Statistics 581, Final Exam Solutions
Wellner; 12/09/2014

1. (48) points) **Define** *each* of the following terms.
   (a) An asymptotically linear estimator T_n of a parameter $\nu(P)$ with influence function ψ.
   (b) A locally regular estimator of a parameter $\nu(P_\theta)$.
   (c) The Hellinger distance between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.
   (d) The total variation distance $d_{TV}(P, Q)$ between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.
   (e) The likelihood ratio test statistic for testing $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ in a regular parametric model.

2. (32 points) **State** *four* of the following six results, providing the appropriate (brief) context for your statement:
   (a) A result giving two ways of computing the Fisher information matrix $I(\theta)$ if a regularity condition holds.
   (b) A theorem about the limiting distribution of the likelihood ratio statistic $2 \log \lambda_n$ for testing $H : \theta = \theta_0$ versus $K : \theta \neq \theta_0$ in a regular parametric model when the true $\theta_n = \theta_0 + tn^{-1/2}$.
   (c) A result relating the Hellinger affinity $\rho(P^n, Q^n)$ to $\rho(P, Q)$ when P^n is the measure corresponding to the product density $\prod_{i=1}^n (x_i)$ of $X_1, \ldots, X_n$ i.i.d. P (and similarly for Q^n).
   (d) Wald's theorem concerning consistency of maximum likelihood estimators.
   (e) LAN (Local Asymptotic Normality) of the local likelihood ratios for a regular parametric model satisfying the Cramér hypotheses.
   (f) The Lindeberg-Feller central limit theorem.
   (g) The basic inverse transformation theorem.

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\theta, \theta)$ where $\theta \in (0, \infty) \equiv \Theta$.
   (a) Find the score for a sample of size $n = 1$ and compute the information matrix.
   (b) Find the MLE $\hat{\theta}_n$ of θ.
   (c) Show that the sequence of MLE's $\{\hat{\theta}_n\}$ is consistent.
   (d) Show that the sequence of MLE's is asymptotically normal and find its asymptotic variance.
   (e) Suggest two alternative inefficient estimators of θ based on the usual $N(\mu, \sigma^2)$ model and compare their asymptotic variances to the variance of the MLE you computed in (d).

   Solution: (a) The density for one observation is

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{(x-\theta)^2}{2\theta} \right),$$

   and hence

$$\log p_\theta(x) = -\frac{1}{2} \log \theta - \frac{(x-\theta)^2}{2\theta}$$

$$= -\frac{1}{2}\log\theta - \frac{1}{2}\left(\frac{x^2}{\theta} - 2x + \theta\right).$$

It follows that the score for θ is given by

$$\dot{\mathbf{l}}_\theta(x) = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2} - \frac{1}{2}.$$

To find the information for θ we calculate

$$\ddot{\mathbf{l}}_{\theta,\theta}(x) = \frac{1}{2\theta^2} - \frac{x^2}{\theta^3},$$

and this yields

$$\begin{aligned} I(\theta) &= -E_\theta \ddot{\mathbf{l}}_{\theta,\theta}(X) = \frac{E_\theta(X^2)}{\theta^3} - \frac{1}{2\theta^2} \\ &= \frac{\theta+\theta^2}{\theta^3} - \frac{1}{2\theta^2} = \frac{2\theta+1}{2\theta^2}. \end{aligned}$$

(b) The likelihood equation can be written as

$$0 = \dot{\mathbf{l}}_\theta(\theta|\underline{X}) = \frac{1}{2\theta^2}\sum_{i=1}^n X_i^2 - \frac{n}{2\theta} - \frac{n}{2},$$

and hence the MLE $\hat{\theta}_n$ is given by the solution of $\hat{\theta}_n^2 + \hat{\theta}_n = n^{-1}\sum_{i=1}^n X_i^2$, or equivalently

$$(\hat{\theta}_n + 1/2)^2 = \overline{X^2} + 1/4,$$

and this yields $\hat{\theta}_n = \left(1/4 + \overline{X^2}\right)^{1/2} - 1/2$.

