Statistics 581, Final Exam Solutions

Wellner; 12/13/2017

1. (40 points) **Define** *any four* of the following eight terms. In each case, provide an appropriate (brief) context for your definition.
   (a) The *Kullback - Leibler* divergence (or information) between a probability measure P and another (sub-)probability measure Q on the same measurable space $(\mathcal{X}, \mathcal{A})$.
   (b) An $n^{1/4}-$consistent preliminary estimator of θ in a parametric model.
   (c) A one-step estimator of a (vector-)parameter θ in a regular parametric model.
   (d) A locally regular estimator of a parameter $\nu(P_\theta) = q(\theta)$ in a parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$.
   (e) An asymptotically linear estimator.
   (f) The vector of score functions for a sample of size one in a regular parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$.
   (g) The *information matrix* for a sample of size one in a regular parametric model $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ with $\Theta \subset \mathbb{R}^d$.
   (h) The efficient influence function $\tilde{\mathbf{l}}_\nu$ for a differentiable parameter $q(\theta) = \nu(P_\theta)$ in a regular parametric model $\mathcal{P}$.

   Solution: See Course Notes, Chapters 1-4.

2. (40 points) **State** *four* of the following five results, providing an appropriate (brief) context for your statements:
   (a) The asymptotic behavior of the likelihood ratio statistic $2 \log \lambda_n$ (assuming the MLE's $\hat{\theta}_n$ exist) for testing a simple null hypothesis $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ under a sequence of local alternatives $\theta_n = \theta_0 + tn^{-1/2}$ in a regular parametric model.
   (b) Two inequalities relating the total variation distance $d_{TV}(P, Q)$ to the Hellinger distance $H(P, Q)$.
   (c) A result relating $H^2(P^n, Q^n)$ to $H^2(P, Q)$ where P^n is the measure corresponding to the product density $\prod_{i=1}^n p(x_i)$ of $X_1, \ldots, X_n$ i.i.d. P (and similarly for Q^n).
   (d) LAN (Local Asymptotic Normality) of the local log-likelihood ratios for a (regular) parametric model which is differentiable in quadratic mean.
   (e) Hájek's convolution theorem.

   Solution: See Course Notes, Chapters 1-4.

Do either problem 3 or problem 4:

3. (40 points) Let $X_1, \ldots, X_n$ be i.i.d. $P_\theta = \text{Normal}(\theta, 1)$.
   (a) Give the Hodges superefficient estimator T_n of θ (with superefficiency at $\theta = 0$).
   (b) What is the limiting distribution of $\sqrt{n}(T_n - \theta)$ as a function of θ?
   (c) What is the limiting distribution of $\sqrt{n}(T_n - \theta_n)$ when sampling from $\theta = \theta_n$ when $\theta_n = cn^{-1/2}$?
   (d) Does the limit distribution in (c) depend on c? Is T_n a locally regular estimator of θ at $\theta = 0$?
   (e) What is the limit of $E_{\theta_n}\{[\sqrt{n}(T_n - \theta_n)]^2\}$ when $\theta_n = cn^{-1/2}$ as in (c)? For what values of c does the limiting risk of T_n exceed the (limiting) risk of $\overline{X}_n$?

   Solution: See Course Notes, Chapter 3.

4. (40 points)
   (a) **State** the Glivenko-Cantelli theorem. Then **prove** that it holds *if it holds* for the case of i.i.d. Uniform$(0,1)$ random variables.
   (b) **Prove** the Glivenko-Cantelli theorem for i.i.d. Uniform$(0,1)$ random variables: if $\xi_1, \ldots, \xi_n, \ldots$ are i.i.d. Uniform$(0,1)$ with empirical distribution function

$$\mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(\xi_i)\,, \qquad \text{then} \qquad \sup_{0 \le t \le 1} |\mathbb{G}_n(t) - t| \to_{a.s.} 0\,.$$

   Solution: See Course Notes, Chapter 2.

