THE AUSTRALIAN NATIONAL UNIVERSITY
DEPARTMENT OF STATISTICS AND ECONOMETRICS

STATISTICAL INFERENCE - STAT3013

Final Examination 2000 - Solutions

Total Marks: 100

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

---

Question 1

(a) FALSE, the posterior may be bimodal.

(b) FALSE, the Cramér-Rao lower bound applies only to unbiased estimators.

(c) TRUE, since T is unbiased, the Cramér-Rao lower bound implies that $Var_{theta}(T) \geq \frac{1}{I(\theta)}$ for any θ. Thus, when $\theta = \hat{\theta}$, we see that $Var_{\hat{\theta}}(T) \geq \frac{1}{I(\hat{\theta})}$ and the right-hand expression is now the normal approximation variance for $\hat{\theta}$. Thus, since the length of the two confidence intervals in question is determined by the variance used, we see that the first interval must be at least as longer than the second.

(d) TRUE, if all we are allowed to change is the cut-off level, then clearly increasing the power means increasing the cut-off (so that more outcomes fall in the rejection region) which must simultaneously increase the chance of a Type I error.

(e) TRUE.

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) We have:

$$\begin{aligned} f_X(x;\theta) &= \theta(x+1)^{-(\theta+1)} \\ &= \exp\{\ln(\theta) - (\theta+1)\ln(x+1)\} \\ &= \exp\{-\theta\ln(x+1) + \ln(\theta) - \ln(x+1)\}, \end{aligned}$$

which clearly has the form of a one-dimensional exponential family with defining functions $a(x) = \ln(x+1)$, $b(\theta) = -\ln(\theta)$, $c_1(\theta) = -\theta$, and $d_1(x) = \ln(x+1)$. Therefore, we know that a complete, sufficient statistic for θ is given by $D = \sum_{i=1}^n d_1(X_i) = \sum_{i=1}^n \ln(X_i+1)$.

(b) To find the Cramér-Rao bound, we first need the Fisher information. From part (a), we see that the log-likelihood for θ based on the sample is $l(\theta) = -\theta\sum_{i=1}^n \ln(X_i+1) + n\ln(\theta) - \sum_{i=1}^n \ln(X_i+1) = -\theta D + n\ln(\theta) - D$, which leads to a second derivative of:

$$l''(\theta) = -\frac{n}{\theta^2}.$$

Thus, the Fisher information is:

$$I(\theta) = -E_\theta\{l''(\theta)\} = \frac{n}{\theta^2}.$$

Moreover, we have $\tau'(\theta) = -(\theta-1)^{-2}$. Finally, then, the Cramér-Rao lower bound for the variance of an unbiased estimator T of $\tau(\theta)$ is:

$$Var_\theta(T) \geq \frac{\{\tau'(\theta)\}^2}{I(\theta)} = \frac{(\theta-1)^{-4}}{n/\theta^2} = \frac{\theta^2}{n(\theta-1)^4}.$$

Now, the variance of $\overline{X}$ is just

$$Var_\theta(\overline{X}) = \frac{1}{n}Var_\theta(X_i) = \frac{\theta}{n(\theta-1)^2(\theta-2)}.$$

A simple bit of algebra shows that

$$\frac{\theta}{n(\theta-1)^2(\theta-2)} - \frac{\theta^2}{n(\theta-1)^4} = \frac{\theta(\theta-1)^2 - \theta^2(\theta-2)}{n(\theta-1)^4(\theta-2)} = \frac{\theta}{n(\theta-1)^4(\theta-2)},$$

which is clearly positive for any $\theta > 2$. Therefore, the variance of $\overline{X}$ does not achieve the Cramér-Rao lower bound. This does not, in general, imply directly that $\overline{X}$ is not *UMVU*; however, in this case we can see that $\overline{X}$ is not *UMVU* since it is not a function of D, the complete and sufficient statistic in this problem. In order to determine an estimator which is *UMVU*, we could find $T = E(\overline{X}|D)$.

(c) The derivative of the log-likelihood given in part (b) is $l'(\theta) = n\theta^{-1} - D$, which clearly implies that $\hat{\theta}_{MLE} = nD^{-1}$, and this means that $\hat{\tau}_{MLE} = \tau(\hat{\theta}_{MLE}) = (nD^{-1} - 1)^{-1} = \frac{D}{n-D}$. Furthermore, the approximate variance is given by

$$\frac{\{\tau'(\hat{\theta}_{MLE})\}^2}{I(\hat{\theta})} = \frac{\hat{\theta}^2}{n(\hat{\theta}-1)^4} = \frac{nD^2}{(n-D)^4}.$$

Therefore, an approximate 95% confidence interval can be constructed as:

$$\frac{D}{n-D} \pm 1.96\frac{\sqrt{n}D}{(n-D)^2}.$$

Alternatively, we might use $-\Phi^{-1}(1-\beta)$ and $\Phi^{-1}(0.95+\beta)$ for some $0 \leq \beta \leq 0.05$ instead of $-\Phi(0.975) = -1.96$ and $\Phi^{-1}(0.975) = 1.96$ to arrive at an asymmetric interval.

