Statistics 581, Problem Set 8

Wellner; 11/21/2001

Reading: Chapter 3, Section 2;
Ferguson, ACLST, Chapter 19, pages 126-132, Chapter 20, pages 133-134;
Lehmann and Casella, pages 113-129, and 439- 443;
begin reading Chapter 4 (to be handed out on Wednesday 11/21).
Due: Wednesday, November 28, 2001.

1. Suppose that $\theta = (\theta_1, \theta_2 \in \Theta \subset R^k$ where $\theta_1 \in R$ and $\theta_2 \in R^{k-1}$. Show that:
A. $l_1^* = \dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2$ is orthogonal to $[\dot{l}_2] \equiv \{a' \dot{l}_2 : a \in R^{k-1}\}$ in $L_2(P_\theta)$.
B. $I_{11\cdot 2} = \inf_{c \in R^{k-1}} E_\theta (\dot{l}_1 - c' \dot{l}_2)^2$ and that the minimum is achieved when $c' = I_{12} I_{22}^{-1}$.
Thus
$$I_{11\cdot 2} = E_\theta (\dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2)^2 = E_\theta [(l_\theta^*)^2] \, .$$
C. Prove the formula (14) on page 16 of Chapter 3 and interpret this formula geometrically.

Solution: A. Note that for any $a \in R^{k-1}$ we have
$$\begin{aligned} E_\theta [l_1^* \dot{l}_2^T a] &= E_\theta \left\{ (\dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2) \dot{l}_2^T a \right\} \\ &= \left\{ E_\theta \left\{ \dot{l}_1 \dot{l}_2^T \right\} - I_{12} I_{22}^{-1} E_\theta \left\{ \dot{l}_2 \dot{l}_2^T \right\} \right\} a \\ &= \{ I_{12} - I_{12} \} a = 0 \, . \end{aligned}$$
Thus l_1^* is orthogonal to $[\dot{l}_2]$ in $L_2(P_\theta)$.
B. Note that for any $c \in R^{k-1}$ we have
$$\begin{aligned} & E_\theta (\dot{l}_1 - c' \dot{l}_2)^2 \\ &= E_\theta (\dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2 + I_{12} I_{22}^{-1} \dot{l}_2 - c' \dot{l}_2)^2 \\ &= E_\theta (\dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2)^2 + E_\theta ((I_{12} I_{22}^{-1} - c') \dot{l}_2)^2 \\ &= I_{11} - I_{12} I_{22}^{-1} I_{21} + E_\theta ((I_{12} I_{22}^{-1} - c') \dot{l}_2)^2 \\ &\geq I_{11\cdot 2} \end{aligned}$$
with equality if and only if $c' = I_{12} I_{22}^{-1}$. Here the second equality uses the orthogonality proved in A.
C. Formula (14) says that
$$\tilde{l}_1 = I_{11}^{-1} \dot{l}_1 - I_{11}^{-1} I_{12} \tilde{l}_2 \, . \tag{0.1}$$
One way to derive this is as indicated on page 17: since $\tilde{l} = I^{-1} \dot{l}$ we have
$$\tilde{l}_1 = I^{11} \dot{l}_1 + I^{12} \dot{l}_2 \qquad \text{and} \qquad \tilde{l}_2 = I^{21} \dot{l}_1 + I^{22} \dot{l}_2 \, .$$
Hence it follows that
$$\begin{aligned} & \tilde{l}_1 + I_{11}^{-1} I_{12} \tilde{l}_2 \\ &= I^{11} \dot{l}_1 + I^{12} \dot{l}_2 + I_{11}^{-1} I_{12} (I^{21} \dot{l}_1 + I^{22} \dot{l}_2) \\ &= I_{11}^{-1} \left\{ (I_{11} I^{11} + I_{12} I^{21}) \dot{l}_1 + (I_{11} I^{12} + I_{12} I^{22}) \dot{l}_2 \right\} \\ &= I_{11}^{-1} \left\{ Ident \cdot \dot{l}_1 + 0 \cdot \dot{l}_2 \right\} \\ &= I_{11}^{-1} \dot{l}_1 \, . \end{aligned}$$

Rearranging yields (0.1). Note that this indentity decomposes the efficient influence function $\tilde{l}_1$ in the larger model with both θ_1 and θ_2 unknown into its projection onto the efficient influence function in the sub-model when θ_2 is known, namely $I_{11}^{-1}\dot{l}_1$, and a term which is orthogonal to $[\dot{l}_1]$.