(c) Since $\overline{X^2}_n \rightarrow_p E_\theta(X^2) = \theta + \theta^2$, and $g(v) = \sqrt{1/4 + v} - 1/2$ is continuous, it follows by the Mann-Wald theorem that

$$\begin{aligned} \hat{\theta}_n &= \left(1/4 + \overline{X^2}\right)^{1/2} - 1/2 \equiv g(\overline{X^2}_n) \\ &\rightarrow_p g(\theta+\theta^2) = \theta. \end{aligned}$$

(d) From our general theory and the information calculation in (a) it follows that

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow_d D \sim N(0, I(\theta)^{-1}) = N(0, 2\theta^2/(2\theta+1)).$$

This result agrees with the application of the delta-method and the Lindeberg CLT applied to $\sqrt{n}(\overline{X^2} - E_\theta X^2)$ after a somewhat tedious calculation: $E_\theta(X^4) = 2\theta^2(1 + 2\theta)$.

(e) One alternative estimator is $\overline{X}_n$. In this case we know that $\sqrt{n}(\overline{X}_n - \theta) \rightarrow_d N(0,\theta)$ by the ordinary CLT. A second alternative estimator is $S_n^2 = n^{-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ in this case we know that

$$\sqrt{n}(S_n^2 - \theta) = \sqrt{n}\left(\overline{(X-\theta)^2} - \theta\right) + o_p(1) \rightarrow_d N(0, Var_\theta[(X-\theta)^2])$$

where $Var_\theta[(X-\theta)^2] = E_\theta(X-\theta)^4 - \{E_\theta(X-\theta)^2\}^2 = \theta^2 \cdot 3 - \theta^2 = 2\theta^2$. Note that $\theta > 2\theta^2/(2\theta+1)$ and $2\theta^2 > 2\theta^2/(2\theta+1)$ for all $\theta > 0$, and the asymptotic relative efficiencies of $\overline{X}_n$ and S_n^2 relative to the MLE are given by

$$\text{Rel-Eff}(\overline{X}_n, MLE)(\theta) = \frac{2\theta^2/(2\theta+1}{\theta} = \frac{2\theta}{2\theta+1},$$

and

$$\text{Rel-Eff}(S_n^2, MLE)(\theta) = \frac{2\theta^2/(2\theta+1)}{2\theta^2} = \frac{1}{2\theta+1}.$$

Thus $\overline{X}_n$ performs well relative to the MLE when θ is large, while S_n^2 performs well relative to the MLE when θ is small, with $\theta = 1/2$ being the break-point.

4. Suppose that $X_1, \ldots, X_n$ are i.i.d. Uniform$(0,\theta)$. Let $X_{(1)} = \min_{1\le i\le n} X_i$ and $X_{(n)} = \max_{1\le i\le n} X_i$.
   (a) Show that $(nX_{(1)}, n(\theta - X_{(n)})) \to_d \theta(U,V)$ where U and V are independent exponential(1) random variables. Hint: begin by computing $P(X_{(1)} > x,\ X_{(n)} \le y)$; then use this to study the limit of $P(nX_{(1)}/\theta > x,\ n(1 - X_{(n)}/\theta) \ge y)$.
   (b) Show that $S_n \equiv (n+1)X_{(n)}/n$ and $T_n \equiv X_{(1)} + X_{(n)}$ are both unbiased estimators of θ.
   (c) Find the joint limiting distribution of $(n(S_n - \theta), n(T_n - \theta))$.
   (d) Which of the two estimators would you prefer? [Hint: compute $\lim_n E\{[n(S_n-\theta)]^2\}$ and $\lim_n E\{[n(T_n-\theta)]^2\}$.]