Do either problem 5 or problem 6:

5. (40 points)
Suppose that $\underline{X}, \underline{X}_1, \ldots, \underline{X}_n$ are i.i.d. $\text{Mult}_k(1, \underline{p})$, so that $\underline{N}_n \equiv \sum_{i=1}^n \underline{X}_i \sim \text{Mult}_k(n, \underline{p})$. Thus

$$\begin{aligned} P_{\underline{p}}(\underline{X} = \underline{x}) &= \prod_{j=1}^k p_j^{x_j} \quad \text{for } \ x_i \in \{0,1\}, \ \sum_1^k x_i = 1\,, \\ P_{\underline{p},n}(\underline{N}_n = \underline{m}) &= \frac{n!}{\prod_{j=1}^k m_j!} \prod_{j=1}^k p_j^{m_j} \quad \text{for } \ m_i \geq 0, \text{ integers } \sum_{j=1}^k m_j = n\,. \end{aligned}$$

(a) Compute $K(P_{\underline{q}}, P_{\underline{p}}) \equiv K(\underline{q}, \underline{p})$ for vectors $\underline{q}$, $\underline{p}$ with $\sum p_j = \sum q_j = 1$.
(b) Evaluate $K(\widehat{\underline{p}}, \underline{p})$ where $\widehat{\underline{p}} = n^{-1}\underline{N}_n$. Relate this to the log-likelihood $\log L_n(\underline{p}|\underline{N}_n)$.
(c) Use the result of (b) to show, without using any calculus, that the MLE of $\underline{p}$ is $\widehat{\underline{p}} = \underline{N}/n$.

Solution: (a) First,

$$\log \frac{p_{\underline{q}}(\underline{x})}{p_{\underline{p}}(\underline{x})} = \log \prod_{j=1}^k \frac{q_j^{x_j}}{p_j^{x_j}} = \sum_{j=1}^k x_j \log(\frac{q_j}{p_j})\,.$$

Thus

$$K(\underline{q}, \underline{p}) = \sum_{j=1}^k q_j \log \frac{q_j}{p_j}\,.$$

(b) From A it follows that

$$K(\widehat{\underline{p}}, \underline{p}) = \sum_{j=1}^k \widehat{p}_j \log \frac{\widehat{p}_j}{p_j} = -\sum_{j=1}^k \widehat{p}_j \log \frac{p_j}{\widehat{p}_j}\,.$$

Now

$$\begin{aligned} \log L_n(\underline{p}|\underline{N}_n) &= \sum_{j=1}^k N_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right) \\ &= n\sum_{j=1}^k \widehat{p}_j \log p_j + \log\left(\frac{n!}{\prod N_j!}\right) \\ &= n\sum_{j=1}^k \widehat{p}_j \log\left(\frac{p_j}{\widehat{p}_j}\right) + n\sum_{j=1}^k \widehat{p}_j \log \widehat{p}_j + \log\left(\frac{n!}{\prod N_j!}\right) \\ &= -nK(\widehat{\underline{p}}, \underline{p}) + \text{terms constant in } \underline{p}\,. \end{aligned}$$

Even more neatly, as several of you noted,

$$\log \frac{L_n(\widehat{\underline{p}}|\underline{N}_n)}{L_n(\underline{p}|\underline{N}_n)} = n\sum_{j=1}^k \{\widehat{p}_j \log \widehat{p}_j - \widehat{p}_j \log p_j\} = nK(\widehat{\underline{p}}, \underline{p})\,.$$

(c) Since $K(\underline{\widehat{p}}, \underline{p}) \ge 0$ with equality if and only if $\underline{p} = \underline{\widehat{p}}$, we see from the identity in B that $L_n(\underline{p}|\underline{N}_n)$ is maximized by $\underline{p} = \underline{\widehat{p}}$.

