Statistics Ph.D. Qualifying Exam: Part II

January 12, 2008

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, X_2, X_3, X_4 be a random sample of size 4 from the normal distribution with mean 0 and variance 1.

   (a) Find the moment generating function of $W = X_1 \times X_2$.

   (b) Find the moment generating function of $V = X_1 \times X_2 - X_3 \times X_4$.

   (c) Using the moment generating function or other method, prove that V in (b) has a double exponential distribution with p.d.f. $g(v) = \frac{1}{2}e^{|v|}, \quad -\infty < v < \infty$.

   Hint: You can use the following facts and other known distributional properties (such as the relation among normal, χ^2, exponential, and gamma distributions) without proof: (1) Let $X \sim N(\mu, \sigma^2)$ with $E(X) = \mu$ and $Var(X) = \sigma^2$. Then the moment generating function of X is $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. (2) Let $Y \sim Gamma(\alpha, \beta)$ with $E(Y) = \alpha\beta$ and $Var(Y) = \alpha\beta^2$. Then the moment generating function of Y is $\left(\frac{1}{1-\beta t}\right)^{\alpha}$.

2. Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a Bernoulli distribution with probability of success θ, where $0 < \theta < 1$.

   (a) Show that $T = \sum_{i=1}^{n} X_i$ is a complete sufficient statistics for θ.

   (b) Find UMVUE of $\theta^r(1-\theta)^s$, where r and s are non-negative integers with $1 \leq r + s < n$.

   (c) Is there unbiased estimator of the odd ratio, $\frac{\theta}{1-\theta}$? Justify your answer.

3. Let $Y_1, Y_2, \cdots, Y_n$ be a random sample of size n satisfy

$$Y_i = \beta x_i + \epsilon_i, \quad i = 1, 2, ..., n,$$

where $x_1, x_2, \cdots, x_n$ are fixed known constants and $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ are iid with $N(0, \sigma^2)$, σ^2 unknown.

We consider three estimators of β: MLE of β,$\hat{\beta}_{MLE}$, $\hat{\beta}_a = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}$, and $\hat{\beta}_b = \frac{\sum_{i=1}^n Y_i/x_i}{n}$.

(a) Find the MLE of β, $\hat{\beta}_{MLE}$.

(b) Prove the all three estimators of β are unbiased.

(c) Compare these three estimators of β in terms of their variances.

4. Let $(X_1, Y_1) \ldots, (X_n, Y_n)$ be a random sample of size n from a bivariate normal distribution with unknown parameters: means μ_1, μ_2, variances σ_1^2, σ_2^2 and correlation coefficient ρ.

   (a) Describe the paired t test procedure for testing $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$.

   (b) Derive the likelihood ratio test for testing $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$.

   (c) Argue that the paired t test in (a) and the LRT in (b) are equivalent.

5. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from independent Exponential distributions with means μ and$\lambda\mu$ respectively.

   (a) Find the MLE of $P(X_1 > Y_1)$ and hence find the MLE of $P(X_{(1)} > Y_{(1)})$, where $X_{(1)} = \min(X_1, \ldots, X_m)$ and $Y_{(1)} = \min(Y_1, \ldots, Y_n)$.

   (b) Find jointly sufficient statistics S for (λ, μ).

   (c) Prove or disprove that S jointly minimally sufficient for (λ, μ).

6. 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 $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$.

(a) Let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix. Prove that under the null hypothesis H_0 : $\boldsymbol{\beta} = \mathbf{0}$,

$$\frac{\mathbf{Y}'\boldsymbol{A}\mathbf{Y}}{\sigma^2} \sim \chi^2_p.$$

if and only if $\boldsymbol{A}$ is an idempotent matrix such that trace$(\boldsymbol{A}) = p$.

(b) Give a comprehensive explanation of how you the result of part a) can be used to develop tests of hypotheses in an analysis of variance.

7. Suppose that $X|n,\theta$ has a binomial distribution with parameter θ. Suppose we put independent prior distributions on n and θ, with n having Poisson(λ) prior and θ having a Beta(α,β) prior, where α and β are known hyperparameters.

   (a) Prove that the posterior density of θ given $X = x$ and n is Beta($x+\alpha, n-x+\beta$).

   (b) Prove that the posterior probability function of $n + X$ given $X = x$ and θ is Poisson$[(1-\theta)\lambda]$.

   (c) Suppose $\alpha = \beta = 1$ and $X = 10$, explain in details how you can obtained 100 samples of n's from the **posterior distribution** of n given $X = 10$.

8. 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.

9. Let $\{X_1, \ldots, X_n\}$ be a random sample from the normal distribution with mean μ_1 and variance σ_x^2 and $\{Y_1, \ldots, Y_m\}$ a random sample from the normal distribution with mean μ_2 and variance σ_y^2. Assume that the X_i's are independently distributed of the Y_j's. Put $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$, $S_y^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2$, $\hat{\sigma}^2 = \frac{1}{n}S_x^2 + \frac{1}{m}S_y^2$.

   (a) What are a and b if one approximates the sampling distribution of $\hat{\sigma}^2$ by $a\chi_b^2$, where χ_b^2 is a central Chi-square random variable with degrees of freedom b.

   (b) Derive a $(1 - \alpha)$ % approximate confidence interval for $\mu_1 - \mu_2$ by using the approximation in (a).

10. Let $\{X_1, \ldots, X_n\}$ be a random sample from the population with density $f(x; \theta_i, i = 1, 2)$, where $f(x; \theta_i, i = 1, 2)$ is given by:

$$
\begin{aligned}
f(x; \theta_i, i = 1, 2) &= \frac{1}{\theta_2 - \theta_1} \text{ if } \theta_1 \leq x \leq \theta_2, \\
&= 0 \text{ if for otherwise.}
\end{aligned}
$$

(a) Show that the statistics $\{X_{(1)} = Min(X_1, \ldots, X_n), X_{(n)} = Max(X_1, \ldots, X_n)\}$ are sufficient and complete statistics for the parameters $\{\theta_i, i = 1, 2\}$.

(b) Derive the UMVUE of $\theta_2 - \theta_1$.

11. Let $\{X_1, \ldots, X_n\}$ be a random sample from the population with density $f(x; \theta, \mu_i, i = 1, 2) = \theta f_1(x; \mu_1) + (1-\theta) f_2(x; \mu_2)$, where $f_i(x; \mu_i)$ is the density of the normal distribution with mean μ_i and variance 1, and $0 < \theta < 1$. Illustrate how to derive a procedure to compute the MLE (Maximum Likelihood Estimator) of $\{\theta, \mu_i, i = 1, 2\}$ by using the EM-algorithm.

12. Let $\{X_1, \ldots, X_m\}$ be a random sample from the normal population with mean μ_1 and variance σ_1^2. Let $\{Y_1, \ldots, Y_n\}$ be a random sample from the normal population with mean μ_2 and variance σ_2^2 independently of $\{X_1, \ldots, X_m\}$.

(a) Derive the size α Likelihood Ratio test for testing $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.

(b) Derive the power function of your test.

(c) Derive a $1 - \alpha$ % 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 this compared with the procedure of (a)?