Statistics 581, Problem Set 7 Solutions; Corrected

Wellner; 11/23/2001

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.

   Solution: A. For $f(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, it follows that $\log f(x) = -x^2/2 +$ *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

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

See the plots on the last page.

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):

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

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. Lehmann and Casella, TPE, Problem 6.6, page 142: If

$$p(x) \equiv p(x;\epsilon,\xi,\tau) \equiv (1-\epsilon)\phi(x-\xi) + \frac{\epsilon}{\tau}\phi\left(\frac{x-\xi}{\tau}\right),$$

find $I(\epsilon,\xi,\tau)$.

Solution: I would prefer to write $\theta = (\xi,\epsilon,\tau)$. The first thing to note about the model

$$\mathcal{P} = \{P_\theta : (dP_\theta/d\lambda)(x) = p(x;\theta) = p(x)\}$$

is that it is a location family: $p(x;\xi,\epsilon,\tau) = p_0(x-\xi;\epsilon,\tau)$ where

$$p_0(x;\epsilon,\tau) = (1-\epsilon)\phi(x) + (\epsilon/\tau)\phi(x/\tau).$$

Moreover, all the distributions in the family $\mathcal{P}_0 = \{P_{0,\epsilon,\tau} : \epsilon \in [0,1], \tau > 0\}$ are *symmetric about* 0. Here are plots of these densities for $\tau = 3$ and $\epsilon = 0, .2, .4, .6, .8, 1$; the second plot is of the densities p_0 for $\epsilon = .2$ and $\tau = 1.5, 4, 8, 16, 32, 64$.

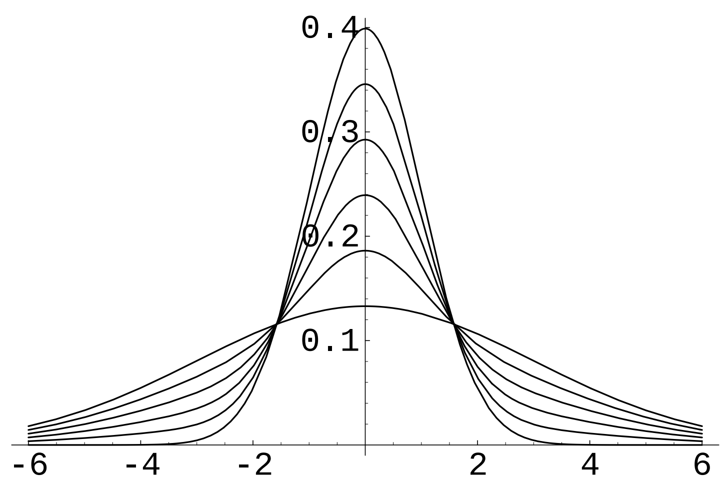

Figure 1: Densities $p(x)$ for $\xi = 0$, $\tau = 3$, $\epsilon = 0, .2, .4, .6, .8, 1.0$

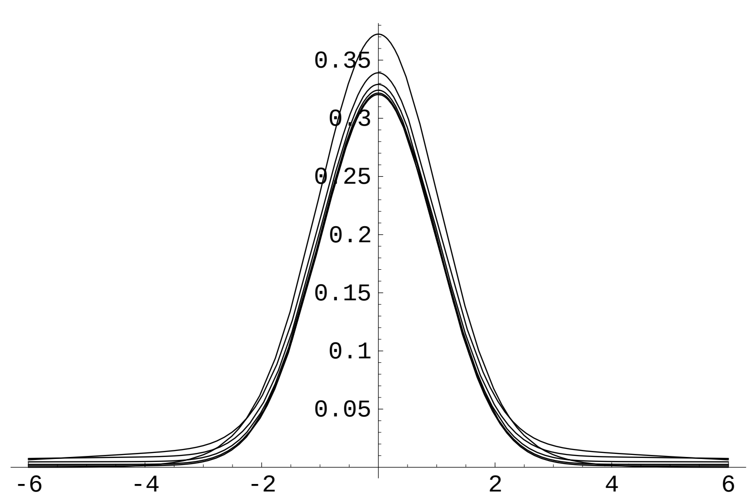

Figure 2: Densities $p(x)$ for $\xi = 0$, $\epsilon = .2$, $\tau = 1.5, 4, 8, 16, 32, 64$