6. (40 points).
Suppose that $P = P_0 = N(0,1)$, $Q = P_\theta = N(\theta, 1)$ on $(\mathbb{X}, \mathcal{A}) = (\mathbb{R}, \mathcal{B})$.
(a) Compute $K(P,Q) = K(P_0, P_\theta)$.
(b) Compute $H^2(P,Q) = 1 - \rho(P,Q)$ and $\rho(P,Q) = \int \sqrt{p(x)q(x)}dx$. [It might be easiest to compute $\rho(P,Q)$ first recalling that if $Z \sim N(0,1)$ then $E\exp(tZ) = \exp(t^2/2)$.]
(c) Compute $d_{TV}(P,Q) = 1 - \eta(P,Q)$ and $\eta(P,Q) = \int p(x) \wedge q(x) dx$. [It might be easiest to compute $\eta(P,Q)$ first.]
(d) Show in general that $K(P,Q) \ge 2H^2(P,Q)$, thereby strengthening the fact $K(P,Q) \ge 0$ that we proved in class. [Hint: write both $K(P,Q)$ and $H^2(P,Q)$ in terms of $Y = (p(X)/q(X))^{1/2}$ and use the inequality $\log(1+x) \ge x/(1+x)$ for $x \ge 0$. You will need to relate $E_Q Y$ and $E_Q Y^2$ to $H^2(P,Q)$.]
(e) Use the results of (a) and (d) to find a lower bound for $K(P^n, Q^n)$ in terms of $H^2(P,Q)$ or $\rho(P,Q)$; here P^n and Q^n are the probability distributions of $X_1, \ldots, X_n$ i.i.d. as P and Q respectively.

Solution: (a) Now $p(x) = \phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and $q(x) = (2\pi)^{-1/2}\exp(-(x-\theta)^2/2)$, so

$$\frac{p(x)}{q(x)} = \exp(-x^2/2 + (x-\theta)^2/2) = \exp(-\theta x + \theta^2/2),$$
$$\log\frac{p}{q}(x) = -\theta x + \theta^2/2,$$

and it follows that

$$K(P,Q) = E_P\left(\log\frac{p}{q}\right) = -\theta E_P(X) + \theta^2/2 = \theta^2/2.$$

(b) We compute $\rho(P,Q)$ first:

$$\begin{aligned}\rho(P,Q) &= \int \sqrt{p(x)q(x)}dx = \int \frac{1}{\sqrt{2\pi}}\exp(-x^2/4)\exp(-(x-\theta)^2/4)dx \\ &= \int \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\exp((2\theta x - \theta^2)/4)\,dx \\ &= \exp(-\theta^2/4)E_P\exp(\theta X/2) = \exp(-\theta^2/4)\exp(\theta^2/8) = \exp(-\theta^2/8).\end{aligned}$$

Hence

$$H^2(P,Q) = 1 - \rho(P,Q) = 1 - \exp(-\theta^2/8).$$

(c) We compute $\eta(P,Q)$ first. Since $\phi(x) \ge \phi(x-\theta)$ if and only if

$$\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \ge \frac{1}{\sqrt{2\pi}}\exp(-(x-\theta)^2/2)$$

or equivalently, if and only if

$$1 \geq \exp(\theta x - \theta^2/2) \qquad \text{iff} \qquad \theta x - \theta^2/2 \leq 0$$

and, for $\theta > 0$, this holds if and only if $x \leq \theta/2$. Hence it follows that

$$\begin{aligned}
\int p(x) \wedge q(x) dx &= \int \phi(x) \wedge \phi(x-\theta) dx \\
&= \int_{-\infty}^{\theta/2} \phi(x-\theta) dx + \int_{\theta/2}^{\infty} \phi(x) dx \\
&= \int_{-\infty}^{-\theta/2} \phi(y) dy + 1 - \Phi(\theta/2) \\
&= \Phi(-\theta/2) + 1 - \Phi(\theta/2)\,, \qquad \text{if } \ \theta > 0 \\
&= 2\Phi(-|\theta|/2), \quad \text{if } \theta > 0.
\end{aligned}$$

When $\theta < 0$, $\theta x - \theta^2/2 < 0$ if and only if $x \geq \theta/2$, and this yields

$$\begin{aligned}
\int p(x) \wedge q(x) dx &= \int_{\theta/2}^{\infty} \phi(x-\theta) dx + \int_{-\infty}^{\theta/2} \phi(x) dx \\
&= \Phi(\theta/2) + 1 - \Phi(-\theta/2) \\
&= 2\Phi(-|\theta|/2), \quad \theta < 0.
\end{aligned}$$