(d) Using the given distributional fact, we know that:

$$Pr\{\chi^2_{(2n)}(0.05-\beta) < Q \leq \chi^2_{(2n)}(1-\beta)\} = (1-\beta) - (0.05-\beta) = 0.95,$$

for any $0 \leq \beta \leq 0.05$. After some straightforward algebraic manipulation (and noting that $Q = 2\theta D$, this probability statement yields a pivotal confidence interval of the form:

$$C = \left\{ \frac{\chi^2_{(2n)}(0.05-\beta)}{2D}, \frac{\chi^2_{(2n)}(1-\beta)}{2D} \right\}.$$

(e) The Neymann-Pearson test of $H_0 : \theta = 1$ versus $H_1 : \theta = 2$ is determined by a rejection region of the form $C = \{\Lambda(X_1, \ldots, X_n) \leq c_\alpha\}$, where c_α is a cut-off value chosen to give the test size α and

$$\begin{aligned}
\Lambda(X_1, \ldots, X_n) &= \frac{L(1)}{L(2)} \\
&= \frac{\prod_{i=1}^n (X_i+1)^{-(1+1)}}{\prod_{i=1}^n 2(X_i+1)^{-(2+1)}} \\
&= \prod_{i=1}^n \frac{1}{2}(X_i+1).
\end{aligned}$$

Thus, the desired test has a rejection region of the form:

$$\begin{aligned}
C &= \left\{ \prod_{i=1}^n \frac{1}{2}(X_i+1) \leq c_\alpha \right\} \\
&= \left\{ \prod_{i=1}^n (X_i+1) \leq c_\alpha 2^n \right\} \\
&= \left\{ \sum_{i=1}^n \ln(X_i+1) \leq \ln(c_\alpha 2^n) \right\},
\end{aligned}$$

which is precisely the required form when $k = \ln(c_\alpha 2^n)$. Finally, to set the size, we note that we need:

$$\alpha = Pr_{H_0}(C) = Pr_{\theta=1}(D \leq k) = Pr_{\theta=1}(2D \leq 2k).$$

Now, since we know that $2\theta D \sim \chi^2_{(2n)}$, we see that $2D$ has a $\chi^2_{(2n)}$ distribution when $\theta = 1$, and thus we need $2k$ to be equal to the α-quantile of the $\chi^2_{(2n)}$ distribution. In other words, we need

$$2k = \{\chi^2_{(2n)}\}^{-1}(\alpha) \qquad \Longrightarrow \qquad k = \frac{1}{2}\{\chi^2_{(2n)}\}^{-1}(\alpha),$$

where $\{\chi^2_{(2n)}\}^{-1}(\cdot)$ is the inverse of the *CDF* of a chi-squared distribution with $2n$ degrees of freedom.

Question 3

(a) The form of the posterior distribution can be found as:

$$\begin{aligned}\pi(\theta|X_1,\ldots,X_{10}) &\propto \pi(\theta)L(\theta;X_1,\ldots,X_{10})\\ &= \theta e^{-\theta}\prod_{i=1}^{10} f_X(X_i;\theta)\\ &= \theta e^{-\theta}\frac{\theta^5}{\{\Gamma(1/2)\}^{10}}\prod_{i=1}^{10}\frac{e^{-\theta(\ln X_i)^2}}{X_i}\\ &= \frac{\theta^6}{\{\Gamma(1/2)\}^{10}\prod_{i=1}^{10}X_i}\exp\left\{-\theta-\theta\sum_{i=1}^{10}(\ln X_i)^2\right\}\\ &\propto \theta^6\exp\left[-\theta\left\{1+\sum_{i=1}^{10}(\ln X_i)^2\right\}\right],\end{aligned}$$

which clearly has the form of a gamma distribution with shape parameter $6+1=7$ and scale parameter $\left\{1+\sum_{i=1}^{10}(\ln X_i)^2\right\}^{-1} = (11.9469)^{-1} = 0.0837$.

(b) The posterior Bayes estimator is simply the mean of the posterior distribution, which means that $\hat{\theta}_\pi = 7(0.0837) = 0.5859$ since the mean of a gamma distribution is just the product of its shape and scale parameters.

