Statistics Ph.D. Qualifying Exam: Part I

November 7, 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 random sample from an exponential distribution with parameter 1, and $N \sim \text{Geometric}(\alpha)$. Let $X_{(1)} < \ldots < X_{(N)}$ be the order statistics.

   (a) Find $\text{Var}(\sum_{i=1}^{N} X_i)$.

   (b) Find $P(X_{(N)} > a)$.

   (c) Find $E(X_{(1)})$.

2. In order to decide the appropriate amount to charge as premium, insurance companies often use the exponential principle defined as follows: If X is the random amount that it will have to pay in claims, then the premium charged by the insurance company should be

$$P = \frac{1}{a} \ln(E[e^{aX}]),$$

where $a > 0$ is a fixed specified constant. Suppose that an insurance company assumes that X has a uniform distribution distribution on $[0, \theta]$.

(a) Find P.

(b) An insurance company wishes to find a maximum likelihood estimator $\hat{P}$ of P, by taking a random sample $X_1, \ldots, X_n$ from a large set of previous payments. Assuming that the X's are a random sample from Uniform$[0, \theta]$, where θ is unknown. Find $\hat{P}$.

3. Let $Y_1, \ldots, Y_n$ be a random sample from Poisson(θ). Let $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ be the sample mean and $S^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)$ be the sample variance.

   (a) Prove that $\bar{Y}$ is a complete and sufficient statistic for θ.

   (b) Prove that both $\bar{Y}$ and S^2 are unbiased estimators of θ.

   (c) Prove that $E(S^2|\bar{Y}) = \bar{Y}$.

   (d) Prove that $\text{Var}(S^2) > \text{Var}(\bar{Y})$.

   (e) Is S^2 an admissible estimator of θ?

4. 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 $a\chi_b^2$ by equating the first two cumulants (i.e, the expected value and the variance), where χ_b^2 is a central Chi-square random variable with degrees of freedom b. (Notice that b may not be an integer.)

   (b) Derive a $100(1 - \alpha)\%$ approximate confidence interval for $\mu_1 - \mu_2$ by using the approximation in (a).

5. Let $\{X_1, \ldots, X_n\}$ be a random sample from the probability distribution with density $f(x, \theta, \phi) = \frac{1}{\phi} e^{-\frac{1}{\phi}(x-\theta)}, x \geq \theta,$ where $\phi > 0$.

   (a) Derive the MLE (Maximum Likelihood Estimator) of θ and the MLE of ϕ. Are these MLEs forming a set of sufficient and complete statistics for (θ, ϕ).

   (b) Derive the UMVUE (Uniformly Mininmun Varainced Unbiased Estimator) of θ. What is the UMVUE of ϕ?

6. Let $\{X_1, \ldots, X_n\}$ be a random sample from the Bernoulli distribution with density $f(x, \theta) = \theta^x (1-\theta)^{1-x}, x = 0, 1$, where $0 < \theta < \frac{1}{2}$.

   (a) Derive the maximum likelihood estimator (MLE) $\hat{\theta}$ of θ.

   (b) Show that the MSE (Mean Squared Error) of $\hat{\theta}$ is uniformly $\leq$ the MSE of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

7. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independently and identically distributed as (X, Y), where (X, Y) follows a bivariate normal distribution with means $E(X) = \mu_1$ and $E(Y) = \mu_2$ and with variances and covariance as $Var(X) = \sigma_1^2$, $Var(Y) = \sigma_2^2$ and $Cov(X, Y) = \rho\sigma_1\sigma_2$, respectively. Put $Z_i = X_i - Y_i, i = 1, \ldots, n$.

   (a) Based on the observed Z values, derive the likelihood ratio test (LRT) 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. What is the p-value of the LRT test?

8. Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution with pdf

$$f(x;\theta) = \theta e^{-\theta x}, x > 0; \theta > 0.$$

Find UMVUEs for

(a) $1/\theta$

(b) θ

(c) $e^{-k\theta}$, where k is a known constant.

9. Let $X_1, X_2, \cdots, X_n$ be a random sample from a uniform distribution with pdf

$$f(x; \alpha, \beta) = \frac{1}{2\beta}, \quad \alpha - \beta < x < \alpha + \beta.$$

Find estimators of α and β based on

(a) method of moments

(b) maximum likelihood principle

(c) UMVUE.

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

11. Let Y be a random variable.

(a) Show that the quantity $E(Y - c)^2$ is minimized when $c = E(Y)$. Thus, the "best" predictor of a random variable's value in the sense of minimizing mean-squared prediction error is the mean of the random variable.

(b) Assume that Y is a continuous random variable with pdf $f(y)$. Show that the quantity $E|Y - c|$ is minimized when $c = m$, median(Y), where $\int_{-\infty}^{m} f(y)dy = 1/2$. Thus, the "best" predictor of a random variable's value in the sense of minimizing mean-absolute prediction error is the median of the random variable.

12. Suppose that the primary endpoint of a clinical trial is survival time T and the distribution of T is exponential istribution with p.d.f. $f(t;\theta) = \frac{1}{\theta}e^{-\frac{t}{\theta}}, t > 0$. Suppose further that the study primary objective is formulated into testing $H_0 : \theta = 3$ vs $H_1 : \theta = 1$. Describe how to find the sample size for the clinical trial to obtain a level of 5% test with power 80%, by using

   (a) the exact method, and

   (b) the normal approximation.