Qualifying Exam Statistical Theory Problem Solutions

August 2005

1. Let X_1, X_2, ..., X_n be iid uniform $U(0, \theta)$, $0 < \theta < \infty$.

(a) Assume a quadratic loss function. Calculate the Bayes estimate for the improper prior

$$\pi(\theta) = 1, \quad 0 < \theta < \infty$$

Verify whether the Bayes estimate is consistent for θ.

(b) Identify (without proof) a complete sufficient statistic T. Using completeness and sufficiency of T, calculate $E(\overline{X}|T)$ where $\overline{X} = \sum_{i=1}^n X_i/n$.

Solution: (a) The likelihood is

$$\frac{1}{\theta^n} I_{(X_{(n)}<\theta)} I_{(X_{(1)}>0)}$$

Since $X_{(1)} > 0$ with probability 1 for any θ, we take this likelihood to be

$$\frac{1}{\theta^n} I_{(X_{(n)}<\theta)}$$

The Bayes estimate is

$$\frac{\int_{X_{(n)}}^{\infty} \theta \frac{1}{\theta^n} d\theta}{\int_{X_{(n)}}^{\infty} \frac{1}{\theta^n} d\theta} = \frac{n-1}{n-2} X_{(n)} \xrightarrow{P} \theta, \quad \forall \theta$$

The Bayes estimate is consistent for θ.

(b) $X_{(n)}$ is a complete and sufficient statistic. By completeness and sufficiency, $E(\overline{X}|X_{(n)}) = \Psi(X_{(n)})$, where Ψ is the unique function with $E_\theta(\Psi(X_{(n)})) = E_\theta(\overline{X}) = \theta/2$. Now, $E_\theta(X_{(n)}) = \theta\frac{n}{n+1}$. Therefore $E(\overline{X}|X_{(n)}) = \frac{n+1}{2n} X_{(n)}$.

2. Let $X_1, ..., X_n$ be a sample from the beta distribution $\beta(\theta, 1)$.

(a) Find the MLE of $1/\theta$. Is it unbiased? Calculate the information inequality lower bound and check whether the MLE achieves the lower bound?

(b) Find an unbiased estimate of $\theta/(\theta + 1)$. Does the unbiased estimate achieve the information inequality lower bound?

Solution: (a) The density function of the beta distribution $\beta(\theta, 1)$ is $f(x; \theta) = \theta x^{\theta-1}$, $\theta > 0$, $0 < x < 1$. The likelihood from the n samples is

$$\begin{aligned} L &= \exp\{\theta \sum_{i=1}^n \log X_i + n \log \theta - \sum_{i=1}^n \log X_i\} \quad \text{and} \\ l &= \log L = \theta \sum_{i=1}^n \log X_i + n \log \theta - \sum_{i=1}^n \log X_i \end{aligned}$$

The MLE of θ is the root of the equation $\sum_{i=1}^n \log X_i + n/\theta = 0$. Therefore MLE $\widehat{\theta} = -n/\sum_{i=1}^n \log X_i$ and the MLE of $1/\theta$ is $-\sum_{i=1}^n \log X_i/n$. Note

$$\begin{aligned} E(-\frac{\sum_{i=1}^n \log X_i}{n}) &= -E(\log X_1) \\ &= -\int_0^1 (\log x)\theta x^{\theta-1}dx \\ &= -\int_0^1 (\log x)d(x^\theta) \\ &= -(\log x)x^\theta|_0^1 + \int_0^1 x^\theta \frac{1}{x}dx \\ &= \int_0^1 x^{\theta-1}dx = \frac{1}{\theta} \end{aligned}$$

hence this MLE is unbiased. Moreover,

$$\begin{aligned} var(\frac{-\sum_{i=1}^n \log X_i}{n}) &= \frac{var(\log X_1)}{n} \\ &= \frac{1}{n}(E(\log X_1)^2 - \frac{1}{\theta^2}) \\ &= \frac{1}{n}(\frac{2}{\theta^2} - \frac{1}{\theta^2}) = \frac{1}{n\theta^2} \end{aligned}$$

and the Fisher information number is

$$I(\theta) = -E(\frac{\partial^2}{\partial\theta^2}\log f(x,\theta)) = \frac{1}{\theta^2}$$

The information inequality lower bound is

$$\frac{\left((\frac{1}{\theta})'\right)^2}{nI(\theta)} = \frac{\frac{1}{\theta^4}}{n\frac{1}{\theta^2}} = \frac{1}{n\theta^2} = var(\frac{-\sum_{i=1}^n \log X_i}{n})$$

Therefore the MLE achieves the lower bound.