It follows that $\eta(P, Q) = 2\Phi(-|\theta|/2)$ for all θ, and hence

$$d_{TV}(P, Q) = 1 - \eta(P, Q) = 1 - 2\Phi(-|\theta|/2).$$

The following Figure shows $K(P_0, P_\theta)$, $H^2(P_0, P_\theta)$, and $d_{TV}(P_0, P_\theta)$ as functions of θ.

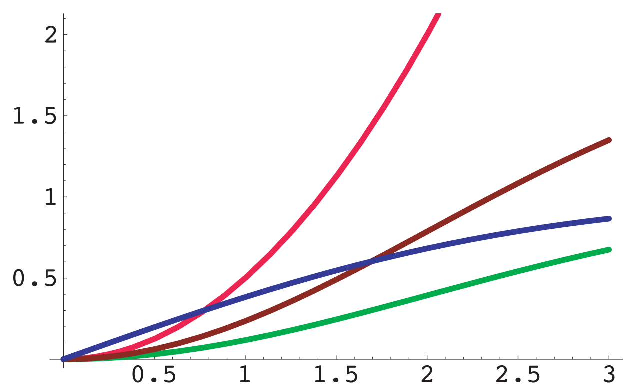

Figure 1: $K(P_0, P_\theta)$ (red), $H^2(P_0, P_\theta)$ (green), $d_{TV}(P_0, P_\theta)$ (blue), and $2H^2(P_0, P_\theta)$ (burgundy) as functions of θ

(d) Let $Y \equiv \sqrt{p/q}$. Then

$$K(P, Q) = 2 \int p \log(\sqrt{p/q}) d\mu = 2E_P \log Y = 2E_P \log(1 + (Y - 1))$$

$$
\begin{aligned}
&\geq 2E_P \frac{Y-1}{1+Y-1} \qquad \text{using } \log(1+x) \geq \frac{x}{1+x}, \\
&= 2E_P \frac{Y-1}{Y} = 2\int p\left\{\sqrt{\frac{p}{q}} - 1\right\}\sqrt{\frac{q}{p}}d\mu \\
&= 2\int p\left\{1 - \sqrt{\frac{q}{p}}\right\}d\mu = 2\left\{1 - \int \sqrt{pq}d\mu\right\} \\
&= 2H^2(P,Q).
\end{aligned}
$$

Here is a another way of organizing the argument with a slightly different choice of Y, as follows. Let $Y \equiv \sqrt{p/q} - 1$. Then $p/q = (1+Y)^2$ and it follows that

$$
\begin{aligned}
K(P,Q) &= \int p\log(p/q)d\mu = \int (p/q)\log(p/q)qd\mu \\
&= 2\int (p/q)\log(p/q)^{1/2}dQ = 2\int (1+Y)^2\log(1+Y)dQ \\
&\geq 2\int (1+Y)^2\frac{Y}{1+Y}dQ \\
&= 2\int Y(1+Y)dQ = 2\left\{\int YdQ + \int Y^2dQ\right\}.
\end{aligned}
$$

But we also have

$$
\begin{aligned}
&\int YdQ = \int \sqrt{pq}d\mu - 1 = -H^2(P,Q), \\
&\int Y^2dQ = \int [\sqrt{p/q} - 1]^2 qd\mu = \int [\sqrt{p} - \sqrt{q}]^2 d\mu = 2H^2(P,Q).
\end{aligned}
$$

By combining the results of these last two displays it follows that

$$K(P,Q) \geq 2\left\{2H^2(P,Q) - H^2(P,Q)\right\} = 2H^2(P,Q).$$

(e) By (d),

$$K(P^n,Q^n) \geq 2H^2(P^n,Q^n) = 2\{1 - \rho(P^n,Q^n)\} = 2\{1 - \rho(P,Q)^n\}$$

in general. For the particular case we began with $\rho(P,Q) = \exp(-\theta^2/8)$, and thus we conclude that in this case

$$K(P^n,Q^n) \geq 2\{1 - \exp(-n\theta^2/8)\}.$$

Another lower bound follows from the identity $K(P^n,Q^n) = nK(P,Q)$: we conclude from this together with the inequality from (d) that

$$K(P^n,Q^n) = nK(P,Q) \geq n2H^2(P,Q) = 2n(1 - \exp(-\theta^2/8)).$$

Do **either** Problem 7 **or** Problem 8.

