Statistics 581, Problem Set 7 Solutions

Wellner; 11/21/2002

1. Compute and plot the *score for location*, $-(f'/f)(x)$ when:
   A. $f(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, (normal or Gaussian);
   B. $f(x) = \exp(-x)/(1+\exp(-x))^2$, (logistic);
   C. $f(x) = \frac{1}{2}\exp(-|x|)$, (double exponential);
   D. $f = t_k$, the $t-$distribution with k degrees of freedom;
   E. $f(x) = \exp(-x)\exp(-\exp(-x))$, Gumbel or extreme value.

   Soluton: A. For $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, it follows that $\log f(x) = -x^2/2 + \textit{constant}$ so that $(-f'/f)(x) = x$, $-1 - x(f'/f)(x) = x^2 - 1$.
   B. For $f(x) = e^{-x}/(1+e^{-x})^2$, $\log f(x) = -x - 2\log(1+e^{-x})$ and

$$-\frac{f'}{f}(x) = \frac{1-e^{-x}}{1+e^{-x}},$$

while

$$-1 - x\frac{f'}{f}(x) = x\frac{1-e^{-x}}{1+e^{-x}} - 1 \sim |x| - 1 \qquad \text{as} \qquad |x| \to \infty.$$

C. For $f(x) = 2^{-1}\exp(-|x|)$,

$$\log f(x) = -|x| + \textit{constant},$$

and

$$-\frac{f'}{f}(x) = \begin{cases} -1 & x < 0 \\ \text{undefined} & x = 0 \\ +1 & x > 0 \end{cases},$$

while

$$-1 - x\frac{f'}{f}(x) = |x| - 1, \qquad \text{for} \qquad x \neq 0.$$

D. For the t_k distribution, $f(x) = \frac{\Gamma(\frac{1}{2}(k+1))}{\Gamma(\frac{1}{2}k)}\frac{1}{\sqrt{\pi k}}(1+\frac{x^2}{k})^{-(k+1)/2}$,

$$\log f(x) = -\frac{k+1}{2}\log(1+\frac{x^2}{k}),$$

and

$$-\frac{f'}{f}(x) = \frac{k+1}{k}\frac{x}{1+\frac{x^2}{k}},$$

while

$$-1 - x\frac{f'}{f}(x) = k\frac{x^2-1}{x^2+k}.$$

E. For $f(x) = \exp(-x)\exp(-\exp(-x))$,

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

and

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

while

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

Plots of these score functions are given in Figure 1. Note that they are *odd functions* in cases A-D, which are all symmetric densities about zero.

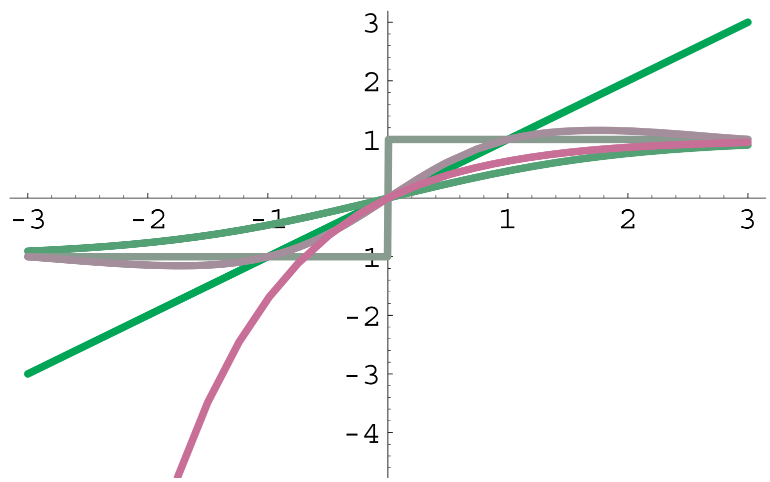

Figure 1: Scores for location.

