Statistics 581, Problem Set 8 Solutions

Wellner; 11/25/2009

1. 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 minimum 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 (16) and (17) 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

$$\tilde{l}_1 = I_{11}^{-1} \dot{l}_1 - I_{11}^{-1} I_{12} \tilde{l}_2 . \tag{0.1}$$

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 (0.1). 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:

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

2. Suppose that $(Y|Z) \sim \text{Poisson}(\lambda e^{\gamma Z})$, and $Z \sim G_\eta$ on R with density g_η with respect to some dominating measure μ. You may assume that

$$a(z) \equiv (\partial/\partial\eta)\log g_\eta(z)$$

exists and $E\{a^2(Z)\} < \infty$. Thus Z is a "covariate" or "predictor variable", γ is a "regression parameter" which affects the intensity of the (conditionally) Poisson variable Y, and $\theta = (\gamma, \lambda, \eta)$.
(a) Find the information matrix for θ. What does the structure of this matrix say about the effect of η being known or unknown about the estimation of γ and λ?
(b) Find the information and information bound for γ if the parameter λ is known.
(c) Find the efficient score function and the efficient influence function for estimation of γ when λ is known.
(d) Find the information and information bound for γ if the parameter λ is unknown, $I_{\gamma\gamma\cdot\lambda}$.
(e) Find the efficient score function and the efficient influence function for estimation of γ when λ is unknown.
(f) In the case when $Z \sim \text{Bernoulli}(\eta)$, compute the ratio of the information for γ when λ is unknown, to the information for γ when λ is known as a function of γ and of η.

Solution: (a) The density of $X = (Y, Z)$ is

$$p_\theta(y,z) = f_{\lambda,\gamma}(y|z)g_\eta(z) = e^{-\lambda e^{\gamma z}}\frac{(\lambda e^{\gamma z})^y}{y!}g_\eta(z)\,,$$

and hence

$$\log p_\theta(y,z) = y\log(\lambda e^{\gamma z}) - \lambda e^{\gamma z} - \log y! + \log g_\eta(z)\,.$$

From this we calculate the scores $\dot{l}_\lambda$, $\dot{l}_\gamma$, and $\dot{l}_\eta$:

$$\begin{aligned} \dot{l}_\lambda(y,z) &= \frac{y}{\lambda} - e^{\gamma z} = \frac{1}{\lambda}(y - \lambda e^{\gamma z})\,, \\ \dot{l}_\gamma(y,z) &= yz - \lambda e^{\gamma z}z = z(y - \lambda e^{\gamma z})\,, \\ \dot{l}_\eta(y,z) &= a(z)\,. \end{aligned}$$

This leads to calculating the entries of the information matrix as follows:

$$I_{\lambda,\lambda} = E_\theta(\dot{l}_\lambda(X)^2) = \lambda^{-2}E[(Y - \lambda e^{\gamma Z})^2]$$

$$
\begin{aligned}
&= \lambda^{-2}E[E[(Y-\lambda e^{\gamma Z})^2|Z]] \\
&= \lambda^{-2}E[\lambda e^{\gamma Z}] = E[e^{\gamma Z}]/\lambda\,, \\
I_{\gamma,\gamma} &= E_\theta(\dot{l}_\gamma(X)^2) = E[Z^2(Y-\lambda e^{\gamma Z})^2] \\
&= E[E[Z^2(Y-\lambda e^{\gamma Z})^2|Z]] \\
&= E[Z^2\lambda e^{\gamma Z}] = \lambda E[Z^2 e^{\gamma Z}]\,, \\
I_{\eta,\eta} &= E_\theta a^2(Z)\,, \\
I_{\lambda,\gamma} &= E_\theta(\dot{l}_\lambda \dot{l}_\gamma(X)) = \lambda^{-1}E_\theta(Z(Y-\lambda e^{\gamma Z})^2) \\
&= \lambda^{-1}E[E[Z(Y-\lambda e^{\gamma Z})^2|Z]] \\
&= \lambda^{-1}E[Z\lambda e^{\gamma Z}]\,, \\
I_{\lambda,\eta} &= E_\theta(\dot{l}_\lambda(X)\dot{l}_\eta(X)) = \lambda^{-1}E[(Y-\lambda e^{\gamma Z})a(Z)] \\
&= E[E[a(Z)(Y-\lambda e^{\gamma Z})|Z]] \\
&= E[a(Z)\cdot 0] = 0\,, \\
I_{\gamma,\eta} &= E_\theta(\dot{l}_\gamma(X)\dot{l}_\eta(X)) = E[Z(Y-\lambda e^{\gamma Z})a(Z)] \\
&= E[E[a(Z)Z(Y-\lambda e^{\gamma Z})|Z]] \\
&= E[Za(Z)\cdot 0] = 0\,.
\end{aligned}
$$