7. (49 points) Suppose that $\{P_\theta : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^d$ is a regular parametric model satisfying the hypotheses A0-A4 of Theorem 4.2.1. Let θ_0 be an interior point of θ and suppose that the hypotheses A0-A4 of A0-A4 hold at θ_0. Let $\theta \in \Theta$ be some other fixed point at which A0-A4 hold. Consider the parameter $q(\theta) = K(P_{\theta_0}, P_\theta)$.
   (a) Express $q(\theta)$ in terms of the densities p_θ and p_{θ_0}.
   (b) Compute the derivatives $(\partial/\partial\theta_i)q(\theta)$ and the gradient vector $\nabla q(\theta)$ for a general θ and for θ_0. (You may assume that interchange of differentiation and integration is allowed.)
   (c) Compute $(\partial/\partial\theta_j)(\partial/\partial\theta_i)q(\theta)$ and thereby the matrix of second derivatives $\ddot{q}(\theta)$ for a general θ and for θ_0.
   (d) Now suppose that $\tilde{\theta}_n$ is a consistent estimator of θ when we observe $X_1, \ldots, X_n$ i.i.d. P_θ with density p_θ. What is the limit in probability of $q(\tilde{\theta}_n)$ $= K(P_{\theta_0}, P_\theta)|_{\theta=\tilde{\theta}_n}$?
   (e) If $\theta = \theta_0$ is true and $\sqrt{n}(\tilde{\theta}_n - \theta_0) \rightarrow_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$ holds, what is the limiting distribution of $\sqrt{n}(q(\tilde{\theta}_n) - q(\theta_0))$?
   (f) If $\theta = \theta_0$ is true and $\sqrt{n}(\tilde{\theta}_n - \theta_0) \rightarrow_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$ holds, what is the limiting distribution of $2n(q(\tilde{\theta}_n) - q(\theta_0))$?
   (g) If $\theta = \theta_0 + tn^{-1/2}$ is true and $\sqrt{n}(\tilde{\theta}_n - \theta_n) \rightarrow_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$, holds, what is the limiting distribution of $2n(q(\tilde{\theta}_n) - q(\theta_0))$?

Solution: (a) Since the probability measures P_{θ_0} and P_θ have densities p_{θ_0} and p_θ with respect to a dominating measure μ it follows that

$$q(\theta) = K(P_{\theta_0}, P_\theta) = \int p_{\theta_0} \log(p_\theta/p_{\theta_0}) d\mu = \int p_{\theta_0} \log p_{\theta_0} d\mu - \int p_{\theta_0} \log p_\theta d\mu.$$

(b) Since we can interchange differentiation and integration, we find that

$$\dot{q}(\theta) = -\int p_{\theta_0} \dot{\mathbf{l}}_\theta d\mu = -E_{\theta_0} \dot{\mathbf{l}}_\theta(X;\theta),$$

and hence $\dot{q}(\theta_0) = -E_{\theta_0} \dot{\mathbf{l}}_\theta(X;\theta_0) = 0$. Note that this makes sense since $q(\theta) = K(P_{\theta_0}, P_\theta)$ is minimized by $\theta = \theta_0$.
(c) By taking another derivative in (b) we find that

$$\ddot{q}(\theta) = -\int p_{\theta_0} \ddot{\mathbf{l}}_{\theta,\theta}(x;\theta) d\mu,$$

