# Statistics Ph.D. Qualifying Exam: Part II

November 14, 2009

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. **Do not** write your answers on the back.

3. Assemble your work in right order and in the original problem order. (Including the ones that you do not select)

1. Let $X_1 \ldots, X_n$ be a sample of size $n$ from $U(0, \theta)$ where $\theta$ is unknown parameter.

   (a) Show that $Y_i = -\ln(X_i/\theta)$ has an exponential distribution with unit mean.

   (b) Show that if $n$ is large, the "geometric mean" of $X_i$,

$$(\Pi_{i=1}^n X_i)^{\frac{1}{n}}$$

   has approximately a $N(\mu, \sigma^2)$ distribution. Find $\mu$ and $\sigma^2$. [hint: use the relation $x = e^{ln(x)}$.]

2. Let $X_1, \ldots, X_n$ be a random sample from the geometric distribution

$$f(x;\theta) = \theta(1-\theta)^x I_{\{0,1,2,\ldots\}}(x)$$

where $0 < \theta < 1$.

(a) Derive $E(X_i)$ and $Var(X_i)$.

(b) Find the probability distribution of $S = X_1 + X_2 + \cdots + X_n$.

(c) Find the Cramér-Rao lower bound for the variance of unbiased estimators of $(1-\theta)/\theta^2$.

3. Consider the regression model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i, i = 1, \ldots, n$$

where $x_i$'s are known and $\epsilon_i$ i.i.d with $N(0, \sigma^2)$ for $i = 1, \ldots, n$.

(a) Derive the MLE $\hat{\beta}_i, i = 0, 1, 2$ and $\hat{\sigma}^2$.

(b) Find the asymptotic distribution of $e^{\hat{\beta}_0 + \hat{\beta}_1}$.

4. Let $X_{(1)} \leq \cdots \leq X_{(n)}$ be the order statistics for a random sample from a continuous distribution with cumulative distribution function $F(x)$ and density $f(x)$. Define $Y_i = F(X_{(i)})$ and $U_i, i = 1, \ldots, n$, by

$$U_i = \frac{Y_i}{Y_{i+1}}, i = 1, \ldots, n-1,$$

and

$$U_n = Y_n.$$

(a) Show that if $X$ has c.d.f. $F(x)$, $Y = F(X) \sim U(0,1)$.

(b) Find the joint p.d.f. of $Y_i$, $i = 1, \ldots, n$.

(c) Find the joint p.d.f. of $U_i$, $i = 1, \ldots, n$.

(d) Show that $U_1, U_2^2, \ldots, U_n^n$ are i.i.d uniform $(0,1)$ random variables.

5. Let $X_1 \ldots, X_n$ be a sample of size $n$ from $U(0, \theta)$ where $\theta$ is unknown parameter. Let $\hat{\theta}_{mme}$ = moment estimator of $\theta$; $\hat{\theta}_{mle}$ = maximum likelihood estimator of $\theta$.

   (a) Derive $\hat{\theta}_{mme}$ and $\hat{\theta}_{mle}$.

   (b) Find the asymptotic distributions of $\hat{\theta}_{mme}$ and $\hat{\theta}_{mle}$.

   (c) Compare $\hat{\theta}_{mme}$ and $\hat{\theta}_{mle}$ as an estimator of $\theta$.

6. Let $X_1, \ldots, X_m$ be a random sample from

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} x^{-(1-\theta)/\theta} & , \quad 0 < x < 1, \quad \theta > 0 \\ 0 & , \quad \text{otherwise.} \end{cases}$$

(a) Find the uniformly most powerful (UMP) test of size $\alpha$ for testing $H_0 : \theta \leq 1$ vs. $H_1 : \theta > 1$.

(b) Explain how to find the power function of the UMP test.

7. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from Exponential($\lambda$) and Exponential($\mu$) populations respectively.

   (a) Construct a likelihood ratio test of

$$H_0 : \lambda = \mu \quad \text{versus} \quad H_1 : \lambda \neq \mu.$$

   (b) Give the critical values of this test in terms of percentiles of one of the standard distributions.