2. Compute $I_f = \int (f'(x)/f(x))^2 f(x)dx$, the information for location, for each of the densities in problem 1.

Solution: A. In this case $I_f = \int x^2 \phi(x)dx = Var(Z) = 1$ where $Z \sim N(0,1)$.
B. For the logistic density the information for location is

$$\begin{aligned} I_f &= \int_{-\infty}^{\infty} (\frac{1-e^{-x}}{1+e^{-x}})^2 dF(x) \\ &= \int_{-\infty}^{\infty} (2F(x)-1)^2 dF(x) \\ &= \int_0^1 (2u-1)^2 du = 4Var(U) \\ &= 4\frac{1}{12} = \frac{1}{3}. \end{aligned}$$

C. For the double-exponential density, $[(-f'/f)(x)]^2 = 1$, so $I_f = 1$.
D. For the $t-$ distribution with k degrees of freedom, by using a change of

variables and letting T_r denote a random variable with the $t-$ distribution with r degrees of freedom,

$$\begin{aligned} I_f &= \int_{-\infty}^{\infty} (\frac{k+1}{k})^2 \frac{x^2}{(1+x^2/k)^2} \frac{\Gamma(k+\frac{1}{2})}{\Gamma(\frac{k}{2})\sqrt{\pi k}} \frac{1}{(1+x^2/k)^{(k+1)/2}} dx \\ &= \frac{(k+1)(k+2)}{(k+4)(k+3)} Var(T_{k+4}) \\ &= \frac{(k+1)(k+2)}{(k+4)(k+3)} \frac{k+4}{k+2} = \frac{k+1}{k+3} \end{aligned}$$

since $Var(T_r) = r/(r-2)$ for $r > 2$.

E. For the extreme value distribution $F(x) = \exp(-\exp(-x))$ and therefore if $X \sim F$, the random variable $Y \equiv \exp(-X) \sim$ exponential(1):

$$\begin{aligned} P(Y \geq y) &= P(\exp(-X) \geq y) = P(X \leq -\log(y)) \\ &= \exp(-\exp(\log(y))) = \exp(-y). \end{aligned}$$

Since $-(f'/f)(x) = -1 + e^{-x}$, it is easy to see that

$$I_f = E[-\frac{f'}{f}(X)]^2 = E[\exp(-X) - 1]^2 = E[Y-1]^2 = Var(Y) = 1.$$

3. Suppose that $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, $\Theta \subset R^k$ is a parametric model satisfying the hypotheses of the multiparameter Cramér - Rao inequality. Partition θ as $\theta = (\nu, \eta)$ where $\nu \in R^m$ and $\eta \in R^{k-m}$ and $1 \leq m < k$. Let $\dot{l} = \dot{l}_\theta = (\dot{l}_1, \dot{l}_2)$ be the corresponding partition of the (vector of) scores $\dot{l}$, and, with $\widetilde{l} \equiv I^{-1}(\theta)\dot{l}$, the *efficient influence function* for θ, let $\widetilde{l} = (\widetilde{l}_1, \widetilde{l}_2)$ be the corresponding partition of $\widetilde{l}$. In both cases, $\dot{l}_1$, $\widetilde{l}_1$ are $m-$vectors of functions, and $\dot{l}_2$, $\widetilde{l}_2$ are $k-m$ vectors. Partition $I(\theta)$ and $I^{-1}(\theta)$ correspondingly as

$$I(\theta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}$$

where I_{11} is $m \times m$, I_{12} is $m \times (k-m)$, I_{21} is $(k-m) \times m$, I_{22} is $(k-m) \times (k-m)$. Also write

$$I^{-1}(\theta) = [I^{ij}]_{i,j=1,2}.$$

Verify that:

A. $I^{11} = I_{11\cdot 2}^{-1}$ where $I_{11\cdot 2} \equiv I_{11} - I_{12} I_{22}^{-1} I_{21}$,
$I^{22} = I_{22\cdot 1}^{-1}$ where $I_{22\cdot 1} \equiv I_{22} - I_{21} I_{11}^{-1} I_{12}$,
$I^{12} = -I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1}$,
$I^{21} = -I_{22\cdot 1}^{-1} I_{21} I_{11}^{-1}$.
This amounts to formulas (5) and (6) of section 3.2, page 15.

B. Verify that
$\widetilde{l}_1 = I^{11}\dot{l}_1 + I^{12}\dot{l}_2 = I_{11\cdot 2}^{-1}(\dot{l}_1 - I_{12} I_{22}^{-1} \dot{l}_2)$, and
$\widetilde{l}_2 = I^{21}\dot{l}_1 + I^{22}\dot{l}_2 = I_{22\cdot 1}^{-1}(\dot{l}_2 - I_{21} I_{11}^{-1} \dot{l}_1)$.
The first of these is (7) on page 15, section 3.2.

