Statistics 581, Problem Set 7 Solutions

Wellner; 11/15/2018

1. Suppose that $X \sim \text{Beta}(\alpha, \beta)$; i.e. X has density p_θ given by

$$p_\theta(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} 1_{(0,1)}(x), \quad \theta = (\alpha, \beta) \in (0,\infty) \times (0,\infty) \equiv \Theta \, .$$

Consider estimation of:
A. $q_A(\theta) \equiv E_\theta X$. B. $q_B(\theta) \equiv F_\theta(x_0)$ for a fixed x_0; here $F_\theta(x) \equiv P_\theta(X \le x)$.
(i) Compute $I(\theta) = I(\alpha, \beta)$; compare Lehmann & Casella page 127, Table 6.1
(ii) Compute $q_A(\theta)$, $q_B(\theta)$, $\dot{q}_A(\theta)$, and $\dot{q}_B(\theta)$.
(iii) Find the efficient influence functions for estimation of q_A and q_B.
(iv) Compare the efficient influence functions you find in (iii) with the influence functions ψ_A and ψ_B of the natural nonparametric estimators $\overline{X}_n$ and $\mathbb{F}_n(x_0)$ respectively. Does $\psi_A \in \dot{\mathcal{P}}$? Does $\psi_B \in \dot{\mathcal{P}}$ hold?

Solution: For the Beta(α, β) density:

$$p_\theta(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} 1_{(0,1)}(x).$$

Thus

$$\log p_\theta(x) = (\alpha-1)\log x + (\beta-1)\log(1-x) + \log\Gamma(\alpha+\beta) - \log\Gamma(\alpha) - \log\Gamma(\beta),$$

and hence

$$\begin{aligned} \dot{l}_\alpha(x) &= \log x + \psi(\alpha+\beta) - \psi(\alpha), \\ \dot{l}_\beta(x) &= \log(1-x) + \psi(\alpha+\beta) - \psi(\beta). \end{aligned}$$

Furthermore,

$$\begin{aligned} \ddot{l}_{\alpha\alpha}(x) &= \psi'(\alpha+\beta) - \psi'(\alpha), \\ \ddot{l}_{\alpha\beta}(x) &= \psi'(\alpha+\beta), \\ \ddot{l}_{\beta\beta}(x) &= \psi'(\alpha+\beta) - \psi'(\beta). \end{aligned}$$

Hence

$$I(\theta) = \begin{pmatrix} \psi'(\alpha) - \psi'(\alpha+\beta) & -\psi'(\alpha+\beta) \\ -\psi'(\alpha+\beta) & \psi'(\beta) - \psi'(\alpha+\beta) \end{pmatrix}. \tag{0.1}$$

This is positive definite for all $\alpha > 0$, $\beta > 0$.
(ii). Now $q_A(\theta) = \alpha/(\alpha+\beta)$, and

$$q_B(\theta) = P_\theta(X \le x_0) \quad = \quad \int_0^{x_0} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\, dx,$$

Therefore

$$
\begin{aligned}
\dot q_A^T(\theta) = \left(\frac{\partial}{\partial\alpha}q_A, \frac{\partial}{\partial\beta}q_A\right) &= (\frac{\beta}{(\alpha+\beta)^2}, -\frac{\alpha}{(\alpha+\beta)^2}) = (\alpha+\beta)^{-2}(\beta,-\alpha) \\
&= \mathrm{Cov}_\theta(X - E_\theta(X), \dot l_\theta^T(X)),
\end{aligned}
$$

while, with

$$
\begin{aligned}
\dot q_B(\theta) &= \begin{pmatrix} E_\theta(1_{(0,x_0]}(X)\log X) + (\psi(\alpha+\beta)-\psi(\alpha))F_\theta(x_0) \\ E_\theta(1_{(0,x_0]}(X)\log(1-X)) + (\psi(\alpha+\beta)-\psi(\beta))F_\theta(x_0) \end{pmatrix} \\
&= \mathrm{Cov}_\theta[(1_{[0,x_0]}(X) - F_\theta(x_0)), \dot l_\theta^T]\,.
\end{aligned}
$$