and hence $\ddot{q}(\theta_0) = -E_{\theta_0} \ddot{\mathbf{l}}_{\theta,\theta}(X;\theta_0) = I(\theta_0)$. Thus $\ddot{q}(\theta_0) = I(\theta_0)$ is the Hessian of $q(\theta)$ at θ_0.
(d) Since $q(\theta) = K(P_{\theta_0}, P_\theta)$ is continuous under A0-A4 and $\tilde{\theta}_n$ is consistent, $q(\tilde{\theta}_n) \rightarrow_p q(\theta) = K(P_{\theta_0}, P_\theta)$.
(e) When θ_0 is true and $\sqrt{n}(\tilde{\theta}_n - \theta_0) \rightarrow_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$, the delta-method (or g'-theorem yields $\sqrt{n}(q(\tilde{\theta}_n) - q(\theta_0)) \rightarrow_d 0 \cdot \underline{D} = 0$ in view of (b).
(f) When θ_0 is true and $\sqrt{n}(\tilde{\theta}_n - \theta_0) \rightarrow_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$, then $q(\theta_0) = 0$ and it follows from (b) and (c) that

$$2n(q(\tilde{\theta}_n) - q(\theta_0)) \rightarrow \underline{D}^T I(\theta_0) \underline{D} \sim \chi^2_d.$$

(g) When $\theta_n = \theta_0 + n^{-1/2}t$ is true and $\sqrt{n}(\tilde{\theta}_n - \theta_n) \rightarrow_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$, then $q(\theta_0) = 0$ and it follows from (b) and (c) that

$$2n(q(\tilde{\theta}_n) - q(\theta_0)) \rightarrow_d (\underline{D} + \underline{t})^T I(\theta_0)(\underline{D} + \underline{t}) \sim \chi^2_d(\delta)$$

where $\delta = t^T I(\theta_0)t$.

It is interesting to note that similar considerations apply to $\tilde{q}(\theta) \equiv K(P_\theta, P_{\theta_0}) \neq K(P_{\theta_0}, P_\theta)$; even though K is not symmetric in its arguments, it is "locally quadratic" in either argument.

8. (49 points) Suppose that X has the Weibull(α, β) density $p_\theta(x)$ studied in Example 3.2.5 of Chapter 3 of the course notes:

$$p_\theta(x) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right) 1_{[0,\infty)}(x)$$

with respect to Lebesgue measure where $\theta = (\alpha, \beta) \equiv (\theta_1, \theta_2) \in (0,\infty) \times (0,\infty) \subset R^2$. For the Weibull family $\mathcal{P}$, $\log p_\theta(x)$ is differentiable at every $\theta \in \Theta$ and the scores are:

$$\begin{aligned} \dot{\mathbf{l}}_\alpha(x) &= \frac{\beta}{\alpha}\left\{\left(\frac{x}{\alpha}\right)^{\beta} - 1\right\} \\ \dot{\mathbf{l}}_\beta(x) &= \frac{1}{\beta} - \frac{1}{\beta}\log\left\{\left(\frac{x}{\alpha}\right)^{\beta}\right\}\left\{\left(\frac{x}{\alpha}\right)^{\beta} - 1\right\}. \end{aligned}$$

Thus $\dot{\mathcal{P}} \equiv [\dot{\mathbf{l}}_\theta]$ is the two-dimensional subspace of $L_2(P_\theta)$ spanned by $\dot{\mathbf{l}}_\alpha$ and $\dot{\mathbf{l}}_\beta$, and the Fisher information matrix is

$$I(\theta) = E\{\dot{\mathbf{l}}_\theta(X)\dot{\mathbf{l}}_\theta^T(X)\} = \begin{pmatrix} \beta^2/\alpha^2 & a/\alpha \\ a/\alpha & b^2/\beta^2 \end{pmatrix}$$

where

$$a = -E\{(Y-1)^2\log(Y)\} = -(1-\gamma), \quad b^2 = E\{[(Y-1)\log(Y)-1]^2\} = \pi^2/6+(1-\gamma)^2.$$

and the computation of $I(\theta)$ was simplified by noting that $Y \equiv (X/\alpha)^\beta \sim$ Exponential(1).

