Second Term Exam STA 3000Y1Y
Friday April 15, 2011

Instructions: Answer all 4 questions in the exam booklets. Be as precise as possible in your answers, stating clearly the theorems and assumptions you are using. Questions are of equal value.

1. Given a model $f(y;\theta)$ for a variable $Y \in R$ and parameter $\theta \in R$, consider a one-to-one reparameterization $\varphi = \varphi(\theta)$. Denote the log-likelihood function for θ by $\ell(\theta)$, and that for φ by $\ell^*(\varphi)$.[1]

   (a) Use the chain rule to show that

   $$i^*(\varphi) = i(\theta)\left(\frac{\partial\theta}{\partial\varphi}\right)^2, \qquad (1)$$

   where $i(\theta)$ and $i^*(\varphi)$ are the expected Fisher information based on $\ell(\theta)$ and $\ell^*(\varphi)$, respectively. Show that a similar result holds for observed Fisher information when it is evaluated at the maximum likelihood estimator, but not otherwise.

   $$\begin{aligned} \ell^*(\varphi) &= l(\theta(\varphi)) \\ \ell^{*'}(\varphi) &= \ell'(\theta)\theta'(\varphi) \\ \ell^{*''}(\varphi) &= \ell''(\theta)\theta''(\varphi) + \ell'(\theta)\theta''(\varphi) \\ E\{-\ell^{*''}(\varphi)\} &= i(\theta)\{\theta'(\varphi)\}^2, \quad \text{because } E\{\ell'(\theta) = 0, \\ -\ell^{*''}(\hat\varphi) &= -\ell''(\hat\theta)\{\theta'(\hat\varphi)\}^2, \quad \text{because } \ell'(\hat\theta) = 0, \end{aligned}$$

   and $\theta(\hat\varphi) = \hat\theta$ by invariance of the maximum likelihood estimator.

   (b) When $\theta \in R^p$, we write the vector version of (1) as

   $$i^*(\varphi) = \left(\frac{\partial\theta}{\partial\varphi}\right)^T i^{-1}(\theta)\left(\frac{\partial\theta}{\partial\varphi}\right).$$

[1] Even though they are essentially the same, i.e. $\ell^*(\varphi;y) = \ell(\theta(\varphi);y))$, where $\theta(\cdot)$ is the inverse function.

Consider the three asymptotic chi-squared pivots for inference on θ:

$$\begin{aligned} w(\theta) &= 2\{\ell(\hat\theta) - \ell(\theta)\}, \\ w_u(\theta) &= \left(\frac{\partial \ell(\theta)}{\partial \theta}\right)^T i(\theta) \frac{\partial \ell(\theta)}{\partial \theta}, \\ w_e(\theta) &= (\hat\theta - \theta)^T i(\theta)(\hat\theta - \theta), \end{aligned}$$

where $\hat\theta$ is the maximum likelihood estimate, which we assume is obtained as the unique solution of the score equation $\partial\ell(\theta)/\partial\theta = 0$.

Show that $w(\theta)$ and $w_u(\theta)$ are invariant under one-to-one reparametrizations of θ, but that $w_e(\theta)$ is not.

$$w = 2\{\ell(\hat\theta) - \ell(\theta)\} = 2\{\ell^*(\varphi(\hat\theta) - \ell^*(\varphi(\theta))\} = 2\{\ell^*(\hat\varphi) - \ell^*(\varphi)\}$$

$$\begin{aligned} w_u^*(\varphi) &= \left(\frac{\partial \ell(\varphi)}{\partial \varphi}\right)^T i^*(\varphi) \frac{\partial \ell(\varphi)}{\partial \varphi} \\ &= \left(\frac{\partial \theta}{\partial \varphi}\frac{\partial \ell}{\partial \theta}\right)^T \left(\frac{\partial \theta}{\partial \varphi}\right)^{-1} i^{-1}(\theta) \left(\frac{\partial \theta}{\partial \varphi}\right)^{-T} \frac{\partial \theta}{\partial \varphi}\frac{\partial \ell}{\partial \theta} \\ &= \left(\frac{\partial \ell(\theta)}{\partial \theta}\right)^T i(\theta) \frac{\partial \ell(\theta)}{\partial \theta} \end{aligned}$$

$$w_e^*(\varphi) = (\hat\varphi - \varphi)^T i^*(\varphi)(\hat\varphi - \varphi) = (\varphi(\hat\theta) - \varphi(\theta)) \left(\frac{\partial \theta}{\partial \varphi}\right)^T i(\theta) \left(\frac{\partial \theta}{\partial \varphi}\right) (\varphi(\hat\theta) - \varphi),$$

which is clearly not invariant, as there is a change of scale.