(iii). The scores are given by

$$
\dot l_\theta(x) = \begin{pmatrix} \dot l_\alpha(x) \\ \dot l_\beta(x) \end{pmatrix} = \begin{pmatrix} \log(x) - (\psi(\alpha) - \psi(\alpha+\beta)) \\ \log(1-x) - (\psi(\beta) - \psi(\alpha+\beta)) \end{pmatrix}
$$

and the information matrix is as given in (0.1) Thus

$$
I^{-1}(\theta) = \frac{1}{\det I(\theta)} \begin{pmatrix} \psi'(\beta) - \psi'(\alpha+\beta) & \psi'(\alpha+\beta) \\ \psi'(\alpha+\beta) & \psi'(\alpha) - \psi'(\alpha+\beta) \end{pmatrix}
$$

where

$$
\det(I(\theta)) = (\psi'(\alpha) - \psi'(\alpha+\beta))(\psi'(\beta) - \psi'(\alpha+\beta)) - \psi'(\alpha+\beta)^2,
$$

and the efficient influence function for estimation of q_A is

$$
\tilde l_A(x) = \dot q_A(\theta)^T I^{-1}(\theta)\dot l_\theta(x) \in \dot{\mathcal{P}}
$$

and hence is a (centered) linear combination of $\log x$ and $\log(1-x)$. Note that $X - E_\theta(X) \notin [\dot l_\theta] = \dot{\mathcal{P}}$, and hence the sample mean is inefficient for estimation of $E_\theta(X)$ in this model.
Similarly, $\tilde l_B(x) = \dot q_B(\theta) I^{-1}(\theta)\dot l_\theta(x)$; unfortunately, this does not simplify much, largely due to the fact that $1_{[0,x_0]}(X) - F_\theta(x_0) \notin [\dot l_\theta] = \dot{\mathcal{P}}$.
(iv) The information bound for estimation of q_A is

$$
\begin{aligned}
I^{-1}(P|q_A, \mathcal{P}) &= \dot q_A^T I^{-1}(\theta)\dot q_A \\
&= (\alpha+\beta)^{-4}(\beta,-\alpha)\frac{1}{\det I(\theta)} \begin{pmatrix} \psi'(\beta) - \psi'(\alpha+\beta) & \psi'(\alpha+\beta) \\ \psi'(\alpha+\beta) & \psi'(\alpha) - \psi'(\alpha+\beta) \end{pmatrix} \begin{pmatrix} \beta \\ -\alpha \end{pmatrix} \\
&< \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta-1)} = Var_\theta(X)
\end{aligned}
$$

where the inequality holds since $\psi_A(X) = (X - E_\theta(X)) \notin \dot{\mathcal{P}}$. Similarly,

$$
I^{-1}(P|q_B, \mathcal{P}) = \dot q_B^T I^{-1}(\theta)\dot q_B,
$$

which does not simplify appreciably because $1_{[0,x_0]}(X) - F_\theta(x_0) \notin [\dot l_\theta] = \dot{\mathcal{P}}$. However, since we know that $\tilde l_B = \Pi(1_{[0,x_0]}(x) - F(x_0)|\dot{\mathcal{P}})$, it follows easily that

$$
I^{-1}(P|q_B, \mathcal{P}) < E_\theta(1_{[0,x_0]}(X) - F_\theta(x_0))^2 = F_\theta(x_0)(1 - F_\theta(x_0));
$$

i.e. it is possible to improve on the natural nonparametric estimators $\overline{X}_n$ and $\mathbb{F}_n(x_0)$ of $q_A(\theta) = E_\theta(X)$ and $q_B(\theta) = F_\theta(x_0)$ when the model holds. (If we had considered $q_C(\theta) = E_\theta \log(X/(1-X))$ or $q_D(\theta) = E_\theta \log X$, this story would change! It is also an instructive exercise to consider the sub-model consisting of the beta densities with $\alpha = \beta$.)

2. Suppose that $X \sim F_\theta = \text{exponential}(\theta)$ with density $f_\theta(x) = \theta e^{-\theta x} 1_{(0,\infty)}(x)$ and $Y \sim G_\eta$ independent of X with densities $\{g_\eta : \eta \in R^+\}$, a regular parametric model on $(0, \infty)$. Consider the following three scenarios for observation of X or functions of X:
(a) Uncensored: we observe X and Y.
(b) Right-censored: we observe

$$T(X,Y) = (X \wedge Y, 1\{X \le Y\} \equiv (\min\{X,Y\}, 1\{X \le Y\}) \equiv (Z, \Delta).$$

(c) Interval-censored (case 1): we observe $S(X,Y) = (Y, 1\{X \le Y\}) \equiv (Y, \Delta)$.
(i) Find the joint density of (X,Y) and joint distributions of $T(X,Y)$ and $S(X,Y)$.
(ii) Find the scores for θ and η in each of the three scenarios (a), (b), and (c). (Let $(\partial/\partial\eta) \log g_\eta(y) \equiv a(y)$ with $a \in L_2^0(G_\eta)$.)
(iii) Compute and compare $I_{X,Y}(\theta)$, $I_{T(X,Y)}(\theta)$, and $I_{S(X,Y)}(\theta)$. Make the comparisons in general and then explicitly by making one or more choices of the family $\{g_\eta\}$.