Since $\phi'(x) = -x\phi(x)$, the score functions for ϵ, ξ, and τ are given by

$$\begin{aligned}
\dot{l}_\xi(x) &= \frac{1}{p(x)}\left\{(x-\xi)(1-\epsilon)\phi(x-\xi) + \frac{x-\xi}{\tau^2}\frac{\epsilon}{\tau}\phi\left(\frac{x-\xi}{\tau}\right)\right\}, \\
\dot{l}_\epsilon(x) &= \frac{1}{p(x)}\left\{\frac{1}{\tau}\phi\left(\frac{x-\xi}{\tau}\right) - \phi(x-\xi)\right\}, \\
\dot{l}_\tau(x) &= \frac{1}{p(x)}\frac{\epsilon}{\tau^2}\phi\left(\frac{x-\xi}{\tau}\right)\left\{\left(\frac{x-\xi}{\tau}\right)^2 - 1\right\}.
\end{aligned}$$

Thus with $\theta \equiv (\xi, \epsilon, \tau)$ and $\underline{\dot{l}}_\theta$, the information matrix $I(\xi, \epsilon, \tau) = I(\theta)$ is given by

$$\begin{aligned}
I(\theta) &= E_\theta\{\underline{\dot{l}}_\theta(X)\underline{\dot{l}}_\theta(X)^T\} \\
&= \left(E_\theta(\dot{l}_i(X)\dot{l}_j(X))\right).
\end{aligned}$$

Hence it becomes clear that all of the elements of $I(\epsilon, \xi, \tau)$ are constant functions of ξ; hence it suffices to compute the information matrix for $\xi = 0$. Thus we take $\xi = 0$ in the rest of the argument. Now note that the scores for ϵ and τ are even functions of x: $\dot{l}_\epsilon(-x) = \dot{l}_\epsilon(x)$ and similarly for $\dot{l}_\tau$. On the other hand, the score function for ξ is an odd function of x: $\dot{l}_\xi(-x) = -\dot{l}_\xi(x)$. It follows that

$$E_\theta\{\dot{l}_\xi(X)\dot{l}_\epsilon(X)\} = 0, \qquad \text{and} \qquad E_\theta\{\dot{l}_\xi(X)\dot{l}_\tau(X)\} = 0\,.$$

Thus the only non-zero entry off the diagonal is $I_{23}(\theta) = I_{\epsilon,\tau}(\theta)$. I do not know any "closed form" results for $I_{11}(\theta)$, $I_{22}(\theta)$, $I_{33}(\theta)$, or $I_{23}(\theta)$, but it is not hard to compute them as functions of $\theta = (\epsilon, \xi, \tau)$ especially since they depend only on (ϵ, τ). See Figures 3-5 below.

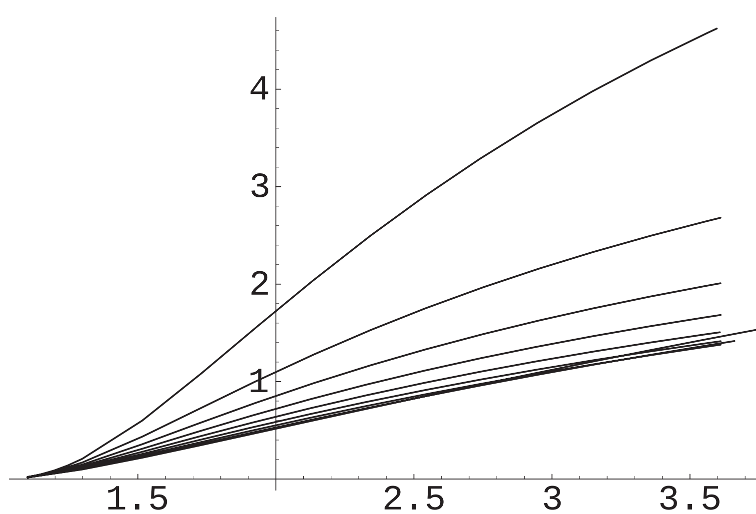

Figure 3: Information for ϵ as a function of τ for $\epsilon = .1, .2, .3, \ldots, .9$