Thus the information matrix $I(\theta) = I(\lambda,\gamma,\eta)$ is given by:

$$
I(\theta) = \begin{pmatrix} \lambda^{-1}E(e^{\gamma Z}) & E(Ze^{\gamma Z}) & 0 \\ E(Ze^{\gamma Z}) & \lambda E(Z^2 e^{\gamma Z}) & 0 \\ 0 & 0 & E_\theta a^2(Z) \end{pmatrix}.
$$

(b) If λ is known, the information for γ is $I_{\gamma,\gamma} = \lambda E(Z^2 e^{\gamma Z})$, and the information bound is $1/I_{\gamma,\gamma} = 1/\{\lambda E(Z^2 e^{\gamma Z})\}$.
(c) When λ is known, the efficient score function for γ is just the score function $\dot{l}_\gamma$, and the efficient influence function is $\tilde{l}_\gamma = \dot{l}_\gamma / I_{\gamma,\gamma}$.
(d) When λ is unknown, the (efficient) information for γ is

$$
\begin{aligned}
I_{\gamma\gamma\cdot\lambda} &= I_{\gamma,\gamma} - I_{\gamma,\lambda}I_{\lambda,\lambda}^{-1}I_{\lambda,\gamma} \\
&= \lambda E(Z^2 e^{\gamma Z}) - \frac{[E(Ze^{\gamma Z})]^2}{E(e^{\gamma Z})/\lambda} \\
&= \lambda\left\{E(Z^2 e^{\gamma Z}) - \left(\frac{E(Ze^{\gamma Z})}{E(e^{\gamma Z})}\right)^2 E(e^{\gamma Z})\right\} \\
&= \lambda E(e^{\gamma Z}) Var_{G_\gamma}(Z)\,,
\end{aligned}
$$

where G_γ is the $\gamma-$tilted distribution corresponding to G given by

$$
G_\gamma(A) = \frac{\int_A e^{\gamma z} dG(z)}{\int e^{\gamma z} dG(z)}\,.
$$

(e) When λ is unknown, the efficient score function for γ is

$$
\begin{aligned}
\dot{l}_\gamma^*(x) &= \dot{l}_\gamma(x) - I_{\gamma\lambda}I_{\lambda\lambda}^{-1}\dot{l}_\lambda(x) \\
&= z(y-\lambda e^{\theta z}) - \frac{\lambda E(Ze^{\gamma Z})}{E(e^{\gamma Z})}\lambda^{-1}(y-\lambda e^{\gamma z}) \\
&= \left(z - \frac{E(Ze^{\gamma Z})}{E(e^{\gamma Z})}\right)(y-\lambda e^{\gamma z})\,.
\end{aligned}
$$

Note that $E[l_\gamma^*(X)^2] = I_{\gamma,\gamma\cdot\lambda}$ which we computed in (d). Thus the efficient influence function for γ is $\tilde{l}_\gamma(x) = l_\gamma^*(x)/I_{\gamma\gamma\cdot\lambda}$.
(f) When $Z \sim$ Bernoulli(η), the ratio of the information when λ is unknown to the information when λ is known is

$$
\begin{aligned}
\frac{I_{\gamma,\gamma\cdot\lambda}}{I_{\gamma,\gamma}} &= \frac{\lambda Var_{G_\gamma}(Z)}{\lambda E(Z^2 e^{\gamma Z})} \\
&= \frac{\eta e^\gamma - \left(\frac{\eta e^\gamma}{\eta e^\gamma + (1-\eta)}\right)^2}{\eta e^\gamma} \\
&= \frac{1-\eta}{1-\eta+\eta e^\gamma}\,.
\end{aligned}
$$

See Figure 1 for a plot of this as a function of η for various values of γ.

Figure 1: Ratio of Information for γ with λ known and unknown.

3. (a) Lehmann and Casella, Problem 2.13, page 501.
(b) Let $R_n(\theta) \equiv nE_\theta(T_n - \theta)^2$ where T_n is the Hodges superefficient estimator as in Example 3.3.1 (so $T_n = \delta_n$ of Example 2.5, Lehmann and Casella pages 440 - 443). Show that $R_n(n^{-1/4}) \to \infty$ as $n \to \infty$.

Solution: (a) (a') First recall that (with $\delta_n = T_n$) since $\sqrt{n}(\overline{X} - \theta) \stackrel{d}{=} Z \sim N(0,1)$ we can write

$$
\begin{aligned}
\sqrt{n}(T_n - \theta) &= \sqrt{n}(\overline{X}_n 1_{[|\overline{X}_n| > n^{-1/4}]} + a\overline{X}_n 1_{[|\overline{X}_n| \le n^{-1/4}]} - \theta) \\
&\stackrel{d}{=} Z 1_{[|Z+\theta\sqrt{n}| > n^{1/4}]} + [aZ + \sqrt{n}\theta(a-1)] 1_{[|Z+\theta\sqrt{n}| \le n^{1/4}]} \\
&= Z + [(a-1)Z + (a-1)\sqrt{n}\theta] 1_{[|Z+\theta\sqrt{n}| \le n^{1/4}]} \\
&= Z - (1-a)[Z + \sqrt{n}\theta] 1_{[|Z+\theta\sqrt{n}| \le n^{1/4}]}.
\end{aligned}
$$

