Statistics Ph.D. Qualifying Exam: Part I

November 11, 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 and Y have joint pdf

$$f(x, y) = \frac{1}{4}(x + 2y), \quad 0 \le x \le 2, \quad 0 < y < 1.$$

(a) Find the marginal pdf of X.

(b) Find the pdf of $Z = (2X - 1)^2$.

(c) Find the pdf of $W = X + Y$.

2. Let X and Y be two independent random variables following exponential distributions with means a and b, respectively. Define the $Z = \min(X, Y)$ and

$$W = \begin{cases} 1, & \text{if } Z = X \\ 0, & \text{if } Z = Y. \end{cases}$$

(a) Find $P(W = 1)$.

(b) Find the p.d.f. of Z.

(c) Find the joint distribution of W and Z. That is, $P(Z \leq z, W = w)$, where $z > 0$ and $w = 0, 1$.

(d) Are W and Z independent ? Justify your answer.

3. Let B and C be two iid random variables with $U(0, 1)$ distribution. Find the probability that $x^2 + Bx + C = 0$ has two real roots.

4. Let $X_1, \ldots, X_n$ be a random sample from a Poisson probability distribution

$$f(x;\theta) = \frac{\theta^x e^{-\theta}}{x!}, x = 0, 1, 2, \cdots,$$

where $\theta > 0$. Let $\bar{X}$ and S^2 be the sample mean and sample variance, respectively.

(a) Show that $\bar{X}$ and S^2 are unbiased estimators of θ.

(b) Show that $E(S^2|\bar{X}) = \bar{X}$ and $Var(S^2) > Var(\bar{X})$. Justify your argument.

5. Let $X_1, \ldots, X_n$ be a random sample from a geometric probability distribution $f(x;\theta) = \theta(1-\theta)^{x-1}$, where $0 < \theta < 1$ and $x = 1, 2, 3, \cdots$.

   (a) Show that X_1 is an unbiased estimator of $1/\theta$.

   (b) Find the Cramér-Rao lower bound for the variance of unbiased estimators of $1/\theta$.

   (c) Find the UMVUE of $1/\theta$.

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from the following distribution

$$f(x|\theta) = \begin{cases} \theta e^{2\theta x}, & x < 0 \\ \frac{\theta}{2} e^{-\theta x}, & x \geq 0, \end{cases}$$

where $\theta > 0$ is unknown parameter. Derive the UMP test for $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$.

7. Let $\{X_1, X_2\}$ be distributed as a 2-dimensional multinomial random vector with parameters $\{n; p_1, p_2\}$, where $0 < p_i, i = 1, 2,\ p_1 + p_2 < 1$ and n an integer. Put $Y_i = \log\{X_i/(n - X_1 - X_2)\}, i = 1, 2$.

   (a) Assuming that n is very large, derive the asymptotic joint pdf (Probability Density Function) of $\{(X_1 - np_1)/\sqrt{n}, (X_2 - np_2)/\sqrt{n}\}$.

   (b) Assuming that n is very large, derive the asymptotic pdf of $Y_i, (i = 1, 2)$. What are the asymptotic mean value and the asymptotic variance of $Y_i (i = 1, 2)$? What is the asymptotic covariance between Y_1 and Y_2 ?

   (c) Assuming $p_1 = p_2$, derive the asymptotic distribution of $Z = (Y_1 - Y_2)^2$.

8. Let $\{X_1, \ldots, X_m\}$ be a random sample from the normal distribution with mean μ_1 and variance $4 \times \sigma^2$. Let $\{Y_1, \ldots, Y_n\}$ be a random sample from the normal distribution with mean μ_2 and variance $9 \times \sigma^2$. Define the statistics $\{\bar{X} = \frac{1}{m}\sum_{i=1}^m X_i, \bar{Y} = \frac{1}{n}\sum_{j=1}^n Y_j, S_X^2 = \sum_{i=1}^m (X_i - \bar{X})^2, S_Y^2 = \sum_{j=1}^n (Y_j - \bar{Y})^2\}$.