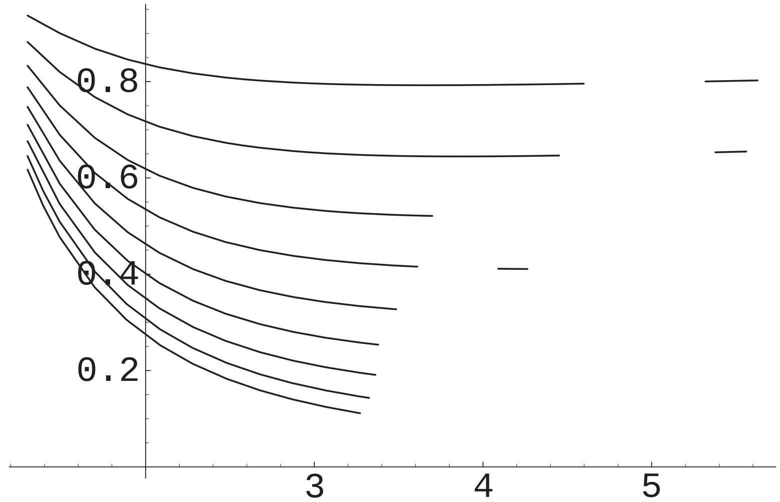

Figure 4: Information for ξ as a function of τ for $\epsilon = .1, .2, .3, \ldots, .9$

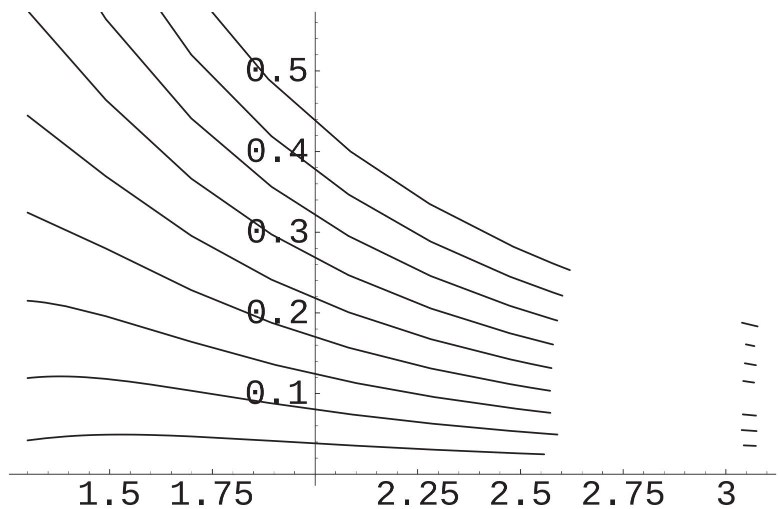

Figure 5: Information for τ as a function of τ for $\epsilon = .1, .2, .3, \ldots, .9$

4. 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 $1 \le m < k$. 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 (3) and (4) of section 3.2, page 14.
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
$\widetilde{l}_2 = I^{21}\dot{l}_1 + I^{22}\dot{l}_2 = I_{22\cdot 1}^{-1}(\dot{l}_2 - I_{21} I_{11}^{-1} \dot{l}_1)$.

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

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

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 $\tilde{l}_1$:

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

5. 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:

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

In class we computed the Cramér-Rao lower bound for unbiased estimates of θ based on $Z = X/Y$, the maximal invariant for the group of scale changes $g(x,y) = (cx, cy)$ with $c > 0$. We also compared this bound to the information bounds for estimation of θ based on observation of (X, Y) when η is known and unknown.