   (c) Is the likelihood ratio test, uniformly most powerful? Why or Why not?

8. Let $X_1, X_2, \ldots X_n$ be a independent Normal$(\mu, \sigma^2)$ random variables.

   (a) Prove that $\bar{X}$ and $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$ are independent.

   (b) Prove that $S^2/\sigma^2$ has a chi-squared distribution.

   (c) Let $g(x)$ be a continuous function of $x$. Find $C$ such that $C(g(\bar{X}) - g(\mu))^2/S^2$ has an F distribution. What are the degrees of freedom?

9. Let $X_1, \ldots, X_n$ be a random sample from a population with density $f(x|\theta)$, and let $\xi(\theta)$ represent a prior density on $\theta, \theta \in \Omega$.

   (a) Using loss function

   $$L(\tau(\theta) - a) = w(\theta)(\tau(\theta) - a)^2,$$

   where $\tau(\theta), w(\theta)$ are functions of $\theta$, with $w(\theta) > 0$, prove that the Bayes estimator of $\tau(\theta)$ is given by

   $$d_B(\mathbf{x}) = \frac{E[w(\theta)\tau(\theta)|\mathbf{X} = \mathbf{x}]}{E[w(\theta)|\mathbf{X} = \mathbf{x}]}.$$

   (b) If $L(\tau(\theta)-a) = \sqrt{\theta}(\tau(\theta)-a)^2$, $f(x|\theta) = \theta e^{-\theta x}$, and $\xi(\theta)$ is chosen to be a conjugate prior, find the Bayes estimator of $\tau(\theta) = e^{-\theta}$.

10. Let $\{(X_{i,1}, \ldots, X_{i,n}), i = 1, 2\}$ be independent random samples from normal distributions with means $\mu_i$ and variance $\sigma^2$ respectively. Let $\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{i,j}, i = 1, 2$ and $S_i^2 = \sum_{j=1}^{n}(X_{i,j} - \bar{X}_i)^2, i = 1, 2$. Put $Y_i = \frac{\bar{X}_i}{\sqrt{\hat{\sigma}^2}}, i = 1, 2$, where $\hat{\sigma}^2 = (S_1^2 + S_2^2)/(2n-2)$.

  (a) Obtain the joint pdf (probability density function) of $\{Y_1, Y_2\}$ under the assumption ($\mu_i = 0, i = 1, 2$).

  (b) What is the pdf of $Y_1 - Y_2$ under the assumption $\mu_1 = \mu_2$?

11. Let $X_1, \ldots, X_n$ be a random sample from the normal distribution with mean $\mu_1$ and variance $\sigma^2$. Let $Y_1, \ldots, Y_m$ be a random sample from the normal distribution with mean $\mu_2$ and variance $4\sigma^2$. Consider the hypotheses $H_0 : \mu_1 - \mu_2 = 4$ versus $H_1 : \mu_1 - \mu_2 \neq 4$.

    (a) Derive the level-$\alpha$ Likelihood Ratio test for testing $H_0$ versus $H_1$.

    (b) What is the sampling distribution of your testing statistic under $H_0$?

12. Consider the following regression model:

$$
\begin{aligned}
Y_j &= \theta_1 + \theta_2(x_j - \bar{x}) + \epsilon_j, j = 1, \ldots, n \\
\text{where} \quad \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i,
\end{aligned}
$$

Assume that the $\epsilon_j$'s are independently distributed as normal random variables with means 0 and variance $\sigma^2$ and that the $x_i$'s are non-stochastic.

(a) Assuming a non-informative prior for $\{\theta_i, i = 1, 2, \sigma^2)$ as $P(\theta_i, i = 1, 2, \sigma^2) \propto (\sigma^2)^{-1}$, derive the posterior distribution of $\{\theta_i, i = 1, 2\}$.

(b) Derive the posterior distribution of $\theta_2$ and a $100(1-\alpha)$ % HPD ( Highest Posterior Density) interval for $\theta_2$. How is this HPD interval comparing with the $100(1-\alpha)$ % confidence interval for $\theta_2$ ?