Solution: A. This is just block inversion/multiplication of matrices:

$$\begin{aligned}
\begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix} \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} &= \begin{pmatrix} I_{11\cdot}^{-1} & -I_{11\cdot 2}^{-1} I_{21} I_{22}^{-1} \\ -I_{22\cdot 1}^{-1} I_{21} I_{11}^{-1} & I_{22\cdot 1} \end{pmatrix} \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} \\
&= \begin{pmatrix} I_{11\cdot 2}^{-1}(I_{11} - I_{12} I_{22}^{-1} I_{21}) & I_{11\cdot 2}^{-1}(I_{12} - I_{12}) \\ I_{22\cdot 1}^{-1}(-I_{21} + I_{21}) & I_{22\cdot 1}^{-1}(-I_{21} I_{11} I_{12} + I_{22}) \end{pmatrix} \\
&= \begin{pmatrix} Ident & 0 \\ 0 & Ident \end{pmatrix} = Identity.
\end{aligned}$$

by using the definition of $I_{11\cdot 2}$ and $I_{22\cdot 1}$.

B. This follows immediately from the formulas for I^{11} and I^{12} by just plugging into the formula $\tilde{l}_1 = I^{11}\dot{l}_1 + I^{12}\dot{l}_2$ for $\tilde{l}_1$:

$$\begin{aligned}
\tilde{l}_1 &= I_{11\cdot 2}^{-1}\dot{l}_1 - I_{11\cdot 2}^{-1} I_{12} I_{22}^{-1}\dot{l}_2 \\
&= I_{11\cdot 2}^{-1}(\dot{l}_1 - I_{12} I_{22}^{-1}\dot{l}_2) = I_{11\cdot 2}^{-1} l_1^* .
\end{aligned}$$

4. Suppose that we want to model the survival of twins with a common genetic defect, but with one of the two twins receiving some treatment. Let X represent the survival time of the untreated twin and let Y represent the survival time of the treated twin. One (overly simple) preliminary model might be to assume that X and Y are independent with Exponential(η) and Exponential($\theta\eta$) distributions, respectively:

$$f_{\theta,\eta}(x,y) = \eta e^{-\eta x} \eta\theta e^{-\eta\theta y} 1_{(0,\infty)}(x) 1_{(0,\infty)}(y)$$

A. One crude approach to estimation in this problem is to reduce the data to $W = X/Y$, the maximal invariant for the group of scale changes $g(x,y) = (cx, cy)$ with $c > 0$. Find the distribution of W, and compute the Cramér-Rao lower bound for unbiased estimates of θ based on W.

B. Find the information bound for estimation of θ based on observation of (X,Y) pairs when η is known and unknown.

C. Compare the bounds you computed in A and B and discuss the pros and cons of reducing to estimation based on the W.

Solution: A. We compute, for $w \geq 0$,

$$\begin{aligned}
P(W > w) &= P(X/Y > w) = P(X > wY) \\
&= \int_0^\infty \int_{wy}^\infty \eta^2 \theta e^{-\eta x} e^{-\eta\theta y} dx dy \\
&= \int_0^\infty \eta\theta e^{-\eta\theta y} \left(\int_{wy}^\infty \eta e^{-\eta x} dx \right) dy \\
&= \int_0^\infty \eta\theta e^{-\eta\theta y} e^{-\eta\theta y} dy \\
&= \eta\theta \int_0^\infty e^{-\eta(\theta+w)y} dy = \frac{\theta}{\theta + w} .
\end{aligned}$$