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

$$\begin{aligned} p_{\theta,a,b}(x, y) &= \int_0^\infty f_{\theta,z}(x, y) g_{a,b}(z) dz \\ &= \frac{\theta}{b^2}\left(\frac{b}{b + x + \theta y}\right)^{a+2} \frac{\Gamma(a + 2)}{\Gamma(a)} \qquad (0.2) \\ &= \frac{\theta}{b^2}\left(\frac{b}{b + x + \theta y}\right)^{a+2} a(a + 1)\,. \qquad (0.3) \end{aligned}$$

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) Here the parameter is $\gamma = (\theta, \eta_1, \ldots, \eta_n) \in \Theta \subset R^{n+1}$. For (X_i, Y_i) the log of the density is

$$\log p_{\theta,\underline{\eta}}(x_i, y_i) = 2 \log \eta_i + \log \theta - \eta_i(x_i + \theta y_i)\,,$$

so the scores are

$$\dot{l}_\theta(x_i, y_i) = \frac{1}{\theta} - \eta_i y_i \qquad \dot{l}_{\eta_i}(x_i, y_i) = \frac{2}{\eta_i} - (x_i + \theta y_i), \qquad \dot{l}_{\eta_j}(x_i, y_i) = 0, \quad j \neq i\,.$$

Thus we find that the information matrix for γ based on (X_i, Y_i) is $I_i = (I_{i,j,k})$ with $I_{i,1,1} = 1/\theta^2$, $I_{i,i,i} = 2/\eta_i^2$, $I_{i,1,i} = I(i, i, 1) = 1/(\theta\eta_i)$, and $I_{i,j,k} = 0$ otherwise (i.e. $j \neq i$ or $k \neq i$). By independence of (X_i, Y_i) for $i = 1, \ldots, n$, the information matrix for γ based on observation of all the data is

$$I_n(\gamma) = \sum_{i=1}^n I_i(\gamma) = \begin{pmatrix} n/\theta^2 & 1/(\theta\eta_1) & \cdots & \cdots & 1/(\theta\eta_n) \\ 1/(\theta\eta_1) & 2/\eta_1^2 & 0 & \cdots & 0 \\ . & 0 & . & . & 0 \\ . & 0 & . & . & 0 \\ 1/(\theta\eta_n) & 0 & \cdots & \cdots & 2/\eta_n^2 \end{pmatrix}.$$

Thus

$$I_{11\cdot2}(\gamma) = I_{11} - I_{12} I_{22}^{-1} I_{21}$$

$$
= \frac{n}{\theta^2} - \frac{1}{\theta^2}(\frac{1}{\eta_1}, \ldots, \frac{1}{\eta_n}) \begin{pmatrix} \eta_1^2/2 & 0 & . & . \\ 0 & \eta_2^2/2 & 0 & . \\ . & . & . & . \\ 0 & . & . & \eta_n^2/2 \end{pmatrix} \begin{pmatrix} \frac{1}{\eta_1} \\ . \\ . \\ \frac{1}{\eta_n} \end{pmatrix}
$$

$$
= \frac{n}{\theta^2} - \frac{n}{2\theta^2} = \frac{n}{2\theta^2} .
$$

(b) 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:

$$
I(\theta, a, b) = \begin{pmatrix} \frac{a+1}{a+3}\frac{1}{\theta^2} & \frac{1}{(a+2)\theta} & \frac{-a}{(a+3)b\theta} \\ \frac{1}{(a+2)\theta} & a^{-2} + (a+1)^{-2} & \frac{-2}{(a+2)b} \\ \frac{-a}{(a+3)b\theta} & \frac{-2}{(a+2)b} & \frac{2a}{(a+3)b^2} \end{pmatrix} . \tag{0.4}
$$

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

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

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

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 (??), 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.