(a) Show that $\mathbf{l}_2^*(X) = \dot{\mathbf{l}}_2 - I_{21}I_{11}^{-1}\dot{\mathbf{l}}_1$, the efficient score function for $\beta = \theta_2$, is orthogonal to $[\dot{\mathbf{l}}_1] = \{c\dot{\mathbf{l}}_1 : c \in \mathbb{R}\}$ in $L_2(P_\theta)$.

(b) Compute $E_\theta \mathbf{l}_2^{*2}(X)$ and relate it to some element in $I(\theta)^{-1}$.

(c) What is the efficient influence function $\tilde{\mathbf{l}}_2$ for estimation of β?

(d) If ψ is the influence function of some general asymptotically linear estimator of β (e.g. based on the method of moments or quantiles or ...), what is the relationship between ψ and the efficient influence function $\tilde{\mathbf{l}}_2$ in (c)?

(e) What is the Rao statistic R_n for testing $H : \beta = 1$ versus $K : \beta \neq 1$?

(f) What is its limiting distribution under H? What is its limiting distribution under $\theta_n = (\alpha + sn^{-1/2}, 1 + tn^{-1/2})$?

Solution: (a) To see that $\mathbf{l}_2^*(X) = \dot{\mathbf{l}}_2 - I_{21}I_{11}^{-1}\dot{\mathbf{l}}_1$, the efficient score function for $\beta = \theta_2$, is orthogonal to $[\dot{\mathbf{l}}_1] = \{c\dot{\mathbf{l}}_1 : c \in \mathbb{R}\}$, we simply compute

$$\begin{aligned} E_\theta\{\mathbf{l}_2^*(X)\dot{\mathbf{l}}_1\} &= E_\theta\{(\dot{\mathbf{l}}_2 - I_{21}I_{11}^{-1}\dot{\mathbf{l}}_1)\dot{\mathbf{l}}_1\} \\ &= E_\theta\{\dot{\mathbf{l}}_2\dot{\mathbf{l}}_1\} - I_{21}I_{11}^{-1}E_\theta\{\dot{\mathbf{l}}_1^2\} \\ &= I_{21} - I_{21}I_{11}^{-1}I_{11} = I_{21} - I_{21} = 0. \end{aligned}$$

(b) It follows from (a) that

$$\begin{aligned} E_\theta\{\dot{\mathbf{l}}_2^{*2}(X)\} &= E_\theta\{(\dot{\mathbf{l}}_2 - I_{21}I_{11}^{-1}\dot{\mathbf{l}}_1)(\dot{\mathbf{l}}_2 - I_{21}I_{11}^{-1}\dot{\mathbf{l}}_1)\} \\ &= E_\theta\{(\dot{\mathbf{l}}_2 - I_{21}I_{11}^{-1}\dot{\mathbf{l}}_1)\dot{\mathbf{l}}_2\} \text{ by the orthogonality proved in (a)} \\ &= E_\theta\{(\dot{\mathbf{l}}_2^2\} - I_{21}I_{11}^{-1}E_\theta\{\dot{\mathbf{l}}_1\dot{\mathbf{l}}_2\} \\ &= I_{22} - I_{21}I_{11}^{-1}I_{12} \\ &= I_{22\cdot 1}. \end{aligned}$$

But $I_{22\cdot 1}^{-1}(\theta) = I^{22}(\theta)$, the lower right entry in the inverse of the information matrix.
(c) The efficient influence function for estimation of $\beta = \theta_2$ is $\tilde{\mathbf{l}}_2(x) = I_{22\cdot 1}^{-1}\dot{\mathbf{l}}_2^*(x)$, and in the particular Weibull case we have

$$I_{22\cdot 1} = \frac{b^2}{\beta^2} - \frac{a^2}{\alpha^2}\frac{\alpha^2}{\beta^2} = \frac{1}{\beta^2}(b^2 - a^2) = \frac{\pi^2/6}{\beta^2}$$

so that the information bound for estimation of β when α is unknown is given by $I_{22\cdot 1}^{-1} = (6/\pi^2)\beta^2$, and the efficient influence function for estimation of β is

$$\tilde{\mathbf{l}}_2(x) = \tilde{\mathbf{l}}_\beta(x) = (6/\pi^2)\beta^2 \left\{ \dot{\mathbf{l}}_\beta(x) - I_{\beta\alpha} I_{\alpha\alpha}^{-1} \dot{\mathbf{l}}_\alpha(x) \right\}.$$

