(x_1|H_0) \times f(x_2|H_0)}{f(x_1|H_1) \times f(x_2|H_1)} \leq k,$$

for some constant k. Assume that $k = 0.2$ is chosen for the test. Using the tables provided below, answer the following questions:

(a) Find the rejection region. That is, list the set of (x_1, x_2) for which H_0 is rejected.

(b) Find the Type I error probability α.

(c) Compute the probability of Type II error for the above test.

L_0	1	2	3	4	5	6	7
1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0094
2	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0094
3	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0094
4	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0094
5	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0094
6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0094
7	0.0094	0.0094	0.0094	0.0094	0.0094	0.0094	0.8836

L_1	1	2	3	4	5	6	7
1	0.0036	0.0030	0.0024	0.0018	0.0012	0.0006	0.0474
2	0.0030	0.0025	0.0020	0.0015	0.0010	0.0005	0.0395
3	0.0024	0.0020	0.0016	0.0012	0.0008	0.0004	0.0316
4	0.0018	0.0015	0.0012	0.0009	0.0006	0.0003	0.0237
5	0.0012	0.0010	0.0008	0.0006	0.0004	0.0002	0.0158
6	0.0006	0.0005	0.0004	0.0003	0.0002	0.0001	0.0079
7	0.0474	0.0395	0.0316	0.0237	0.0158	0.0079	0.6241

$\lambda = L_0/L_1$	1	2	3	4	5	6	7
1	0.0278	0.0333	0.0417	0.0556	0.0833	0.1667	0.1983
2	0.0333	0.0400	0.0500	0.0667	0.1000	0.2000	0.2380
3	0.0417	0.0500	0.0625	0.0833	0.1250	0.2500	0.2975
4	0.0556	0.0667	0.0833	0.1111	0.1667	0.3333	0.3966
5	0.0833	0.1000	0.1250	0.1667	0.2500	0.5000	0.5949
6	0.1667	0.2000	0.2500	0.3333	0.5000	1.0000	1.1899
7	0.1983	0.2380	0.2975	0.3966	0.5949	1.1899	1.4158

7. Consider the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y}$ is $(n \times 1)$, $\boldsymbol{\epsilon}$ is $(n \times 1)$, $\boldsymbol{X}$ is $(n \times p)$, and where $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}$.

(a) State and prove the Gauss Markov Theorem.

(b) Explain how you would extend the result in (a) to the case where $\text{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is a known symmetric positive definite matrix.

8. Suppose (X, Y) have a trinomial distribution with parameters n, θ_1, θ_2, where n is the numbers of trials, and $0 \leq \theta_1 + \theta_2 \leq 1$. That is,

$$P(X = x, Y = y|\theta_1, \theta_2) = \frac{n!}{x!y!(n-x-y)!}\theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y}.$$

Suppose we put the Dirichlet density

$$\pi(\theta_1, \theta_2) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)}\theta_1^{a-1}\theta_2^{b-1}(1-\theta_1-\theta_2)^{c-1} \quad 0 \leq \theta_1 \leq 1, 0 \leq \theta_2 \leq 1; 0 \leq \theta_1+\theta_2 \leq 1,$$

as prior for (θ_1, θ_2).

(a) Let r, s, t be positive. Find $E\theta_1^r \theta_2^s (1-\theta_1-\theta_2)^t$.

(b) Find the posterior distribution of $((\theta_1, \theta_2)|X = x, Y = y)$. Is the Dirichlet distribution a conjugate prior for this problem?

(c) For any function h of θ_1 and θ_2, defining the Bayes estimator of $h(\theta_1, \theta_2)$ by $d_B(\mathbf{X}) = E(h(\theta_1, \theta_2)|\mathbf{X})$, find the Bayes estimators of θ_2, and $\theta_1\theta_2(1-\theta_1-\theta_2)$.

9. Let X and Y be random variables such that $Y|X = x \sim$ Poisson(λx), and X has density

$$f_X(x) = \frac{\theta^\theta x^{\theta-1} e^{-\theta x}}{\Gamma(\theta)}, \quad x \geq 0.$$

   (a) Prove that

      i. $E(Y) = \lambda$ and $Var(Y) = \lambda + \theta\lambda^2$.

      ii. Y has density

$$f_Y(y; \lambda) = \frac{\Gamma(\theta + y)\lambda^y \theta^\theta}{\Gamma(\theta) y! (\theta + \lambda)^{\theta+y}}, \quad y = 0, 1, 2, \ldots$$

   (b) Now suppose that $Y_1, \ldots, Y_n$ are independent random variables from the distribution given above, with Y_i having mean λ_i, and $\log(\lambda_i) = \beta z_i$, where z_i's are known covariates, $i = 1, \ldots, n.$, and assume that $\theta = 1$. Write a Fisher scoring algorithm for computing the MLE of β, and discuss its properties.

10. Let $\{X_1, \ldots, X_n\}$ be a random sample from a normal density with mean 0 and variance σ^2. Define the quadratic forms $Q_i = (X_1, \ldots, X_n)A_i(X_1, \ldots, X_n)', i = 1, 2$, where the A_i's are symmetric matrices of given real numbers. Let $rank\ A_i = m_i, i = 1, 2$.

   (a) Show that if $A_1A_2 = 0$, then Q_1 and Q_2 are independently distributed of one another.

   (b) Show that if $A_i^2 = A_i$, then $Y_i = Q_i/\sigma^2$ is distributed as a central chi-square random variable with degrees of freedom m_i, $i = 1, 2$, respectively.

   (c) Using results from (a) and (b), show that if $m_i + m_2 = n$, then $(Y_i = Q_i/\sigma^2, i = 1, 2)$ are independently distributed as central chi-square random variables with degrees of freedom m_i respectively.

11. Let $\{(X_{i,1}, \ldots, X_{i,n_i}), i = 1, 2, 3\}$ be independent random samples from normal distributions with means μ_i and variance $2^i\sigma^2 (i = 1, 2, 3)$ respectively. Assume that $\{\mu_1 = \alpha + \beta + \gamma, \mu_2 = 2\alpha + \beta - \gamma, \mu_3 = \alpha + 2\beta - \gamma\}$.

   (a) Obtain the MLE (Maximum Likelihood Estimator) of $\{\alpha, \beta, \gamma, \sigma^2\}$. What are the optimal properties of these MLE's?

   (b) Derive the 0.95% confidence interval for $\alpha - \beta$.

   (c) Assuming a non-informative prior $P\{\alpha, \beta, \gamma, \sigma^2\} \propto \{\sigma^2\}^{-1}$, derive the 95% HPD (Highest Posterior Density) Bayesian interval for $\alpha - \beta$. How is this Bayesian HPD interval compared with the confidence interval derived in (b)?

12. Let $\{(X_i, Y_i), i = 1, \ldots, n\}$ be a random from a bivariate normal distribution with means $EX = \mu_1$ and $EY = \mu_2$ and with variances and covariance as $Var(X) = \sigma_1^2$, $Var(Y) = \sigma_2^2$ and $Cov(X, Y) = \rho\sigma_1\sigma_2$.

   (a) Derive the size-α likelihood ratio testing procedure for testing $H_0 : \mu_1 = \mu_2$ against the alternative hypothesis $H_1 : \mu_1 \neq \mu_2$.

   (b) Derive the probability distribution of your test statistic under H_0.

   (c) Derive the 100(1-α)% confidence interval for $\mu_1 - \mu_2$. If you use this confidence interval to test the hypothesis H_0, how is it compared with the procedure derived in (a)?