Solution: (a) First note that $X_1/\theta, \ldots, X_n/\theta$ are i.i.d. Uniform$(0,1)$. Thus we compute, for $0 \le x \le y \le 1$

$$\begin{aligned} P(X_{(1)} > x, X_{(n)} \le y) &= P(X_{(1)}/\theta > x/\theta,\ X_{(n)}/\theta \le y/\theta) \\ &= P(x/\theta < X_i/\theta \le y/\theta \text{ for all } 1 \le i \le n) \\ &= P(x/\theta < X_1/\theta \le y/\theta)^n = (y-x)^n/\theta^n. \end{aligned}$$

This implies that for all $x, y > 0$ we have, for n so large that $(x+y)/n \le 1$,

$$\begin{aligned} P\left(nX_{(1)}/\theta > x,\ n(1 - X_{(n)}/\theta) \ge y\right) &= P\left(X_{(1)} > x\theta/n,\ X_{(n)} \le 1 - y\theta/n\right) \\ &= \left(1 - \frac{y}{n} - \frac{x}{n}\right)^n \to \exp(-(x+y)) \\ &= \exp(-x)\exp(-y). \end{aligned}$$

It follows that $(nX_{(1)}, n(\theta - X_{(n)})) \to_d \theta(U,V)$ where U,V are independent exponential(1) random variables.
(b) First, $E_\theta(X_{(1)}/\theta) = 1/(n+1)$ and $E_\theta(X_{(n)}/\theta) = n/(n+1)$. Thus it follows that

$$E_\theta(S_n) = E_\theta\left(\frac{n+1}{n} X_{(n)}\right) = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta,$$

and

$$E_\theta(T_n) = E_\theta\left(X_{(n)} + X_{(1)}\right) = \frac{n}{n+1}\theta + \frac{1}{n+1}\theta = \theta.$$

(c) We see from (a) that

$$\begin{aligned} n(S_n-\theta) &= \frac{n+1}{n} n(X_{(n)}-\theta) + \left(\frac{n+1}{n} n - n\right)\theta \\ &= n(X_{(n)}-\theta) + \theta + o_p(1) \\ &\to_d (-\theta V) + \theta = \theta(1-V), \end{aligned}$$

while

$$n(T_n-\theta) = n(X_{(n)}-\theta) + nX_{(1)} \to -\theta V + \theta U = \theta(U-V).$$

Since $n(T_n-\theta)$ and $n(S_n-\theta)$ are both linear transformations of $nX_{(1)}$ and $n(\theta - X_{(n)})$, the joint convergence follows from the continuous mapping theorem (or Mann-Wald theorem) with $g(u,v) = (-v+\theta, u-v)$.

(d) Now from (c) together with uniform square integrability of $\{n(S_n-\theta)\}$ and $\{n(T_n-\theta)\}$ we see that since $Var(U) = Var(V) = 1$

$$E_\theta\{n^2(S_n-\theta)^2\} \to E\{\theta^2(1-V)^2\} = \theta^2 \cdot 1$$

while

$$E_\theta\{n^2(T_n-\theta)^2\} \to E\{\theta^2(U-V)^2\} = \theta^2 \cdot 2.$$

Thus the asymptotic mean square error of T_n is twice that of S_n; hence I would prefer S_n.

5. (40 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(\theta, 1)$. Let $|a| < 1$. Then Hodges' (superefficient) estimator T_n of θ is given by

$$\begin{aligned} T_n &= \begin{cases} \overline{X}_n, & \text{if } |\overline{X}_n| > n^{-1/4}, \\ a\overline{X}_n, & \text{if } |\overline{X}_n| \le n^{-1/4}, \end{cases} \\ &= \overline{X}_n 1_{[|\overline{X}_n| > n^{-1/4}]} + a\overline{X}_n 1_{[|\overline{X}_n| \le n^{-1/4}]}. \end{aligned}$$

(a) Show that $\sqrt{n}(T_n-\theta) \to_d N(0, V(\theta))$ where $V(\theta) = 1\{\theta \ne 0\} + a^2 1\{\theta = 0\}$.

(b) Show that if $\theta = \theta_n = tn^{-1/2}$ is true, then $\sqrt{n}(T_n - \theta_n) \to_d aZ + t(a-1) \sim N(t(a-1), a^2)$.

(c) Define a locally regular estimator $\{T_n\}$ of a parameter $\theta \in R^k$. Is Hodges estimator locally regular at $\theta = 0$?