Solution: (i) In case (a) when we observe X and Y the joint density of X, Y is simply $f_\theta(x) g_\eta(y) = \theta \exp(-\theta x) g_\eta(y)$. In case (b) the joint density $p(z,\delta) = p(z,\delta;\theta,\eta)$ (with respect to Lebesgue measure on $(0,\infty)$ times counting measure on $\{0,1\}$) is given by

$$p(z,\delta) = \{(1 - G_\eta(z)) f_\theta(z)\}^\delta \{(1 - F_\theta(z)) g_\eta(z)\}^{1-\delta}.$$

In case (c) the joint density $p(y,\delta) = p(y,\delta;\theta,\eta)$ of $S(X,Y) = (Y,\Delta)$ given by

$$p(y,\delta) = F_\theta(y)^\delta (1 - F_\theta(y))^{1-\delta} g_\eta(y).$$

(ii) In case (a),

$$\log p_{X,Y}(x,y;\theta,\eta) = \log f_\theta(x) + \log g_\eta(y) = \log\theta - \theta x + \log g_\eta(y),$$

and hence the scores for θ and η are

$$\dot{l}_\theta(x,y) = \theta^{-1} - x,$$
$$\dot{l}_\eta(x,y) = a(y).$$

In case (b) we find that

$$\begin{aligned} \log p(z,\delta;\theta,\eta) &= \delta(\log f_\theta(z) + \log(1 - G_\eta(z))\} + (1-\delta)\{\log g_\eta(z) + \log(1 - F_\theta(z))\} \\ &= \delta \log f_\theta(z) + (1-\delta)\log(1 - F_\theta(z)) + (1-\delta) g_\eta(z) + \delta(1 - G_\eta(z)). \end{aligned}$$

Thus the scores for θ and η are given by

$$\dot{l}_\theta(z,\delta) = \delta(\theta^{-1} - z) + (1-\delta)(-z) = \theta^{-1}\delta - z,$$
$$\dot{l}_\eta(z,\delta) = (1-\delta) a(z) + \delta(1 - G_\eta(z))^{-1} \int_z^\infty a(y) dG_\eta(y).$$

In case (c),

$$\log p(y, \delta; \theta, \eta) = \delta F_\theta(y) + (1-\delta)(1 - F_\theta(y)) + \log g_\eta(y).$$

Thus the scores for θ and η are given by

$$\begin{aligned}
\dot{l}_\theta(y, \delta) &= \left\{ \frac{\delta}{F_\theta(y)} \frac{\partial}{\partial\theta} F_\theta(y) + \frac{(1-\delta)}{1 - F_\theta(y)} (-\frac{\partial}{\partial\theta} F_\theta(y) \right\} \\
&= \left\{ \frac{\delta}{F_\theta(y)} - \frac{(1-\delta)}{1 - F_\theta(y)} \right\} \frac{\partial}{\partial\theta} F_\theta(y) \\
&= \left\{ \frac{\delta}{F_\theta(y)} - \frac{(1-\delta)}{1 - F_\theta(y)} \right\} (y \exp(-\theta y) \\
&= \{\delta - F_\theta(y)\} \frac{y(1 - F_\theta(y))}{F_\theta(y)(1 - F_\theta(y))}, \\
\dot{l}_\eta(y, \delta) &= a(y).
\end{aligned}$$

(iii) In case (a), the information matrix for (θ, η) is given by

$$I_{X,Y}(\theta, \eta) = \begin{pmatrix} \theta^{-2} & 0 \\ 0 & Ea^2(Y) \end{pmatrix},$$

and hence the information for θ is simply θ^{-2}.
(b) In case (b),

$$\begin{aligned}
I_{11}(\theta, \eta) &= E_{\theta,\eta} \dot{l}_\theta^2(Z, \Delta) \\
&= E_{\theta,\eta}(\theta^{-1}\Delta - Z)^2.
\end{aligned}$$