[Alternatively, $\eta X \sim \text{Exp}(1)$, $\theta\eta Y \sim \text{Exp}(1)$ are independent so $2\eta X \sim \chi_2^2$, $2\theta\eta Y \sim \chi_2^2$ are independent. Thus $W/\theta = (2\eta X/2)/(2\eta\theta Y/2) \sim F_{2,2}$ with density given by (1.2.13).] Thus the density of W is given by

$$f_W(w;\theta) = \frac{\theta}{(\theta + w)^2} 1_{(0,\infty)}(w) .$$

Hence the score for θ based on observation of W is

$$\dot{\mathbf{l}}_\theta(w) = \frac{1}{\theta} - \frac{2}{\theta + w},$$

and the information for θ based on W is

$$\begin{aligned} I_W(\theta) &= E_\theta(\dot{\mathbf{l}}_\theta(W)^2) = -E_\theta \ddot{\mathbf{l}}_\theta \\ &= \frac{1}{\theta^2} - 2\int_0^\infty \frac{\theta}{(\theta+w)^4} dw = \frac{1}{3\theta^2}. \end{aligned}$$

Hence the information bound for estimation of θ based on observation of W is $3\theta^2$.
B. When we observe (X, Y), the scores for θ and η are given by

$$\dot{\mathbf{l}}_\theta(x,y) = \frac{1}{\theta} - \eta y, \qquad \dot{\mathbf{l}}_\eta(x,y) = \frac{2}{\eta} - (x + \theta y),$$

and the second derivatives are

$$\ddot{\mathbf{l}}_{\theta\theta}(x,y) = -\theta^{-2}, \qquad \ddot{\mathbf{l}}_{\eta\eta}(x,y) = -2/\eta^2, \qquad \text{and} \qquad \ddot{\mathbf{l}}_{\theta\eta}(x,y) = -y.$$

Hence the information matrix for (θ, η) is given by

$$I(\theta,\eta) = \begin{pmatrix} 1/\theta^2 & 1/(\theta\eta) \\ 1/(\theta\eta) & 2/\eta^2 \end{pmatrix}.$$

Thus when η is known, the information for θ is $1/\theta^2$ and the information bound based on observation of (X, Y) is θ^2. When η is unknown the information for θ is

$$\begin{aligned} I_{\theta\theta\cdot\eta} &= I_{11\cdot 2} = I_{11} - I_{12} I_{22}^{-1} I_{21} \\ &= 1/\theta^2 - (\theta\eta)^{-2}\eta^2/2 = 1/(2\theta^2), \end{aligned}$$

and the information bound for estimation of θ is $2\theta^2$. Thus lack of knowledge of η costs a factor of two in the bound.
C. Reduction to W cost a factor of 3 in the bound as compared to the bound based on (X, Y) when η is known and a factor of $3/2$ in the bound based on (X, Y) when η unknown. Thus reduction to W does not seem to be advisable. We can do better by basing estimation on *both* X and Y!

5. This is a continuation of the preceding problem. A more realistic model involves assuming that the common parameter η for the two twins varies across sets of twins. There are several different ways of modeling this: one approach involves supposing that each pair of twins observed (X_i, Y_i) has its own fixed parameter η_i, $i = 1, \ldots, n$. In this model we observe (X_i, Y_i) with density f_{θ,η_i} for $i = 1, \ldots, n$; i.e.

$$f_{\theta,\eta_i}(x_i, y_i) = \eta_i e^{-\eta_i x_i} \eta_i \theta e^{-\eta_i \theta y_i} 1_{(0,\infty)}(x_i) 1_{(0,\infty)}(y_i). \tag{0.1}$$

This is sometimes called a "functional model" (or model with incidental nuisance parameters).

Another approach is to assume that $\eta \equiv Z$ has a distribution, and that our observations are from the mixture distribution. Assuming (for simplicity) that

$Z = \eta \sim \text{Gamma}(a, b)$ with density $g_{a,b}(\eta)$, it follows that the (marginal) distribution of (X, Y) is

$$\begin{aligned} p_{\theta,a,b}(x,y) &= \int_0^\infty f_{\theta,z}(x,y) g_{a,b}(z) dz \\ &= \frac{\theta}{b^2} \left(\frac{b}{b + x + \theta y} \right)^{a+2} \frac{\Gamma(a+2)}{\Gamma(a)} . \end{aligned} \tag{0.2}$$

This is sometimes called a "structural model" (or mixture model).

(a) Find the information for θ in the functional model.

(b) Find the information for θ in the structural model.

(c) Compare the information bounds you computed in (a) and (b). When is the information for θ in the functional model larger than the information for θ in the structural model?

