Statistics and Biostatistics Ph.D. Theory Qualifier 2005

Answer any four problems. If you answer more than four questions, only the first four will be graded.

Total time: 3 hours. This is a closed book closed notes test.

1. Consider the family of distribution on $(0, \infty)$ with pdf

$$f(x, \theta) \propto \exp(-\theta x^a),$$

$\theta > 0$, where $a > 1$ is fixed.

a. Find the normalizing constant for the density.

b. Does the pdf belong to the regular exponential family? Justify your answer.

c. Find the MLE of θ based on an independent sample of size n from this distribution.

d. Derive the asymptotic distribution of the MLE.

2. a. Consider a Bayes decision problem with scalar parameter θ. An estimate is required for $\phi = \phi(\theta)$, with loss function

$$L(d, \phi) = (d - \phi)^2.$$

Find the form of the Bayes estimator ϕ.

b. Let $X_1, X_2, ...X_n$ be $i.i.d.$ exponential θ distribution ($f(x, \theta) = \theta \exp(-\theta x), x > 0$), where θ is an unknown parameter. Let Z denote some hypothetical future observation from the same distribution, and suppose we wish to estimate $\phi(\theta) = P(Z > z)$, for some given z.

Suppose we assume a $Gamma(\alpha, \beta)$ prior ($\tau(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}, \theta > 0$) for θ. Find the posterior distribution of θ,and show that the Bayes estimator of ϕ is

$$\widehat{\phi} = \left(\frac{\beta + S_n}{\beta + S_n + z}\right)^{\alpha+n}$$

where $S_n = \sum_{i=1}^n X_i$.

3. a. Define the term "unbiased test."

b. Describe briefly the notion of UMP unbiased test.

c. A town supervisor suspects that the traffic condition have become more hazardous in Newtown than in neighboring Oldtown, so she records the number of accidents occurring in each place in the course of one month. Assume that the random variables X and Y denote the number of accidents occurring in Newtown and Oldtown in one month period, and that X and Y are independent Poisson distributions with parameters λ and μ respectively; construct an UMP unbiased test of size α for the hypothesis:

$$H_0 : \lambda \leq \mu \qquad \text{vs.} \qquad H_a : \lambda > \mu$$

(Hint: Compute the conditional distribution of X given $X + Y$)

d. Carry out the test when $X = 5$ and $Y = 2$, and $\alpha = 0.05$.

4. a. Let $Y_1, ..., Y_n$ be $i.i.d.$ with joint pdf $f(y, \theta)$, with θ a scalar parameter. Suppose a one-to-one differentiable transformation is made from θ to $\phi(\theta)$. What is the relationship between the Fisher information $I(\phi)$ about ϕ contained in the sample and $I(\theta)$, the Fisher information about θ?

b. In Bayesian inference about θ,it was suggested by Jeffreys that a noninformative uniform prior should be assumed for θ that is proportional to $\sqrt{I(\theta)}$. Show that this implies that the prior for ϕ is also proportional to $\sqrt{I(\phi)}$. This is known as the invariance property of Jeffreys' prior under one-to-one transformation.

c. Suppose X be a single observation from $Binomial(k, \theta)$. What is the Jeffreys' prior for θ?

b. Find the Bayes estimator of θ with respect to the Jeffreys' prior under squared error loss.

5. Let $X_1, ..., X_n$ be $i.i.d.$negative binomial (r, p); assume r is known, and we are interested in computing a confidence interval for p.

a. Let $Y = \sum_{i=1}^{n} X_i$, argue that Y is a complete and sufficient statistics.

b. If n is large, use normal approximation to construct a $(1 - \alpha)$ confidence interval for p.

c. Another approximation that does not rely on large n, but rather on small p, is as follows:

First show that as $p \longrightarrow 0$, $2pY \longrightarrow \chi^2_{2nr}$ in distribution, and hence $2pY$ is an approximate pivot. Use this result to construct a $(1 - \alpha)$ confidence interval for p.

(Hint: For this part, consider using MGF)

d. Is the confidence interval in part (c) based on the (approximate) pivot narrowest? If not, show how to choose the endpoints in order to obtain a minimum length $(1 - \alpha)$ confidence interval for p.

6. Let $X_i = (\theta/2)t_i^2 + \varepsilon_i, i = 1, 2, ...n$, where ε_i are independent normal random variables with mean 0 and known variance σ^2.

a. Let $T(X) = T(X_1, ..., X_n)$ be any unbiased estimator of θ. Derive the Cramer-Rao lower bound for $T(X)$.

b. Show that $\widetilde{\theta} = \frac{2\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} t_i^2}$ is an unbiased estimator of θ. Nonetheless, show that the variance of $\widetilde{\theta}$ does not attain Cramer-Rao lower bound.

c. Is there an unbiased estimator $\widehat{\theta}$ of θ that attains Cramer-Rao lower bound? If yes, derive the estimator and prove that it attains the Cramer-Rao lower bound; if not, justify your answer.

d. Find a pivot based on $\widehat{\theta}$ (if you have answer yes in part c) or $\widetilde{\theta}$ (if you have answered no in part c), and find a fixed length level $(1 - \alpha)$ confidence interval for θ.

e. If $0 \leq t_i \leq 1, i = 1, 2, .., n$, but we may otherwise choose t_i freely, what values should we use for t_i so as to make the interval in (d) as short as possible?