2. Suppose that $(Y|Z) \sim \text{Poisson}(\lambda e^{\gamma Z})$, and $Z \sim G_\eta$ on R with density g_η with respect to some dominating measure μ. You may assume that

$$a(z) \equiv (\partial/\partial\eta)\log g_\eta(z)$$

exists and $E\{a^2(Z)\} < \infty$. Thus Z is a "covariate" or "predictor variable", γ is a "regression parameter" which affects the intensity of the (conditionally) Poisson variable Y, and $\theta = (\lambda, \gamma, \eta)$.
(a) Find the information matrix for θ. What does the structure of this matrix say about the effect of η being known or unknown about the estimation of λ and γ?
(b) Find the information and information bound for γ if the parameter λ is known.
(c) Find the efficient score function and the efficient influence function for estimation of γ when λ is known.
(d) Find the information and information bound for γ if the parameter λ is unknown, $I_{\gamma\gamma\cdot\lambda}$.
(e) Find the efficient score function and the efficient influence function for estimation of γ when λ is unknown.
(f) In the case when $Z \sim \text{Bernoulli}(\eta)$, compute the ratio of the information when λ is unknown, to the information when λ is known as a function of γ and of η.

Solution: (a) The density of $X = (Y, Z)$ is

$$p_\theta(y,z) = f_{\lambda,\gamma}(y|z)g_\eta(z) = e^{-\lambda e^{\gamma z}}\frac{(\lambda e^{\gamma z})^y}{y!}g_\eta(z)\,,$$

and hence

$$\log p_\theta(y,z) = y\log(\lambda e^{\gamma z}) - \lambda e^{\gamma z} - \log y! + \log g_\eta(z)\,.$$

From this we calculate the scores $\dot{l}_\lambda$, $\dot{l}_\gamma$, and $\dot{l}_\eta$:

$$\begin{aligned}
\dot{l}_\lambda(y,z) &= \frac{y}{\lambda} - e^{\gamma z} = \frac{1}{\lambda}(y - \lambda e^{\gamma z})\,,\\
\dot{l}_\gamma(y,z) &= yz - \lambda e^{\gamma z}z = z(y - \lambda e^{\gamma z})\,,\\
\dot{l}_\eta(y,z) &= a(z)\,.
\end{aligned}$$

This leads to calculating the entries of the information matrix as follows:

$$\begin{aligned}
I_{\lambda,\lambda} &= E_\theta(\dot{l}_\lambda(X)^2) = \lambda^{-2}E[(Y-\lambda e^{\gamma Z})^2]\\
&= \lambda^{-2}E[E[(Y-\lambda e^{\gamma Z})^2|Z]]\\
&= \lambda^{-2}E[\lambda e^{\gamma Z}] = E[e^{\gamma Z}]/\lambda\,,\\
I_{\gamma,\gamma} &= E_\theta(\dot{l}_\gamma(X)^2) = E[Z^2(Y-\lambda e^{\gamma Z})^2]\\
&= E[E[Z^2(Y-\lambda e^{\gamma Z})^2|Z]]\\
&= E[Z^2\lambda e^{\gamma Z}] = \lambda E[Z^2 e^{\gamma Z}]\,,\\
I_{\eta,\eta} &= E_\theta a^2(Z)\,,
\end{aligned}$$