Solution: (a) The density of the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ is given by

$$p_{\theta,\underline{\eta}}(\underline{x}, \underline{y}) = \prod_{i=1}^n \left\{ \eta_i^2 \theta \exp(-\eta_i x_i) \exp(-\eta_i \theta y_i) \right\} = \theta^n \prod_{i=1}^n \eta_i^2 \exp(-\eta_i x_i) \exp(-\eta_i \theta y_i) .$$

Hence we calculate

$$\dot{\mathrm{l}}_\theta(\underline{x}, \underline{y}) = \frac{n}{\theta} - \sum_{i=1}^n \eta_i y_i , \qquad \dot{\mathrm{l}}_{\eta_i}(\underline{x}, \underline{y}) = \frac{2}{\eta_i} - (x_i + \theta y_i) ,$$

$$\ddot{\mathrm{l}}_{\theta\theta}(\underline{x}, \underline{y}) = \frac{-n}{\theta^2} , \qquad \ddot{\mathrm{l}}_{\eta_i\eta_i}(\underline{x}, \underline{y}) = \frac{-2}{\eta_i^2} , \qquad \ddot{\mathrm{l}}_{\theta\eta_i}(\underline{x}, \underline{y}) = -y_i .$$

It follows easily that the information matrix for $(\theta, \underline{\eta})$ is given by

$$I_n(\theta, \underline{\eta}) = I_{\underline{X},\underline{Y}}(\theta, \underline{\eta}) = \begin{pmatrix} n/\theta^2 & 1/(\gamma\theta_1) & \cdots & \cdots & 1/(\theta\eta_n) \\ 1/(\theta\eta_1) & 2/\eta_1^2 & 0 & \cdots & 0 \\ . & 0 & . & . & 0 \\ . & . & . & . & . \\ . & . & . & . & 0 \\ 1/(\theta\eta_n) & 0 & . & 0 & 2/\eta_n^2 \end{pmatrix} ,$$

Thus it follows that

$$\begin{aligned} I_{11\cdot 2} &= I_{11} - I_{12} I_{22}^{-1} I_{21} \\ &= \frac{n}{\theta^2} - \frac{1}{\theta^2} \left(\frac{1}{\eta_1}, \ldots, \frac{1}{\eta_n} \right) \begin{pmatrix} \eta_1^2/2 & 0 & \ldots & 0 \\ 0 & \eta_2^2/2 & . & 0 \\ . & . & . & . \\ 0 & . & . & \eta_n^2/2 \end{pmatrix} \begin{pmatrix} 1/\eta_1 \\ . \\ . \\ 1/\eta_n \end{pmatrix} \\ &= \frac{n}{\theta^2} - \frac{n}{\theta^2} \frac{1}{2} = \frac{n}{2\theta^2} , \end{aligned}$$

and this is the information for θ in the presence of the nuisance parameters $\eta_1, \ldots, \eta_n$.

(b) and (c): For the structural model, first note that $\Gamma(a+2)/\Gamma(a) = a(a+1)$. Then we compute the scores:

$$\begin{aligned}
\dot{l}_\theta(x,y) &= \frac{1}{\theta} - \frac{(a+2)y}{b+x+\theta y}, \\
\dot{l}_a(x,y) &= \frac{\Gamma'}{\Gamma}(a+2) - \frac{\Gamma'}{\Gamma}(a) + \log\left(\frac{b}{b+x+\theta y}\right) \\
&= \frac{1}{a} + \frac{1}{a+1} + \log\left(\frac{b}{b+x+\theta y}\right), \\
\dot{l}_b(x,y) &= \frac{a}{b} - \frac{a+2}{b+x+\theta y}.
\end{aligned}$$

Furthermore, the second derivatives of the scores are:

$$\begin{aligned}
\ddot{l}_{\theta\theta}(x,y) &= -\frac{1}{\theta^2} + \frac{(a+2)y^2}{(b+x+\theta y)^2}, \\
\ddot{l}_{aa}(x,y) &= \psi'(a+2) - \psi'(a), \qquad \text{where } \psi(x) = \frac{\Gamma'}{\Gamma}(x) \\
&= -\frac{1}{a^2} - \frac{1}{(a+1)^2}, \\
\ddot{l}_{bb}(x,y) &= -\frac{a}{b^2} + \frac{a+2}{(b+x+\theta y)^2}, \\
\ddot{l}_{\theta a}(x,y) &= -\frac{y}{b+x+\theta y}, \\
\ddot{l}_{\theta b}(x,y) &= \frac{(a+2)y}{(b+x+\theta y)^2}, \\
\ddot{l}_{ba}(x,y) &= \frac{1}{b} - \frac{1}{b+x+\theta y}.
\end{aligned}$$