(b) Note that

$$E(X) = \int_0^1 x\theta x^{\theta-1}dx = \int_0^1 \theta x^\theta dx = \frac{\theta}{\theta+1}x^{\theta+1}|_0^1 = \frac{\theta}{\theta+1}$$

therefore $\overline{X}$ is the unbiased estimate of $\theta/(\theta+1)$. Its variance is

$$\frac{var(X)}{n} = \frac{1}{n}\left(E(X^2) - \left(\frac{\theta}{\theta+1}\right)^2\right) = \frac{1}{n}\left(\frac{\theta}{\theta+2} - \frac{\theta^2}{(\theta+1)^2}\right) = \frac{\theta}{n(\theta+2)(\theta+1)^2}$$

The lower bound is

$$\frac{\left((\frac{\theta}{\theta+1})'\right)^2}{nI(\theta)} = \frac{\left(\frac{1}{(\theta+1)^2}\right)^2}{n\frac{1}{\theta^2}} = \frac{\theta^2}{n(\theta+1)^4}$$

It is easy to show that

$$\frac{\theta}{(\theta+1)^2} < \frac{1}{\theta+2}$$

and therefore the lower bound

$$\frac{\theta^2}{n(\theta+1)^4} < \frac{\theta}{n(\theta+2)(\theta+1)^2} = var(\overline{X})$$

Remark: For an estimate to attain the lower bound, it is necessary and sufficient that there is a linear relation between the estimate and the derivative of the log likelihood. This shows the first estimate attains the lower bound but the second does not. However (a) requires the calculation of the lower bound.

3. Let $\mathbf{X} = (X_1, ..., X_n)$ be an iid sample from an exponential density with mean θ. Consider testing $H_0 : \theta = \theta_0$ vs. $H_1 : \theta > \theta_0$. Let $P(\mathbf{X})$ = your p-value for an appropriate test.

(a) What is $E_{\theta_0}(P(\mathbf{X}))$? Derive your answer explicitly.

(b) Derive $E_\theta(P(\mathbf{X}))$ for $\theta \neq \theta_0$. Specifically, assuming only one sample, i.e. $n = 1$, calculate $E_\theta(P(\mathbf{X}))$ as explicitly as possible for $\theta \neq \theta_0$.

(c) When there is only one sample, is $E_\theta(P(\mathbf{X}))$ a decreasing function of θ? In general, can you prove your result for an arbitrary MLR family?

Solution: (a) Let T be the test statistic. For the exponential density, $T = \sum_{i=1}^n X_i$. Let the distribution function of T be $F_\theta(t)$, which is actually $gamma(n, \theta)$ distribution. The p-value equals $1 - F_{\theta_0}(T)$. Since $F_{\theta_0}(T)$ is a standard uniform random variable under the null hypothesis (this can be easily proved), $1 - F_{\theta_0}(T)$ is also uniform in $(0, 1)$. Therefore $E_{\theta_0}(P(\mathbf{X})) = E(1 - F_{\theta_0}(T)) = 1/2$.