$$
\begin{aligned}
I_{\lambda,\gamma} &= E_\theta(\dot{l}_\lambda \dot{l}_\gamma(X)) = \lambda^{-1} E_\theta(Z(Y-\lambda e^{\gamma Z})^2) \\
&= \lambda^{-1} E[E[Z(Y-\lambda e^{\gamma Z})^2 | Z]] \\
&= \lambda^{-1} E[Z\lambda e^{\gamma Z}]\,, \\
I_{\lambda,\eta} &= E_\theta(\dot{l}_\lambda(X)\dot{l}_\eta(X)) = \lambda^{-1} E[(Y-\lambda e^{\gamma Z})a(Z)] \\
&= E[E[a(Z)(Y-\lambda e^{\gamma Z})|Z]] \\
&= E[a(Z)\cdot 0] = 0\,, \\
I_{\gamma,\eta} &= E_\theta(\dot{l}_\gamma(X)\dot{l}_\eta(X)) = E[Z(Y-\lambda e^{\gamma Z})a(Z)] \\
&= E[E[a(Z)Z(Y-\lambda e^{\gamma Z})|Z]] \\
&= E[Za(Z)\cdot 0] = 0\,.
\end{aligned}
$$

Thus the information matrix $I(\theta) = I(\lambda, \gamma, \eta)$ is given by:

$$
I(\theta) = \begin{pmatrix} \lambda^{-1}E(e^{\gamma Z}) & E(Ze^{\gamma Z}) & 0 \\ E(Ze^{\gamma Z}) & \lambda E(Z^2 e^{\gamma Z}) & 0 \\ 0 & 0 & E_\theta a^2(Z) \end{pmatrix}.
$$

(b) If λ is known, the information for γ is $I_{\gamma,\gamma} = \lambda E(Z^2 e^{\gamma Z})$, and the information bound is $1/I_{\gamma,\gamma} = 1/\{\lambda E(Z^2 e^{\gamma Z})\}$.
(c) When λ is known, the efficient score function for γ is just the score function $\dot{l}_\gamma$, and the efficient influence function is $\tilde{l}_\gamma = \dot{l}_\gamma / I_{\gamma,\gamma}$.
(d) When λ is unknown, the (efficient) information for γ is

$$
\begin{aligned}
I_{\gamma\gamma\cdot\lambda} &= I_{\gamma,\gamma} - I_{\gamma,\lambda} I_{\lambda,\lambda}^{-1} I_{\lambda,\gamma} \\
&= \lambda E(Z^2 e^{\gamma Z}) - \frac{[E(Ze^{\gamma Z})]^2}{E(e^{\gamma Z})/\lambda} \\
&= \lambda \left\{ E(Z^2 e^{\gamma Z}) - \left(\frac{E(Ze^{\gamma Z})}{E(e^{\gamma Z})} \right)^2 E(e^{\gamma Z}) \right\} \\
&= \lambda E(e^{\gamma Z}) Var_{G_\gamma}(Z)\,,
\end{aligned}
$$

where G_γ is the $\gamma-$tilted distribution corresponding to G given by

$$
G_\gamma(A) = \frac{\int_A e^{\gamma z} dG(z)}{\int e^{\gamma z} dG(z)}\,.
$$

(e) When λ is unknown, the efficient score function for γ is

$$
\begin{aligned}
\dot{l}_\gamma^*(x) &= \dot{l}_\gamma(x) - I_{\gamma\lambda} I_{\lambda\lambda}^{-1} \dot{l}_\lambda(x) \\
&= z(y - \lambda e^{\theta z}) - \frac{\lambda E(Ze^{\gamma Z})}{E(e^{\gamma Z})} \lambda^{-1}(y - \lambda e^{\gamma z}) \\
&= \left(z - \frac{E(Ze^{\gamma Z})}{E(e^{\gamma Z})} \right) (y - \lambda e^{\gamma z})\,.
\end{aligned}
$$

Note that $E[l_\gamma^*(X)^2] = I_{\gamma,\gamma\cdot\lambda}$ which we computed in (d). Thus the efficient influence function for γ is $\tilde{l}_\gamma(x) = l_\gamma^*(x)/I_{\gamma\gamma\cdot\lambda}$.

(f) When $Z \sim \text{Bernoulli}(\eta)$, the ratio of the information when λ is unknown to the information when λ is known is

$$\begin{aligned}\frac{I_{\gamma,\gamma\cdot\lambda}}{I_{\gamma,\gamma}} &= \frac{\lambda Var_{G_\gamma}(Z)}{\lambda E(Z^2 e^{\gamma Z})} \\ &= \frac{\eta e^\gamma - \left(\frac{\eta e^\gamma}{\eta e^\gamma + (1-\eta)}\right)^2}{\eta e^\gamma} \\ &= \frac{1-\eta}{1-\eta+\eta e^\gamma}\,.\end{aligned}$$

See Figure 1 for a plot of this as a function of η for various values of γ.