But we can also calculate

$$\ddot{l}_{\theta,\theta}(z, \delta) = -\theta^{-2}\delta,$$

and hence

$$\begin{aligned}
I_{11}(\theta, \eta) &= -E_{\theta,\eta}\ddot{l}_{\theta,\theta}(Z, \Delta) = \theta^{-2} P_{\theta,\eta}(\Delta = 1) && (0.2) \\
&= \theta^{-2} \int_0^\infty F_\theta dG_\eta = \theta^{-2} E_\eta g(\theta Y) \le \theta^{-2} && (0.3)
\end{aligned}$$

where $g(v) \equiv 1 - e^{-v}$ where the inequality is strict if $P_\eta(Y < \infty) > 0$. Note that

$$\ddot{l}_{\theta,\eta}(z, \delta) = 0,$$

and hence $I_{12}(\theta, \eta) = I_{21}(\theta, \eta) = 0$. Thus we conclude that the information for θ is simply $I_{11}(\theta, \eta) = \theta^{-2} P_{\theta,\eta}(\Delta = 1)$ as calculated in (0.3). When $Y \sim$ Exponential(η) this yields

$$\begin{aligned}
I_{11 \cdot 2}(\theta, \eta) &= \theta^{-2} \int_0^\infty (1 - \exp(-\theta y)\eta \exp(-\eta y) dy \\
&= \theta^{-2} \left\{ 1 - \eta \int_0^\infty \exp(-(\theta + \eta) y) \right\} \\
&= \theta^{-2} \left\{ 1 - \frac{\eta}{\theta + \eta} \right\} \\
&= \theta^{-2} \frac{\theta}{\theta + \eta} = \theta^{-2} \frac{1}{1 + r} \quad \text{with } r \equiv \eta/\theta.
\end{aligned}$$

(c) In case (c), since $(\Delta|Y) \sim \text{Bernoulli}(F_\theta(Y))$ we calculate conditionally on Y to find that

$$
\begin{aligned}
I_{11}(\theta,\eta) &= E_{\theta,\eta}\dot{l}_\theta(Y,\Delta)^2 \\
&= E_{\theta,\eta}\{F_\theta(Y)(1-F_\theta(Y))\}\frac{Y^2(1-F_\theta(Y))^2}{F_\theta(Y)^2(1-F_\theta(Y))^2} \\
&= \theta^{-2}E_{\theta,\eta}\frac{(\theta Y)^2(1-F_\theta(Y))}{F_\theta(Y)} \\
&= \theta^{-2}E_\eta h(\theta Y)
\end{aligned}
$$

where $h(v) \equiv v^2e^{-v}/(1-e^{-v})$ is a bounded function vanishing at 0 and ∞ and $\|h\|_\infty \le .65$; see Figure yy.

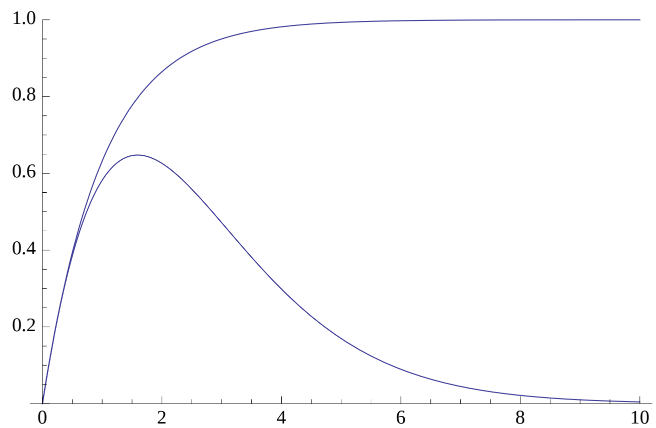

Figure 1: The functions $g(v) = 1 - e^{-v}$ and $h(v) = v^2e^{-v}/(1-e^{-v})$.

Again by computing conditionally we see that

$$
\begin{aligned}
I_{12}(\theta,\eta) &= E_{\theta,\eta}\dot{l}_\theta(Y,\Delta)\dot{l}_\eta(Y,\Delta) \\
&= E\left\{E\left\{(\Delta - F_\theta(Y))\frac{Ya(Y)}{F_\theta(Y)}|Y\right\}\right\} \\
&= E\left\{\frac{Ya(Y)}{F_\theta(Y)}E\left\{(\Delta - F_\theta(Y))|Y\right\}\right\} \\
&= 0.
\end{aligned}
$$

Thus the information for θ based on observation of $S(X,Y) = (Y,\Delta)$ is

$$
I_{11}(\theta,\eta) = \theta^{-2}E_\eta h(\theta Y) \le \theta^{-2}E_\eta g(\theta Y) \le \theta^{-2}
$$

where $h(v) \equiv v^2e^{-v}/(1-e^{-v}) \le 1 - e^{-v} \equiv g(v)$; to see this last inequality note it holds if and only if

$$
v^2e^{-v} \le (1-e^{-v})^2 = 1 - 2e^{-v} + e^{-2v},
$$

or, if and only if

$$
(2+v^2)e^{-v} \le 1 + e^{-2v} \quad \text{or, if and only if} \quad 2 + v^2 \le e^v + e^{-v},
$$

or, if and only if

$$1 + \frac{v^2}{2} \le \frac{1}{2}(e^v + e^{-v});$$

and this last inequality is indeed true.