(b) Let T_0 denote a random variable with distribution F_{θ_0} and let T_0 be independent of T. Then

$$\begin{aligned} E_\theta(P(\mathbf{X})) &= \int P(T_0 \geq t | T = t) f_\theta(t) dt = P(T_0 > T) \\ &= \int_0^\infty \int_t^\infty \frac{s^{n-1} e^{-s/\theta_0}}{\theta_0^n \Gamma(n)} ds \frac{t^{n-1} e^{-t/\theta}}{\theta^n \Gamma(n)} dt \end{aligned}$$

where the density of $gamma(n, \theta)$ is used. When there is only one sample, the above integral is simplified:

$$\begin{aligned} E_\theta(P(\mathbf{X})) &= \int_0^\infty \int_t^\infty \frac{e^{-s/\theta_0}}{\theta_0} ds \frac{e^{-t/\theta}}{\theta} dt = \int_0^\infty \frac{1}{\theta} e^{-t(\frac{1}{\theta}+\frac{1}{\theta_0})} dt \\ &= \frac{\theta_0}{\theta_0 + \theta} \end{aligned}$$

(c) When there is only one sample, $E_\theta(P(\mathbf{X})) = \frac{\theta_0}{\theta_0+\theta}$ is a decreasing function of θ. Suppose we define MLR as meaning $f_\theta(t)/f_{\theta'}(t)$ is nondecreasing function of t, for $\theta > \theta'$. Then the expectation of p-value is nondecreasing. To prove this note that p-value $= 1 - F(T)$, where $F()$ is the distribution function of T under null. This is a decreasing function of T hence its expectation is a descreasing function of θ. (This is a property of MLR families. It is given as a problem in Cassella Berger. The proof is nontrivial. We give this problem just to see if any one can do it. This part should carry less than half of the points for part (c)).

4. Let X_1, X_2 be iid uniform on $\theta - \frac{1}{2}$ to $\theta + \frac{1}{2}$.

(a) Show that for any given $0 < \alpha < 1$, you can find $c > 0$ such that

$$P_\theta\{\overline{X} - c < \theta < \overline{X} + c\} = 1 - \alpha$$

(b) Show that for ϵ positive and sufficiently small

$$P_\theta\{\overline{X} - c < \theta < \overline{X} + c \,|\, |X_2 - X_1| \geq 1 - \epsilon\} = 1$$

(c) The statement in (a) is used to assert that $\overline{X} \pm c$ is a $100(1-\alpha)\%$ confidence interval for θ. Does the assertion in (b) contradict this? If your sample observations are $X_1 = 1$, $X_2 = 2$, would you use the confidence interval in (a)?

Solution: (a) Note that

$$P_\theta\{\overline{X} - C < \theta < \overline{X} + C\} = P_\theta\{-C < \overline{X} - \theta < C\} = P_{\theta=0}\{-C < \overline{X} < C\} \qquad (*)$$

which is a continuous increasing function of C and varies between zero (when $C = 0$) and one (when $C = \infty$). Hence one can find $C > 0$ s.t. the above probability $= 1 - \alpha$, for any $0 < \alpha < 1$.

(b) Note that for the order statistics $X_{(1)} = \min(X_1, X_2)$ and $X_{(2)} = \max(X_1, X_2)$, $\theta - \frac{1}{2} \leq X_{(1)} < X_{(2)} \leq \theta + \frac{1}{2}$, i.e.

$$X_{(2)} - \frac{1}{2} \leq \theta \leq X_{(1)} + \frac{1}{2} \qquad (**)$$

If $X_{(2)} - X_{(1)} > 1 - \epsilon$, then (**) implies

$$\begin{aligned} \frac{1}{2} + \overline{X} - X_{(2)} &\geq \overline{X} - \theta \geq \overline{X} - X_{(1)} - \frac{1}{2} \\ \text{i.e. } \frac{1}{2} - \left(\frac{X_{(2)} - X_{(1)}}{2}\right) &\geq \overline{X} - \theta \geq \left(\frac{X_{(2)} - X_{(1)}}{2}\right) - \frac{1}{2} \\ \text{i.e. } \epsilon &\geq \overline{X} - \theta \geq -\epsilon \end{aligned}$$

Hence for sufficient small ϵ, $0 < \epsilon < C$ the above event will imply $|\overline{X} - \theta| < C$, i.e., $\theta \in (\overline{X} - C, \overline{X} + C)$. Hence the conditional probability =1.