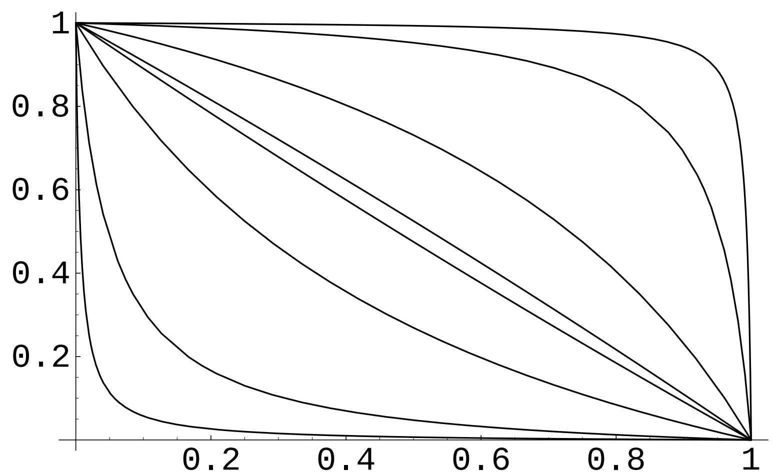

Figure 1: Ratio of Information for γ with λ known and unknown.

3. Information for location-scale families. Example 6.5, TPE page 126.
A. Confirm that Lehmann's information matrix for (regular) location-scale families is correct.
B. Verify that the off-diagonal term $I_{12} = 0$ when the location -scale family is from a density f that is symmetric about 0, and interpret this geometrically in terms of the scores for location $\mu = \theta_1$ and for scale $\sigma = \theta_2$.
C. Produce an example of a location-scale family which is not symmetric about 0 and hence for which $I_{12} \neq 0$. Compute the informatio matrix $I(\theta)$ as explicitly as possible in this case.

Solution: A. Suppose that

$$p_\theta(x) = \frac{1}{\theta_2} f\left(\frac{x-\theta_1}{\theta_2}\right)$$

where $\int [(f'(y))^2/f(y)]dy < \infty$ and $\int y[(f'(y))^2/f(y)]dy < \infty$. Then the scores for θ_1 and θ_2 are given by

$$\begin{aligned}\dot{l}_1(x) &= -\frac{f'((x-\theta_1)/\theta_2)}{f((x-\theta_1)/\theta_2)}\frac{1}{\theta_2}\,, \\ \dot{l}_2(x) &= -\frac{1}{\theta_2} - \frac{f'((x-\theta_1)/\theta_2)}{f((x-\theta_1)/\theta_2)}\frac{(x-\theta_1)}{\theta_2^2}\,.\end{aligned}$$

Hence it follows that

$$\begin{aligned} I_{11} &= \int \dot{l}_1^2(x) p_\theta(x) dx \\ &= \int \left\{ -\frac{f'((x-\theta_1)/\theta_2)}{f((x-\theta_1)/\theta_2)} \frac{1}{\theta_2} \right\}^2 \frac{1}{\theta_2} f\left(\frac{x-\theta_1}{\theta_2} \right) dx \\ &= \frac{1}{\theta_2^2} \int \left\{ -\frac{f'(y)}{f(y)} \right\}^2 f(y) dy \equiv \frac{1}{\theta_2^2} I_{location}(f) \, . \end{aligned}$$

Similarly,

$$\begin{aligned} I_{22} &= \int \dot{l}_2^2(x) p_\theta(x) dx \\ &= \int \left\{ -\frac{1}{\theta_2} - \frac{f'((x-\theta_1)/\theta_2)}{f((x-\theta_1)/\theta_2)} \frac{x-\theta_1}{\theta_2^2} \right\}^2 \frac{1}{\theta_2} f\left(\frac{x-\theta_1}{\theta_2} \right) dx \\ &= \frac{1}{\theta_2^2} \int \left\{ -1 - y \frac{f'(y)}{f(y)} \right\}^2 f(y) dy \equiv \frac{1}{\theta_2^2} I_{scale}(f) \, . \end{aligned}$$