It follows (after some computation; I used Mathematica), that the information matrix for (θ, a, b) is:

$$I(\theta,a,b) = \begin{pmatrix} \frac{a+1}{a+3}\frac{1}{\theta^2} & \frac{1}{(a+2)\theta} & \frac{-a}{(a+3)b\theta} \\ \frac{1}{(a+2)\theta} & \frac{1}{a^2} + \frac{1}{(a+1)^2} & \frac{-2}{(a+2)b} \\ \frac{-a}{(a+3)b\theta} & \frac{-2}{(a+2)b} & \frac{2a}{(a+3)b^2} \end{pmatrix}. \tag{0.3}$$

Hence the information for θ in the structural model is, with

$$D \equiv b^2 \det(I_{22}) = \left(\frac{2a}{a+3}(a^{-2} + (a+1)^{-2}) - \frac{4}{(a+2)^2}\right),$$

$$\begin{aligned}
&I_{11\cdot 2}(\theta,a,b) = I_{11} - I_{12} I_{22}^{-1} I_{21} \\
&= \frac{a+1}{(a+3)\theta^2} \\
&\quad - \begin{pmatrix} \frac{1}{(a+2)\theta} & \frac{-a}{(a+3)b\theta} \end{pmatrix} \frac{1}{D} \begin{pmatrix} \frac{2a}{(a+3)b^2} & \frac{2}{(a+2)b} \\ \frac{2}{(a+2)b} & a^{-2} + (a+1)^{-2} \end{pmatrix} \begin{pmatrix} \frac{1}{(a+2)\theta} \\ \frac{-a}{(a+3)b\theta} \end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{a+1}{(a+3)\theta^2} \\
&\quad - \begin{pmatrix} \frac{1}{(a+2)\theta} & \frac{-a}{(a+3)b\theta} \end{pmatrix} \frac{1}{D} \begin{pmatrix} \frac{2a}{(a+3)b^2} & \frac{2}{(a+2)b} \\ \frac{2}{(a+2)b} & \frac{a+2}{a^2(a+1)^2} \end{pmatrix} \begin{pmatrix} \frac{1}{(a+2)\theta} \\ \frac{-a}{(a+3)b\theta} \end{pmatrix} \\
&= \frac{1}{\theta^2} \left\{ \frac{a+1}{a+3} - \frac{2a}{(a+3)(a+2)^2} \left(\frac{(a+2)}{2a^2(a+1)^2} \frac{a}{a+3}(a+2)^2 - 1 \right) \frac{1}{D} \right\} \\
&= \frac{1}{2\theta^2} \frac{2+a}{3+a}
\end{aligned}
$$

after a bit of algebra (I used Mathematica again). Note that this equals $1/(3\theta^2)$ when $a = 0$, and it increases to $1/(2\theta^2)$ as $a \to \infty$.

For the semiparametric generalization of the mixture (structural) model given by (0.2), we have

$$
p_{\theta,a,b}(x, y) = \int_0^\infty f_{\theta,z}(x, y) dG(z)
$$

where G is an *arbitrary* (mixing) distribution on $(0, \infty)$. In fact, the information for θ in this larger model has the same qualitative feature as in the Gamma-mixture submodel:

$$
I_{11\cdot 2}(\theta) = \frac{1}{3\theta^2} + \frac{1}{12\theta^2} I_{p_T}(scale)
$$

where $I_{p_T}(scale)$ is the information for scale in for the density

$$
p_T(t) \equiv t \int_0^\infty z^2 \exp(-tz) dG(z) \,.
$$

It is easily seen that this information is always greater than $1/(3\theta^2)$ and always less than or equal to $1/(2\theta^2)$. See Bickel, Klaassen, Ritov, and Wellner (1993), pages 134 - 135 for this calculation. Section 4.5 of BKRW has much more on information calculations for semiparametric mixture models.