(c) Say, $X_{(2)} - X_{(1)} = 1$. Since $0 \leq X_{(2)} - X_{(1)} \leq (\theta + \frac{1}{2}) - (\theta - \frac{1}{2}) = 1$ we now must have $X_{(2)} = \theta + \frac{1}{2}$, $X_{(1)} = \theta - \frac{1}{2}$. So we know for sure $\theta = \overline{X}$. To say that we have confidence $1 - \alpha < 1$ that θ lies in our interval is counterintuitive.

5. The inverse Gaussian distributions have the following density function,

$$f(y; \theta, \sigma) = (2\pi\sigma)^{-1/2} y^{-3/2} \exp\{-(2\sigma y)^{-1}(y\theta^{-1} - 1)^2\}, \quad y > 0, \theta > 0, \sigma > 0$$

with mean θ and variance $\theta^3\sigma$. Let $Y_{11}, Y_{12}, \ldots, Y_{1n}$ be an iid sample drawn from $f(y:\theta_1,\sigma)$ and $Y_{21}, Y_{22}, \ldots, Y_{2n}$ an iid sample drawn from $f(y:\theta_2,\sigma)$. Assume that $\theta_i^{-1} = \mu+\alpha_i$, for $i = 1,2$; and $\alpha_1+\alpha_2 = 0$.
(a). Write down the likelihood function and derive the MLEs for μ, α_1, α_2 and σ.
(b). Derive the likelihood ratio test statistic (LRTS) for $H_0 : \alpha_1 = \alpha_2 = 0$.
Solution: (a) The log-likelihood function is

$$\ell = c - n\log\sigma - \frac{1}{2\sigma}\sum_i\sum_j Y_{ij}^{-1}(Y_{ij}(\mu+\alpha_i)-1)^2$$

Taking the direvatives of l with respect to μ and α_i, we have

$$\frac{\partial\ell}{\partial\mu} = -\frac{1}{\sigma}\sum_i\sum_j(Y_{ij}(\mu+\alpha_i)-1);$$
$$\frac{\partial\ell}{\partial\alpha_i} = -\frac{1}{\sigma}\sum_j(Y_{ij}(\mu+\alpha_i)-1) \text{ for } i=1,2.$$

Let $Y_{..} = \sum_i\sum_j Y_{ij}$ and $Y_{i.} = \sum_j Y_{ij}$. Setting $\frac{\partial l}{\partial\mu}$ and $\frac{\partial l}{\partial\alpha_i}$ to be zeros and considering that $\alpha_1+\alpha_2=0$, we have

$$\hat{\mu} = \frac{n}{2}\left(\frac{1}{Y_{1.}}+\frac{1}{Y_{2.}}\right);$$
$$\hat{\alpha}_1 = \frac{n}{2}\left(\frac{1}{Y_{1.}}-\frac{1}{y_{2.}}\right);$$
$$\hat{\alpha}_2 = \frac{n}{2}\left(-\frac{1}{Y_{1.}}+\frac{1}{Y_{2.}}\right).$$

Taking the derivative of l with respect to σ and setting it to be zero, we have

$$\frac{\partial\ell}{\partial\sigma} = -\frac{n}{\sigma}+\frac{1}{2\sigma^2}\sum_i\sum_j Y_{ij}^{-1}(Y_{ij}(\mu+\alpha_i)-1)^2 = 0.$$

Then

$$\hat{\sigma} = \frac{1}{2n}\sum_i\sum_j Y_{ij}^{-1}(Y_{ij}(\hat{\mu}+\hat{\alpha}_i)-1)^2$$
$$= \frac{1}{2n}\sum_i\sum_j\frac{1}{Y_{ij}} - \frac{n}{2Y_{1.}} - \frac{n}{2Y_{2.}}$$

(b) Under $H_0 : \alpha_1 = \alpha_2 = 0$, the log-likelihood function is

$$\ell_0 = c - n\log\sigma - \frac{1}{2\sigma}\sum_i\sum_j Y_{ij}^{-1}(Y_{ij}\mu-1)^2,$$

and the MLEs for μ and σ are

$$\tilde{\mu} = \frac{2n}{Y_{..}};$$

$$\tilde{\sigma} = \frac{1}{2n}\sum_i\sum_j Y_{ij}^{-1}(Y_{ij}\tilde{\mu}-1)^2 = \frac{1}{2n}\sum_i\sum_j \frac{1}{Y_{ij}} - \frac{2n}{Y_{..}}.$$