(d) If ψ is the influence function of some (generally inefficient) asymptotically linear estimator T_n of β, then $\tilde{\mathbf{l}}_\beta$ is the projection of ψ onto the tangent space $\dot{\mathcal{P}}$ of the model which is given by the linear span of the two score functions $\dot{\mathbf{l}}_1 = \dot{\mathbf{l}}_\alpha$ and $\dot{\mathbf{l}}_1 = \dot{\mathbf{l}}_\alpha$. This means that $\psi - \tilde{\mathbf{l}}_\beta$ is orthogonal to $\underline{a}^T \dot{\mathbf{l}}_\theta$ in $L_2(P_\theta)$: that is

$$E_\theta\{(\psi - \tilde{\mathbf{l}}_\beta)\dot{\mathbf{l}}_\theta\} = 0,$$

and the asymptotic variance of T_n is given by

$$\begin{aligned} E_\theta \psi^2(X) &= E_\theta\{(\psi - \tilde{\mathbf{l}}_\beta)^2\} + E_\theta\{(\tilde{\mathbf{l}}_\beta)^2\} \\ &= E_\theta\{(\psi - \tilde{\mathbf{l}}_\beta)^2\} + I_{22\cdot 1}^{-1} \\ &\geq I_{22\cdot 1}^{-1}. \end{aligned}$$

(e) The Rao statistic for testing $H : \beta = 1$ versus $K : \beta \neq 1$ is given by

$$R_n = Z_n^T(\hat{\theta}_n^0)\hat{I}_n^{-1}(\hat{\theta}_n^0)Z_n(\hat{\theta}_n^0)$$

where $\underline{Z}_n(\theta) = n^{-1/2}\sum_{i=1}^n \underline{\dot{\mathbf{l}}}_\theta(X_i|\theta)$ and where $\hat{\theta}_n^0$ is the MLE of α in the smaller model specified by the null hypothesis $\beta = 1$. But when $\beta = 1$, the maximum likelihood estimator of α for the Weibull model is just $\hat{\alpha}_n^0 = \overline{X}_n$, the sample mean and the resulting estimator of $\theta^0 \in \Theta^0$ is $\hat{\theta}_n^0 = (\overline{X}_n, 1)$, and we have

$$\underline{Z}_n(\hat{\theta}_n^0) = (0, Z_{n,2}(\hat{\theta}_n^0))^T.$$

(f) Now under P_{θ^0} with $\theta^0 = (\alpha, 1)$ for some $\alpha > 0$ we have

$$\begin{aligned} R_n &= Z_{n,2}^2(\hat{\theta}_n^0)\hat{I}_{22\cdot 1}^{-1}(\hat{\theta}_n^0) \\ &= \left(Z_{n,2}(\theta^0) - I_{21}(\theta^0)I_{22}^{-1}(\theta^0) + o_p(1)\right)^2 (I_{22\cdot 1}^{-1}(\theta^0) + o_p(1)) \\ &\to_d (Z_2 - I_{21}I_{11}^{-1}Z_1)^2 I_{22\cdot 1}^{-1} \sim \chi_1^2. \end{aligned}$$

Similarly, if $\theta_n = (\alpha_s n^{-1/2}, 1 + tn^{-1/2})$, then under P_{θ_n} we have

$$R_n \to_d (Z_2 - I_{21}I_{11}^{-1}Z_1 + I_{22\cdot 1}t)^2 I_{22\cdot 1}^{-1} \sim \chi_1^2(\delta)$$

where $\delta = t^2 I_{22\cdot 1} = t^2\pi^2/6$.