(d) Find the limit of $R_n(\theta_n) = E_{\theta_n}\{n(T_n-\theta_n)^2\}$ where $\theta_n = tn^{-1/2}$ as in (b) and T_n is Hodges' estimator.

Hint: Note that $\sqrt{n}(\overline{X}_n - \theta) \stackrel{d}{=} Z$ for every n.

Solution: See Course Notes, Chapter 3, pages 25-26.

6. Suppose that $\{P_\theta : \theta \in \Theta\}$ where $\Theta \subset \mathbb{R}^d$ is a regular parametric model satisfying the hypotheses A0-A5 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 $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$ holds, what is the limiting distribution of $\sqrt{n}(q(\tilde{\theta}_n) - q(\theta_0))$?

(f) If $\theta = \theta_0$ is true and $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$ 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 $\sqrt{n}(\tilde{\theta}_n - \theta_n) \to_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$, 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 $\dot{q}(\theta_0) = -E_{\theta_0} \dot{\mathbf{l}}_\theta(X;\theta_0) = 0$.

(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 $\ddot{q}(\theta_0) = -E_{\theta_0} \ddot{\mathbf{l}}_{\theta,\theta}(X;\theta_0) = I(\theta_0)$.

(d) Since $q(\theta) = K(P_{\theta_0}, P_\theta)$ is continuous under A0-A4 and $\tilde{\theta}_n$ is consistent, $q(\tilde{\theta}_n) \to_p q(\theta) = K(P_{\theta_0}, P_\theta)$.

(e) When θ_0 is true and $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$, the delta-method (or g'-theorem yields $\sqrt{n}(q(\tilde{\theta}_n) - q(\theta_0)) \to_d 0 \cdot \underline{D} = 0$ in view of (b).

(f) When θ_0 is true and $\sqrt{n}(\tilde{\theta}_n - \theta_0) \to_d \underline{D} \sim N_d(0, I^{-1}(\theta_0))$, then $q(\theta_0) = 0$ and it follows from (b) and (c) that

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

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

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

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.

7. 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:

$$\begin{aligned} \dot{\mathbf{l}}_1(x) = \dot{\mathbf{l}}_\alpha(x) &= \frac{\beta}{\alpha}\left\{\left(\frac{x}{\alpha}\right)^{\beta} - 1\right\} \\ \dot{\mathbf{l}}_2(x) = \dot{\mathbf{l}}_\beta(x) &= \frac{1}{\beta} - \frac{1}{\beta}\log\left\{\left(\frac{x}{\alpha}\right)^{\beta}\right\}\left\{\left(\frac{x}{\alpha}\right)^{\beta} - 1\right\}. \end{aligned}$$

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 $\mathbf{l}_2^*(X) = \dot{\mathbf{l}}_2 - I_{22}I_{11}^{-1}\dot{\mathbf{l}}_1$, the efficient score function for $\beta = \theta_2$, is orthogonal to $[\dot{\mathbf{l}}_1] = \{c\dot{\mathbf{l}}_1 : c \in \mathbb{R}\}$.

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

(c) What is the efficient influence function $\widetilde{\mathbf{l}}_2$ 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 of ψ and the efficient influence function $\widetilde{l}_2$ 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 $\theta_n = (\alpha + sn^{-1/2}, 1 + tn^{-1/2})$? What does $n^{-1}R_n$ converge to in probability when $\beta \neq 1$ is true?

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 $\widetilde{\mathbf{l}}_2(x) = I_{22\cdot 1}^{-1}\dot{\mathbf{l}}_2^*(x)$, and in the particular Weibull case we have

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

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

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

(d) If ψ is the influence function of some (generally inefficient) asymptotically linear estimator T_n of β, then $\widetilde{\mathbf{l}}_\beta$ 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 $\psi - \widetilde{\mathbf{l}}_\beta$ is orthogonal to $\underline{a}^T \dot{\mathbf{l}}_\theta$ in $L_2(P_\theta)$: that is

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

and the asymptotic variance of T_n is given by

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

(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 $\underline{Z}_n(\theta) = n^{-1/2}\sum_{i=1}^n \dot{\underline{\mathbf{l}}}_\theta(X_i|\theta)$ 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,