Hence the logarithm of the likelihood ratio is

$$\lambda = \ell(\hat{\mu},\hat{\alpha}_1,\hat{\alpha}_2,\hat{\sigma}) - \ell_0(\tilde{\mu},\tilde{\sigma}) = -n\log\frac{\tilde{\sigma}}{\hat{\sigma}}$$

$$= -n\log\frac{\frac{1}{2n}\sum_i\sum_j \frac{1}{Y_{ij}} - \frac{2n}{Y_{..}}}{\frac{1}{2n}\sum_i\sum_j \frac{1}{Y_{ij}} - \frac{n}{2Y_{1.}} - \frac{n}{2Y_{2.}}}.$$

6. Let $X^1, X^2, \ldots, X^n$ denote a sample of n independent observations generated from the following model

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \lambda \\ \lambda \end{pmatrix} F + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

where μ_1, μ_2 and λ are unknown parameters and F, ε_1 and ε_2 are random variables independent and identically distributed as $N(0,1)$. Let $\bar{X}$ be the sample mean and

$$S = \frac{1}{n}\sum(X^i-\bar{X})(X^i-\bar{X})' = \begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix}$$

be the sample covariance matrix.

(a). Write down the log-likelihood function for μ_1, μ_2 and λ.

(b). Show that $S_{11}+S_{22}+2S_{12}$ and $\bar{X}$ are sufficient statistics for λ, μ_1 and μ_2.

Solution: (a) It is easy to see that X follows a bivariate normal distribution with mean $\mu = \binom{\mu_1}{\mu_2}$ and variance matrix

$$\Sigma = \begin{pmatrix} \lambda \\ \lambda \end{pmatrix}(\lambda,\lambda) + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+\lambda^2 & \lambda^2 \\ \lambda^2 & 1+\lambda^2 \end{pmatrix}.$$

So the log-likelihood function is

$$\ell = c - \frac{n}{2}\log|\Sigma| - \sum_{i=1}^{n}\frac{1}{2}(X^i-\mu)'\Sigma^{-1}(X^i-\mu).$$

(b) The inverse of Σ is

$$\Sigma^{-1} = \frac{1}{1+2\lambda^2}\begin{pmatrix} 1+\lambda^2 & -\lambda^2 \\ -\lambda^2 & 1+\lambda^2 \end{pmatrix}.$$

ℓ can be simplied as follows:

$$\begin{aligned}
\ell &= c - \frac{n}{2}\log|\Sigma| - \frac{n}{2}tr(\Sigma^{-1}S) - \frac{n}{2}(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu) \\
&= c - \frac{n}{2}\log(1+2\lambda^2) - \frac{n}{2}\frac{(1+\lambda^2)(S_{11}+S_{22})-2\lambda^2 S_{12}}{1+2\lambda^2} - \frac{n}{2}(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu) \\
&= c - \frac{n}{2}\log(1+2\lambda^2) - n(S_{11}+S_{22}-S_{12}) - \frac{n}{4}\frac{S_{11}+S_{22}+2S_{12}}{1+2\lambda^2} - \frac{n}{2}(\bar{X}-\mu)'\Sigma^{-1}(\bar{X}-\mu).
\end{aligned}$$

By the factoriazation theorem, the sufficient statistics for μ and λ are $\bar{X}$ and $S_{11}+S_{22}+2S_{12}$.