2. Suppose Y follows a Poisson distribution with probability mass function

$$f(y;\theta) = \theta^y e^{-\theta}/y!, \quad y = 0, 1, 2, \ldots; \theta > 0.$$

Assume a Gamma prior for θ:

$$\pi(\theta) = \frac{1}{\Gamma(\beta)}\alpha^\beta \theta^{\beta-1} e^{-\alpha\theta};$$

this prior distribution has mean $E(\theta) = \beta/\alpha$ and variance $\text{var}(\theta) = \beta/\alpha^2$.

(a) Find the Bayes estimator of θ under squared error loss.

(Most people assumed a sample of size n, but this was not required.) The Bayes estimator under squared-error loss is the mean of the posterior:

$$\pi(\theta \mid y) \propto \theta^{y+\beta-1} e^{-(\alpha+1)\theta},$$

which is the kernel of a Gamma distribution with shape $y+\beta$ and scale $\alpha+1$. The mean of this distribution is

$$\tilde{\theta}_B(y) = E(\theta \mid y) = \frac{y+\beta}{\alpha+1}.$$

(b) Find the Bayes risk of the estimator in (a), as a function of α and β.

The Bayes risk is $r_B = \int R(\theta, \tilde{\theta}_B(y))\pi(\theta)d\theta$, and $R(\theta, \tilde{\theta}_B) = E_{Y|\theta}\{(\tilde{\theta}_B - \theta)^2\} = \text{var}(\tilde{\theta}_B(Y)) + E^2(\tilde{\theta}_B(Y))$. Thus

$$\begin{aligned} r_B &= \int \text{var}_{Poisson}(\frac{y+\beta}{\alpha+1}) + (E_{Poisson}(\frac{y+\beta}{\alpha+1} - \theta)^2\pi(\theta)d\theta \\ &= \int \{\frac{\theta}{(\alpha+1)^2} + (\frac{\theta+\beta}{\alpha+1} - \theta)^2\}\pi(\theta)d\theta \\ &= \frac{E_{Gamma}(\theta)}{(\alpha+1)^2} + \frac{E_{Gamma}\theta^2(-\alpha^2-2\alpha)}{(\alpha+1)^2} + \frac{E_{Gamma}(\theta)(2\beta)}{(\alpha+1)^2} + \frac{\beta^2}{(\alpha+1)^2}, \end{aligned}$$

, which doesn't simplify into anything very nice (sorry).

(c) How could you use this result to find the minimax estimator of θ, if it exists?

A Bayes estimator which has constant frequentist risk is minimax: i.e. can we choose α, β so that $R(\theta, \tilde{\theta}_B) = \text{c}$. Since this expression is a quadratic in θ, we need the coefficient of θ^2 and θ to be zero. The coefficient of θ^2 is $1 - 2\alpha$, but the coefficient of θ is $-2\beta\alpha$, so only $\beta = 0$ would work, which gives an improper prior.

3. Consider a linear regression model

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, \ldots, n$$

where the ϵ_i are independent and follow a $N(0, \sigma^2)$ distribution. Thus the joint density is

$$f(\underline{y}; \alpha, \beta, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\{-\frac{1}{2\sigma^2}\sum(y_i - \alpha - \beta x_i)^2\}.$$

We will consider inference for σ^2.

(a) Give an expression for the profile log-likelihood ratio statistic

$$w_{\mathrm{p}}(\sigma^2) = 2\{\ell_{\mathrm{p}}(\hat{\sigma}^2) - \ell_{\mathrm{p}}(\sigma^2)\}$$

where $\ell_{\mathrm{p}}(\sigma^2) = \ell(\hat{\alpha}_{\sigma^2}, \hat{\beta}_{\sigma^2}, \sigma^2)$ is the profile log-likelihood function.

(The hint should have appeared further up in the question.) It is easily verified that the maximum likelihood equations for α and β do not depend on σ^2, thus

$$\hat{\alpha}_{\sigma^2} = \hat{\alpha} = \bar{y}, \quad \hat{\beta}_{\sigma^2} = \hat{\beta} = S_{xy}/S_{xx}.$$

The maximum likelihood estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2.$$

Thus

$$\begin{aligned}
\ell_{\mathrm{p}}(\sigma^2) &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}n\hat{\sigma}^2 \\
\ell_{\mathrm{p}}(\hat{\sigma}^2) &= -\frac{n}{2}\log\hat{\sigma}^2 - \frac{n}{2} \\
w_{\mathrm{p}}(\sigma^2) &= n\log\sigma^2 + n\hat{\sigma}^2/\sigma^2 - n\log\hat{\sigma}^2 - n \\
&= n\left\{\log\left(\frac{\sigma^2}{\hat{\sigma}^2}\right) - \left(1 - \frac{\hat{\sigma}^2}{\sigma^2}\right)\right\}
\end{aligned}$$

(b) Show that σ^2 is orthogonal to α and β with respect to expected Fisher information.