   (a) Show that $\{\bar{X}, \bar{Y}, S_X^2, S_Y^2\}$ is a set of jointly sufficient but not complete statistics for $\{\mu_1, \mu_2, \sigma^2\}$. Obtain a set of jointly sufficient and complete statistics for $\{\mu_1, \mu_2, \sigma^2\}$.

   (b) Derive the UMVUE (Uniformly Minimum-Varianced Unbiased estimator) of $\mu_1 - \mu_2$. What is the UMVUE of σ^2 ?

   (c) Assuming a non-informative prior $P(\mu_1, \mu_2, \sigma^2) \propto \{\sigma^2\}^{-1}$ for $\{\mu_i, i = 1, 2, \sigma^2\}$, derive the Bayesian estimator of $\mu_1 - \mu_2$ and the Bayesian estimator of σ^2 under squared loss function.

9. Let $\{X_1, \ldots, X_m\}$ be a random sample from the population with density $f(x; \beta_1, \sigma_1^2) = \frac{1}{\sigma_1^2} e^{-\frac{1}{\sigma_1^2}(x-\beta_1)}, x > \beta_1$. Let $\{Y_1, \ldots, Y_n\}$ be a random sample from the population with density $f(y; \beta_2, \sigma_2^2) = \frac{1}{\sigma_2^2} e^{-\frac{1}{\sigma_2^2}(y-\beta_2)}, y > \beta_2$, independently of $\{X_1, \ldots, X_m\}$.

   (a) Derive the size α likelihood ratio testing procedure for testing $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$, $(0 < \alpha < 1)$.

   (b) Derive the power function of your test.

   (c) Derive a $100(1-\alpha)$ % confidence interval for $\theta = \sigma_1^2/\sigma_2^2$. If you use this confidence interval to test the above hypothesis H_0, how is it compared with the procedure derived in (a)?

10. Let $X_1, \dots,$ be a sequence of independent random variables with the logistic (cumulative) distribution function

$$F(x) = \frac{1}{1+e^{-x}}, \qquad -\infty < x < \infty.$$

Let N be a geometric random variable with probability function

$$P(N = n) = p(1-p)^{n-1}, n = 1, 2, \dots$$

Assume that N is independent of the X's.

(a) Prove that $X_{(N)} + \log p$ also has a logistic distribution function, where $X_{(N)} = \max(X_1, \dots, X_N)$.

(b) Using the fact that the logistic distribution given above is symmetric about 0, what does the above result say about the distribution of $\min(X_1, \dots, X_N)$?

11. Let $Y_1, \dots, Y_n$ be random variables satisfying the linear model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \dots, n,$$

where $x_1, \dots, x_n$ are fixed constants, and $\epsilon_1, \dots, \epsilon_n$ are i.i.d Normal$(0, \sigma^2)$. Assume that σ^2 is unknown.

(a) Find the jointly sufficient statistics for $(\beta_0, \beta_1, \sigma^2)$

(b) Find the MLE of β_1 and σ^2 and show whether or not they are biased or unbiased.

(c) Construct a likelihood ratio test of $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ at level of significance α. Give the critical value of the rejection region in terms of a percentile of a tabulated distribution.

(d) Explain how you would compute the power of the test.

12. (a) Write a careful proof of the Neyman-Pearson Lemma.

(b) Let $X_1 \ldots, X_{2n+1}$ be a random sample from a population with density

$$f(x;\theta) = e^{-(x-\theta)}, \quad x \geq \theta.$$

You wish to test $H_0 : \theta = 0$ versus $H_1 : \theta = 1$

i. For each of the following procedures, compute the probability of Type I and the power of the tests at $\theta = 1$.

A. Reject H_0 if $X_{(1)} > .5$.

B. Reject H_0 if $0.5 < X_{(n)} < 1$

ii. Construct the most powerful test of these hypotheses, when you set level of significance at 0.05.