When $Y \sim$ Exponential(η) this becomes

$$I_{11}(\theta, \eta) = \theta^{-2} 2 \frac{\eta}{\theta} \zeta(3, 1 + \eta/\theta) = \theta^{-2} 2r\zeta(3, 1+r)$$

where $\zeta(s,a) = \sum_{k=0}^{\infty} (k+a)^{-s}$ is the generalized zeta function and (again) $r \equiv \eta/\theta$. Figure 4 shows $I_{X,Y}(\theta)/I_{T(X,Y)}(\theta)$ and $I_{X,Y}(\theta)/I_{S(X,Y)}(\theta)$ when $Y \sim$ Exponential(η) as a function of $r \equiv \eta/\theta$.

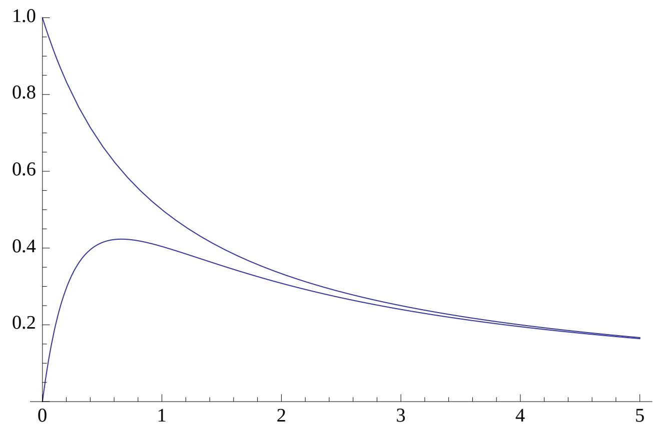

Figure 2: ARE's $I_{X,Y}(\theta)/I_{T(X,Y)}(\theta)$ and $I_{X,Y}(\theta)/I_{S(X,Y)}(\theta)$ as a function of r

3. 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:

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

Compute 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 $c > 0$. Compared this bound to the information bounds for estimation of θ based on observation of (X,Y) when η is known and unknown.

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

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

[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 is unknown. Thus reduction to W does not seem to be advisable. We can do better by basing estimation on *both* X and Y!

4. 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. $\mathbf{l}_1^* = \dot{\mathbf{l}}_1 - I_{12} I_{22}^{-1} \dot{\mathbf{l}}_2$ is orthogonal to $[\dot{\mathbf{l}}_2] \equiv \{a'\dot{\mathbf{l}}_2 : a \in R^{k-1}\}$ in $L_2(P_\theta)$.
B. $I_{11\cdot 2} = \inf_{c\in R^{k-1}} E_\theta(\dot{\mathbf{l}}_1 - c'\dot{\mathbf{l}}_2)^2$ and that the infimum is achieved when $c' = I_{12} I_{22}^{-1}$. Thus

$$I_{11\cdot 2} = E_\theta(\dot{\mathbf{l}}_1 - I_{12} I_{22}^{-1} \dot{\mathbf{l}}_2)^2 = E_\theta[(\mathbf{l}_\theta^*)^2]\,.$$

C. Prove the formulas (15) and (16) on page 21 of the Chapter 3 notes and interpret these formulas 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 (16) says that

$$\tilde{l}_1 = I_{11}^{-1} \dot{l}_1 - I_{11}^{-1} I_{12} \tilde{l}_2 \,. \tag{0.4}$$

One way to derive this is as indicated on page 21: 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 (0.4). 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 $[\dot{l}_1]$. Formula (17) follows immediately from (16) in view of orthogonality of the two terms:

$$\begin{aligned} I_{11 \cdot 2}^{-1} &= E[\tilde{l}_1 \tilde{l}_1^T] = E[I_{11}^{-1} \dot{\mathbf{l}}_1 \dot{\mathbf{l}}_1^T I_{11}^{-1}] + I_{11}^{-1} I_{12} E[\tilde{\mathbf{l}}_2 \tilde{\mathbf{l}}_2^T] I_{21} I_{11}^{-1} \\ &= I_{11}^{-1} + I_{11}^{-1} I_{12} I_{22 \cdot 1}^{-1} I_{21} I_{11}^{-1} . \end{aligned}$$