7. Let $\beta = (\widehat{\beta}_1, ..., \widehat{\beta}_p)^T$ be the set of LSE under the general model

$$E(Y) = \mathbf{X}\beta = \sum_{i=1}^{p} \beta_i \mathbf{x}_i \qquad \Sigma_Y = \sigma^2 I$$

where Y and $\mathbf{x}_i$ are d-dimensional vectors and $\mathbf{x}_i$ is the ith column of $\mathbf{X}$. Let H denote the hypothesis that

$$E(Y) = \sum_{i=1}^{m} \beta_i \mathbf{x}_i \quad \text{or} \qquad \beta_{m+1} = ... = \beta_p = 0$$

Show that $(\widehat{\beta}_1, ..., \widehat{\beta}_m)$ is a set of LSE for $(\beta_1, ..., \beta_m)$ under the hypothesis H if and only if

$$\mathbf{x}_i \perp \sum_{j=m+1}^{p} \widehat{\beta}_j \mathbf{x}_j \quad i = 1, ..., m$$

<u>**Solution**</u>: Let V denote the space spanned by all columns of X, and $V_1 \subset V$ be the space spanned by the first m columns of X. Moreover, let P be the orthogonal projection from to R^n to V, and P_1 be the orthogonal projection from V to V_1, and lastly, let P_2 be the orthogonal projection from R^n to V_1. Clearly, $P_2 = P_1 P$, so if we denote $(\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_m^*)$ be the LSE of $(\beta_1, \beta_2, \ldots, \beta_m)$ under H, then

$$P_1 \left[\sum_{j=1}^{m} \hat{\beta}_j \mathbf{x}_j + \sum_{j=m+1}^{p} \hat{\beta}_j \mathbf{x}_j \right] = \sum_{j=1}^{m} \hat{\beta}_j^* \mathbf{x}_j;$$

since $P_1(\sum_{j=1}^m \hat{\beta}_j \mathbf{x}_j) = (\sum_{j=1}^m \hat{\beta}_j \mathbf{x}_j)$, we conclude:

$$P_1 \left(\sum_{j=m+1}^{p} \hat{\beta}_j \mathbf{x}_j \right) = \sum_{j=1}^{m} (\hat{\beta}_j^* - \beta_j) \mathbf{x}_j.$$

Now $(\hat{\beta}_1, \ldots, \hat{\beta}_m)$ is LSE is equivalent to that $(\hat{\beta}_1, \ldots, \hat{\beta}_m) = (\hat{\beta}_1^*, \ldots, \hat{\beta}_m^*)$ and is equivalent to

$$P_1 \left[\sum_{j=m+1}^{p} \hat{\beta}_j \mathbf{x}_j \right] = 0,$$

which is then equivalent to $\mathbf{x}_i \perp \sum_{j=m+1}^p \hat{\beta}_j \mathbf{x}_j$ for $i = 1, 2, \ldots, m$; this finishes the proof.

8. For a multiple linear regression $Y = X\beta + \epsilon$, where X is n by p full rank matrix $(p < n)$, let the hat matrix $H = X(X'X)^{-1}X'$. Moreover, write $X = [W, V]$, and let H^* be the new hat matrix of the linear model $Y = W\beta + \epsilon$.

a). Show that $H_{ij}^2 \leq H_{ii} \cdot H_{jj}$ for any $1 \leq i, j \leq n$.

b). Show that $H - H^*$ is symmetric and semi-definite.

c). If the first column of X are all 1's, show that $H_{ii} \geq 1/n$ for all $1 \leq i \leq n$.

<u>Solution</u>: (a). Letting x_i be the i-th row of X, notice that $H_{ij} = x_i(X'X)^{-1}x_j$ which defines an inner product $< x_i, x_j >$; (a) follows directly.

(b). To abuse the symbol a little, use X, V, and W for the spaces spanned by the column vectors of X, V and W correspondingly. It is sufficient to show that for any non-zero vector $\xi \in V$, $(I - H^*)\xi \neq 0$. Split $\xi = \eta + \tau$ where $\eta \perp W$ and $\tau \in W$, clearly $(I - H^*)\xi = \eta \neq 0$; the last inequality follows from the full rank assumption of X.

(c). Clearly H and H^* commute with each other so $H - H^*$ itself is symmetric and idempotent and thus semi-definite. (d). Let H^* be the hat matrix with W be the first column of X, then $H_{ii} \geq H^*_{ii} = 1/n$.