It probably follows from the result above $\hat{\alpha}_{\sigma^2} = \hat{\alpha}, \hat{\beta}_{\sigma^2} = \hat{\beta}$, but is easily proved directly:

$$\begin{aligned}
\frac{\partial\ell(\alpha,\beta,\sigma^2)}{\partial\alpha} &= \frac{1}{\sigma^2}\sum(y_i - \alpha - \beta x_i) \\
\frac{\partial^2\ell(\alpha,\beta,\sigma^2)}{\partial\alpha^2} &= -\frac{1}{\sigma^4}\sum(y_i - \alpha - \beta x_i) \\
\frac{\partial\ell(\alpha,\beta,\sigma^2)}{\partial\beta} &= \frac{1}{\sigma^2}\sum(y_i - \alpha - \beta x_i)x_i \\
\frac{\partial^2\ell(\alpha,\beta,\sigma^2)}{\partial\beta^2} &= -\frac{1}{\sigma^4}\sum(y_i - \alpha - \beta x_i)x_i
\end{aligned}$$

and the second derivatives have mean 0 since $E(y_i) = \alpha + \beta x_i$.

(c) Give an expression for the adjusted profile log-likelihood ratio statistic

$$w_{\mathrm{a}}(\sigma^2) = 2\{\ell_{\mathrm{a}}(\hat{\sigma}^2) - \ell_{\mathrm{a}}(\sigma^2)\}$$

where

$$\ell_{\mathrm{a}}(\sigma^2) = \ell_{\mathrm{p}}(\sigma^2) - \frac{1}{2}\log|j_{\lambda\lambda}(\hat{\alpha}_{\sigma^2}, \hat{\beta}_{\sigma^2}, \sigma^2)|,$$

where $j_{\lambda\lambda}(\theta)$ is the submatrix of the observed Fisher information matrix for the nuisance parameter $\lambda = (\alpha, \beta)$.

First we calculate $\ell_{\alpha\alpha} = -n\alpha/\sigma^2$, $\ell_{\alpha\beta} = -\sum x_i/\sigma^2 = 0$, $\ell_{\beta\beta} = -\sum x_i^2/\sigma^2$, so

$$|j_{\lambda\lambda}(\hat{\alpha}, \hat{\beta}, \sigma^2)| \propto \sigma^{-4},$$

and thus

$$\begin{aligned} \ell_{\mathrm{a}}(\sigma^2) &= -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}n\hat{\sigma}^2 - \frac{1}{2}\log(\sigma^{-4}) \\ &= -\frac{n-2}{2}\log\sigma^2 - \frac{n\hat{\sigma}^2}{2\sigma^2} \\ w_a(\sigma^2) &= (n-2)\log\left(\frac{\sigma^2}{\hat{\sigma}^2}\right) - n\left(1 - \frac{\hat{\sigma}^2}{\sigma^2}\right) \end{aligned}$$

(d) Compare the result in (b) to the exact marginal likelihood for σ^2 obtained from the distribution of the residual sum of squares

$$\sum(y_i - \hat{\alpha} - \hat{\beta}x_i)^2.$$

Hint: Simplify some calculations by assuming that $\sum x_i = 0$.

The density of a χ^2_ν distribution is

$$\frac{1}{\Gamma(\nu/2)}\frac{1}{2^{\nu/2}}x_{\nu/2-1}e^{-x/2}.$$

This should have been included on the exam. Thus, since

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2},$$

it has density proportional to

$$(\sigma^{-2})^{(n-2)/2}\exp\{-\frac{n\hat{\sigma}^2/2\sigma^2}{\}},$$

and the log-likelihood for this distribution is

$$\ell_{\mathrm{m}}(\sigma^2) = -\left(\frac{n-2}{2}\right)\log\sigma^2 - \frac{n\hat{\sigma}^2}{\sigma^2},$$

which is equal to $\ell_{\mathrm{a}}(\sigma^2)$ above. The marginal likelihood for σ^2 is sometimes called "REML", for restricted maximum likelihood, and can be extended to normal theory linear models with fixed and random effects.