Finally, noting that $\int \dot{l}_1(x) p_\theta(x) dx = 0$, we have

$$\begin{aligned} I_{12} &= \int \dot{l}_1(x) \dot{l}_2(x) p_\theta(x) dx \\ &= \int \left\{ -\frac{f'((x-\theta_1)/\theta_2)}{f((x-\theta_1)/\theta_2)} \frac{1}{\theta_2} \right\} \left\{ -\frac{1}{\theta_2} - \frac{f'((x-\theta_1)/\theta_2)}{f((x-\theta_1)/\theta_2)} \frac{x-\theta_1}{\theta_2^2} \right\} \frac{1}{\theta_2} f\left(\frac{x-\theta_1}{\theta_2} \right) dx \\ &= \frac{1}{\theta_2^2} \int y \left\{ -\frac{f'(y)}{f(y)} \right\}^2 f(y) dy \, . \end{aligned}$$

B. If the density f is symmetric about 0, then f is an even function (so $f(-x) = f(x)$ for all $x \geq 0$), $-f'$ is an odd function (so $f'(-x) = -f'(x)$ for all $x \geq 0$). Hence $-f'/f$ is an odd function and $(-f'/f)^2$ is an even function. Since the integrand in the last expression for I_{12} is the product of an even function, $(-f'/f)^2 f \equiv h$, and an odd function, $g(y) = y$, the integral equals zero in this case:

$$\begin{aligned} \int g(y) h(y) dy &= \int_{-\infty}^0 g(y) h(y) dy + \int_0^\infty g(y) h(y) dy \\ &= \int_0^\infty g(-y) h(-y) dy + \int_0^\infty g(y) h(y) dy \\ &= -\int_0^\infty g(y) h(y) dy + \int_0^\infty g(y) h(y) dy = 0 \, . \end{aligned}$$

Of course this means that $\dot{l}_1$ and $\dot{l}_2$ are orthogonal when f is symmetric.

C. If $f(x) = \exp(-x)\exp(-\exp(-x))$, the Gumbel (or extreme value distribution) of part E of problem 1, problem set 7, then ... For $f(x) = \exp(-x)\exp(-\exp(-x))$,

$$\log f(x) = -x - \exp(-x),$$

and

$$-\frac{f'}{f}(x) = 1 - \exp(-x),$$

while

$$-1 - x\frac{f'}{f}(x) = -1 + x(1 - \exp(-x)).$$

In problem set 7, problem 2, we computed $I_{location}(f) = 1$. Recall that in showing this we used the fact that $\exp(-X) \equiv Y \sim$ Exponential(1). Now

$$\begin{aligned} I_{scale}(f) &= \int (-1 + x(1 - \exp(-x)))^2 \exp(-x) \exp(-\exp(-x)) dx \\ &= \int_0^\infty (-1 + \log(y)(1 - y))^2 \exp(-y) dy \\ &= \pi^2/6 + (1 - \gamma)^2 \end{aligned}$$

just as in our calculations of the information for β in the Weibull family. (In fact, Weibull and the Gumbel, or double exponential extreme-value distribution are related by a log transformation. Finally,

$$\begin{aligned} \int y \left\{ -\frac{f'(y)}{f(y)} \right\}^2 f(y) dy &= \int (1 - \exp(-x))(-1 + x(1 - \exp(-x)))e^{-x} \exp(-e^{-x}) dx \\ &= \int_0^\infty (1 - y)(-1 - \log(y)(1 - y))e^{-y} dy \\ &= -(1 - \gamma) \neq 0\,, \end{aligned}$$

just as in the Weibull example. Thus the information matrix for $\theta = (\theta_1, \theta_2)$ is, in this case,

$$I(\theta) = \frac{1}{\theta_2^2} \begin{pmatrix} 1 & a \\ a & b^2 \end{pmatrix}$$

where $a \equiv -(1 - \gamma)$ and $b^2 \equiv \pi^2/6 + (1 - \gamma)^2$.