THE AUSTRALIAN NATIONAL UNIVERSITY
DEPARTMENT OF STATISTICS AND ECONOMETRICS

STATISTICAL INFERENCE - STAT3013

Final Examination 2000

Total Marks: 100

Reading Period: 15 Minutes
Time Allowed: Three Hours
Permitted Materials: Course Brick, Lecture Notes, Non-Programmable Calculator

---

Question 1

Indicate whether each statement is either TRUE or FALSE (please write the entire word, as hasty F's can sometimes look like T's). Two marks will be awarded for each correct response, two marks will be deducted for each incorrect response, and zero marks will be given for no response.

(a) A Bayesian highest posterior density region will be the same as the shortest Bayesian credibility interval whenever the posterior distribution is symmetric.

(b) If T is any estimator of θ, then the Cramér-Rao lower bound implies that $Var_\theta(T) \geq \frac{1}{I(\theta)}$.

(c) For a given dataset from a distribution belonging to a parametric probability model indexed by θ, if T is an unbiased estimator of $\tau(\theta)$ which is approximately normally distributed and $\hat{\theta}$ is the *MLE* of θ, then an approximate 95% confidence interval of the form $T \pm 1.96\sqrt{Var_{\hat{\theta}}(T)}$ cannot be shorter in length than the approximate 95% confidence interval based on the asymptotic normal approximation to the distribution of the *MLE* $\hat{\tau} = \tau(\hat{\theta})$.

(d) If we restrict attention to hypothesis tests determined by critical regions of the form $C = \{t(X_1, \ldots, X_20) \leq k\}$ for a given statistic $T = t(X_1, \ldots, X_{20})$ and some cut-off value k, then it is generally only possible to increase our power of detecting H_1 to be true by simultaneously increasing our Type I error rate.

(e) One of the benefits of the bootstrap percentile method confidence interval is that it is range-respecting; that is, it will never contain values which are outside the allowable parameter space.

[10 marks]

Question 2

Let $X_1, \ldots, X_n$ be an *iid* sample from a distribution with density function of the form:

$$f_X(x;\theta) = \theta(x+1)^{-(\theta+1)}, \qquad x > 0,$$

for some parameter value $\theta > 2$.

(a) Show that these densities form an exponential family and find a complete and sufficient statistic for θ. **[8 marks]**

(b) It can be shown that $E(X_i) = \frac{1}{\theta-1}$ and $Var(X_i) = \frac{\theta}{(\theta-1)^2(\theta-2)}$. Clearly, then, $\overline{X} = n^{-1}\sum_{i=1}^n X_i$ is an unbiased estimator of $\tau(\theta) = \frac{1}{\theta-1}$. Find the Cramér-Rao lower bound for the variance of unbiased estimators of $\tau(\theta)$. Does the variance of $\overline{X}$ achieve this bound? Is $\overline{X}$ an *UMVU* estimator? If not, discuss how you might find an *UMVU* estimator. **[10 marks]**

(c) Find an approximate 95% confidence interval for $\tau(\theta)$ using the normal approximation to the distribution of the *MLE*, $\hat{\tau} = \tau(\hat{\theta})$. **[8 marks]**

(d) It can further be shown that $Q = 2\theta\sum_{i=1}^n \ln(X_i+1)$ has a chi-squared distribution with $2n$ degrees of freedom. Use this fact to construct a 95% pivotal confidence interval in terms of appropriate chi-squared quantiles. **[8 marks]**

(e) If $0 < \theta \leq 2$, then $f_X(x;\theta)$ is still a proper density function, however, its variance does not exist. Moreover, if $\theta \leq 1$, then the mean does not exist either. As such, we may want to expand our allowable parameter space and test whether $\theta = 1$ or $\theta = 2$. Show that the Neymann-Pearson most powerful test of $H_0 : \theta = 1$ versus $H_1 : \theta = 2$ is determined by a rejection region of the form $C = \left\{\sum_{i=1}^n \ln(X_i+1) \leq k\right\}$. Further, determine the value of k which gives this test size α. [HINT: Recall the fact given in part (*d*) regarding the quantity Q.] **[6 marks]**

Question 3

Suppose that we observe 10 values from a distribution with density:

$$f_X(x;\theta) = \frac{\sqrt{\theta}}{x\Gamma(1/2)}\exp\{-\theta(\ln x)^2\}.$$

Furthermore, suppose that we have a prior distribution for θ given by $\pi(\theta) = \theta e^{-\theta}$. Finally, suppose that we have been given the following summary values regarding our observed data:

$$\sum_{i=1}^{10} X_i = 18.6061; \qquad \sum_{i=1}^{10} X_i^2 = 85.9106; \qquad \sum_{i=1}^{10} \ln(X_i) = 0.3068; \qquad \sum_{i-1}^{10} \{\ln(X_i)\}^2 = 10.9469.$$

(a) Show that the posterior distribution of θ given the data has the form of a Gamma distribution with shape parameter 7 and scale parameter 0.0837. **[6 marks]**

(b) Determine the posterior Bayes estimator of θ based on the observed data and the given prior distribution. **[6 marks]**

(c) The posterior Bayes estimator from part (*b*) is, of course, equivalent to the Bayes estimate with respect to squared-error loss. Discuss how you would go about finding the Bayes estimator with respect to some other loss function, $\ell(t;\theta)$. **[9 marks]**

(d) It can be shown that $Var_\theta\{\ln(X_i)\} = 0.5\theta^{-1}$. Discuss how you would construct a highest posterior density Bayesian credibility region for the parameter $\tau(\theta) = 0.5\theta^{-1}$. **[9 marks]**

Question 4

Suppose that we observe 5 ordered pairs:

$$(X_1, Y_1) = (9.73, 22.69), \quad (X_2, Y_2) = (9.18, 21.13), \quad (X_3, Y_3) = (10.86, 21.89),$$
$$(X_4, Y_4) = (10.21, 18.15), \quad (X_5, Y_5) = (10.67, 20.28),$$

and we fit a linear regression model

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

and arrive at the usual least-squares estimates $\hat{\alpha} = 24.928$, $\hat{\beta} = -0.405$ and observed residuals $e_1 = 1.700$, $e_2 = -0.083$, $e_3 = 1.358$, $e_4 = -2.646$, $e_5 = -0.329$. Moreover, we note that

$$\sum_{i=1}^{5}(X_i - \overline{X})^2 = 1.893, \qquad \sum_{i=1}^{5} e_i^2 = 11.848.$$

Now, when we fit the linear regression model without including the first data point, we observe an estimate of the slope of $\hat{\beta}_{(1)} = 0.0971$. Similarly, when we remove each of the other datapoints in turn, we observe slope estimates of $\hat{\beta}_{(2)} = -0.533$, $\hat{\beta}_{(3)} = -1.414$, $\hat{\beta}_{(4)} = -0.265$, and $\hat{\beta}_{(5)} = -0.259$.

(a) Calculate the Jackknife estimate of bias and variance of $\hat{\beta}$ and comment on your results [in particular, recall that $\hat{\beta}$ is known to be unbiased and to have a variance of $Var(\epsilon)/\sum_{i=1}^{5}(X_i - \overline{X})^2$, provided the ϵ_i's are mean-zero, independent and homoscedastic]. **[7 marks]**

(b) Discuss two methods of constructing confidence intervals based on bootstrap re-samples of the 5 ordered pairs. Discuss the general properties of each. **[7 marks]**

(c) What is the probability that a bootstrap re-sample will consist of five identical pairs? What kind of problems would such a re-sample cause for the bootstrap procedure? **[6 marks]**

END OF EXAMINATION