4. Suppose $Y_1, \ldots, Y_n$ are independent and identically distributed from a model $f(y;\theta), y \in R, \theta \in R$, and that $\pi(\theta)$ is a proper prior density (with respect to Lebesgue measure on R). Denote by $\hat{\theta}_\pi$ the posterior mode:

$$\hat{\theta}_\pi = \arg\sup_\theta \pi(\theta \mid \underline{y})$$

which we assume is obtained as the unique root of the equation

$$\frac{d}{d\theta}\log\pi(\hat{\theta}_\pi \mid \underline{y}) = 0. \tag{2}$$

This question was downweighted because it involved too much calculation. This was the question I meant to ask on HW3, and I should have asked the HW3 question here, because it is much easier.

(a) Write the posterior density in the form

$$\pi(\theta \mid \underline{y}) = \frac{\exp\{\ell(\theta) + \log\pi(\theta)\}}{\int \exp\{\ell(\theta) + \log\pi(\theta)\}d\theta},$$

and expand the integrand in the denominator about $\hat{\theta}_\pi$ to show that the asymptotic posterior distribution of $\hat{\theta}_\pi$ is normal with mean θ. Give an expression for the asymptotic variance.

(I should have said "expand the exponent in the numerator and denominator".)

Let $\ell_\pi(\theta) = \ell(\theta) + \log\pi(\theta)$, and write

$$\begin{aligned}
\ell_\pi(\theta) &= \ell_\pi(\hat{\theta}_\pi) + (\theta - \hat{\theta}_\pi)\ell_\pi'(\theta) + \frac{1}{2}(\theta - \hat{\theta}_\pi)^2\ell_\pi''(\hat{\theta}_\pi) + R_n \\
\exp\{\ell_\pi(\theta)\} &= \exp\{\ell_\pi(\hat{\theta}_\pi)\}\exp\{\frac{1}{2}(\theta - \hat{\theta}_\pi)^2\ell_\pi''(\hat{\theta}_\pi) + R_n\} \\
&= \exp\{\ell_\pi(\hat{\theta}_\pi)\}\sqrt{2\pi}|-\ell_\pi''(\hat{\theta}_\pi)|^{-1/2}\exp\{\frac{1}{2}(\theta - \hat{\theta}_\pi)^2\ell_\pi''(\hat{\theta}_\pi)\}(1 + r_n);
\end{aligned}$$

on inserting this into the numerator and denominator, and assuming $r_n = o_p(1)$, we have

$$\pi(\theta \mid y) \dot{\sim} N(\hat{\theta}_\pi, j^{-1}(\hat{\theta}_\pi)).$$

We will have $r_n \xrightarrow{p} 0$ if $\hat{\theta}_\pi - \theta \xrightarrow{p} 0$ and the 3rd derivative of ℓ_π is bounded in expectation, which follows from the usual assumptions on ℓ and some smoothness constraints on the prior.

(b) Show that

$$\hat{\theta}_\pi - \hat{\theta} = O_p(\frac{1}{n}).$$

Hint: Expand (2) and $\ell'(\hat{\theta})$ in a Taylor series around θ.

From $\ell'(\hat{\theta}) = 0$ and $\ell'_\pi(\hat{\theta}_\pi) = 0$ we have

$$\ell'_\pi(\theta) + (\hat{\theta}_\pi - \theta)\ell''_\pi(\theta) \doteq \ell'(\theta) + (\hat{\theta} - \theta)\ell''(\theta)$$

from which we can write

$$(\hat{\theta}_\pi - \hat{\theta})\ell''(\theta) = \theta g''(\theta) - g'(\theta),$$

where $g(\theta) = \log \pi(\theta)$. Since $\ell''(\theta) = O_p(n)$, we have

$$\hat{\theta}_\pi - \hat{\theta} = O_p(1/n)$$

as long as $\theta g''(\theta) - g'(\theta) = O(1)$, i.e. is bounded.

Actually we should be looking as well at the remainder terms in the two expansions, say

$$R_{1n} = \frac{1}{2}(\hat{\theta}_\pi - \theta)^2 \ell'''_\pi(\theta^*), R_{2n} = \frac{1}{2}(\hat{\theta} - \theta)^2 \ell'''(\theta^{**})$$

where θ^*, θ^{**} are between θ and $\hat{\theta}_\pi$, $\hat{\theta}$, respectively. These remainder terms are $O_p(1)$ under the usual assumptions, since, for example, $(\hat{\theta} - \theta)^2 = O_p(1/n)$ and $(1/n)\ell'''(\theta^{**})$ converges to its expected value.