(c) To find a more general Bayes estimate with respect to the loss function $\ell(t;\theta)$, we must choose the estimate, $t = t(X_1,\ldots,X_n)$, which will minimise the Bayes risk $r(t) = E_\pi\{R_t(\theta)\}$, where $R_t(\theta) = E_\theta\{\ell(t;\theta)\}$.

(d) To find a highest posterior density for $\tau(\theta)$, we reparameterise the posterior distribution for θ, $\pi(\theta|X_1,\ldots,X_{10})$, to the posterior distribution for τ using the standard change of variable formula (since τ is a monotonic function of θ):

$$\pi_\tau(\tau|X_1,\ldots,X_{10}) = \frac{1}{|\tau'(\theta)|}\pi\{\theta^{-1}(\tau)|X_1,\ldots,X_{10}\} = \frac{1}{2\tau}\pi(0.5\tau^{-1}|X_1,\ldots,X_{10}),$$

and then finding the highest posterior density region, C_τ, as that set of τ values which satisfied the criteria $\int_{C_\tau}\pi_\tau(t|X_1,\ldots,X_{10})dt = 1-\alpha$ and $(\tau\in C_\tau)$ & $\{\pi_\tau(\tau'|X_1,\ldots,X_{10}) > \pi_\tau(\tau|X_1,\ldots,X_{10})\} \implies \tau'\in C_\tau$. In other words, C_τ is that set of τ values having posterior probability $1-\alpha$ which have the highest posterior density values; thus, the highest posterior density region can be determined graphically by sliding a horizontal line down across a graph of the posterior density until the region corresponding to the τ values having posterior density value larger than the cut-off line is $1-\alpha$. [NOTE: Alternatively, since the relationship between τ and θ is monotonic in this case, we could find a highest posterior density region for θ, say C_θ and then transform this region into one for τ by simply appropriately transforming each element so that $C_\tau = \{\tau : 0.5\tau^{-1}\in C_\theta\}$.]

Question 4

Suppose that we observe 5 ordered pairs:

$$(X_1,Y_1) = (9.73, 22.69), \qquad (X_2,Y_2) = (9.18, 21.13), \qquad (X_3,Y_3) = (10.86, 21.89),$$

$$(X_4, Y_4) = (10.21, 18.15), \qquad (X_5, Y_5) = (10.67, 20.28),$$

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) The Jackknife estimate of bias is:

$$\hat{B}_J = (5-1)\left(\frac{1}{5}\sum_{i=1}^{5}\hat{\beta}_{(i)} - \hat{\beta}\right) = 4(-0.47478 + 0.405) = -0.27912.$$

Similarly, the Jackknife estimate of variance is:

$$\frac{5-1}{5}\sum_{i=1}^{5}(\hat{\beta}_{(i)} - \hat{\beta}_\bullet)^2 = \frac{4}{5}(1.30314) = 1.0425.$$

Now, we know that the true bias of $\hat{\beta}$ is 0 and the true variance is given by $Var(\epsilon)/1.893$, which can be approximated here as $MSE/1.893 = (11.848/3)/1.893 = 2.0863$ (since the *MSE* is just the sum of the squared errors divided by the appropriate degrees of freedom, $n - 2 = 3$ in this case). As such, we see that the Jackknife estimates are not that good here.

(b) One method is based on using the bootstrap to estimate the mean and variance of $\hat{\beta}$ and then employing these values in the normal approximation to the distribution of $\hat{\beta}$. The advantage of this method is that it generally requires fewer re-samples to accurately estimate the mean and variance then to estimate distributional quantiles (which are required for the percentile method, which is the other bootstrap confidence interval procedure). However, this method is neither range-respecting (not an issue here, since slopes are not range-restricted parameters) nor parameterisation equivariant. The second method is the bootstrap percentile method, which constructs an interval based on the quantiles of the estimated quantiles of the distribution of $\hat{\beta}$ as approximated by a large number of re-sampled estimates $\hat{\beta}_b^\star$. This interval is range-respecting and parameterisation equivariant. However, it is more computationally expensive as noted above.

(c) Clearly, any given re-sample has a probability of 5^{-5} of occuring. As there are 5 possible re-samples with all pairs identical, we see that the desired probability is $5^{-4} = 0.0016$. In this case, such a re-sample is quite problematic for the bootstrap, since a dataset with 5 identical pairs cannot be used to fit a regression. As such, our collection of $\hat{\beta}_b^\star$'s will have about 1 in 1000 values which are undefined. The typical solution to this problem is a pragmatic one; namely, we ignore the undefined values, but the consequences of this on the distribution should be recognised. Fortunately, the chance of undefined slopes is small and thus will not be much of a problem, particularly in larger, more realistic, datasets.

END OF EXAMINATION