Thus

$$\begin{aligned}
b_n(\theta) &= E_\theta(T_n) - \theta \\
&= n^{-1/2}\left\{EZ - (1-a)E[Z+\sqrt{n}\theta]1_{[|Z+\theta\sqrt{n}|\le n^{1/4}]}\right\} \\
&= -\frac{1-a}{\sqrt{n}}E[Z+\sqrt{n}\theta]1_{[|Z+\theta\sqrt{n}|\le n^{1/4}]} \\
&= -\frac{1-a}{\sqrt{n}}\int_{-n^{1/4}}^{n^{1/4}} x\phi(x-\sqrt{n}\theta)dx
\end{aligned}$$

since $Z + \theta\sqrt{n} \sim N(\theta\sqrt{n}, 1)$.
(b') Differentiating the result in (a') gives

$$\begin{aligned}
b_n'(\theta) &= -\frac{1-a}{\sqrt{n}}\int_{-n^{1/4}}^{n^{1/4}} x\phi'(x-\sqrt{n}\theta)(-\sqrt{n})\ dx \\
&= -(1-a)\int_{-n^{1/4}}^{n^{1/4}} x(x-\sqrt{n}\theta)\phi(x-\sqrt{n}\theta)\ dx \qquad \text{since} \quad \phi'(x) = -x\phi(x) \\
&\to 0 \qquad \text{if} \quad \theta \neq 0
\end{aligned}$$

by the dominated convergence theorem since $x(x-\sqrt{n}\theta)\phi(x-\sqrt{n}\theta)1_{[-n^{1/4},n^{1/4}]}(x) \to 0$ for each fixed x and is dominated by the integrable function $4e^{-1}\phi(x)/(|\theta| \wedge 1)$ (for $n \ge (3/|\theta|)^4$).
Details of this domination: For $|x| \le n^{1/4}$ it follows that

$$|x||x-\sqrt{n}\theta| \le n^{1/4}|-n^{1/4}-\sqrt{n}\theta| \le n^{1/2} + n^{3/4}|\theta| \le 2n^{3/4}(|\theta| \vee 1)$$

while

$$\begin{aligned}
\phi(x-\sqrt{n}\theta) &= \phi(x)\exp(\sqrt{n}\theta x - n\theta^2/2) \\
&\le \phi(x)\exp(|\theta|n^{3/4} - n\theta^2/2) \\
&= \phi(x)\exp(|\theta|n^{3/4}(1 - n^{1/4}|\theta|/2)) \\
&\le \phi(x)\exp(-\frac{1}{2}|\theta|n^{3/4}) \qquad \text{if} \quad 1 - n^{1/4}|\theta|/2 < -1/2
\end{aligned}$$

or, equivalently, when $n > (3/|\theta|)^4$. Combining these two bounds yields

$$\begin{aligned}
|x||x-\sqrt{n}\theta|\phi(x-\sqrt{n}\theta) &\le \phi(x)n^{3/4}2(|\theta| \vee 1)\exp(-|\theta|n^{3/4}/2) \\
&= \phi(x)\begin{cases} 2n^{3/4}\exp(-|\theta|n^{3/4}/2) & \text{if} \quad |\theta| < 1 \\ 2n^{3/4}|\theta|\exp(-|\theta|n^{3/4}/2) & \text{if} \quad |\theta| \ge 1 \end{cases} \\
&= \phi(x)\begin{cases} (4/|\theta|)(n^{3/4}|\theta|/2)\exp(-|\theta|n^{3/4}/2) & \text{if} \quad |\theta| < 1 \\ 4(n^{3/4}|\theta|/2)\exp(-|\theta|n^{3/4}/2) & \text{if} \quad |\theta| \ge 1 \end{cases} \\
&\le \frac{4e^{-1}}{|\theta| \wedge 1}\phi(x).
\end{aligned}$$

When $\theta = 0$

$$b_n'(0) = -(1-a)\int_{-n^{1/4}}^{n^{1/4}} x^2\phi(x)\ dx \to -(1-a)\int_{-\infty}^{\infty} x^2\phi(x)dx = -(1-a).$$

(c') The information inequality implies that

$$Var_\theta(\sqrt{n}(T_n - \theta)) \geq \frac{(b_n'(\theta)+1)^2}{I(\theta)} = (b_n'(\theta)+1)^2$$

since $I(\theta) = 1$. At the point $\theta = 0$ the right side converges to a^2, while the limit inferior of the left side is the variance of the limiting distribution at $\theta = 0$, namely a^2. Thus there is no contradiction with the information inequality.

(b) Using the distributional identity in (a) yields

$$\begin{aligned}
R_n(\theta) &= 1 + (1-a)^2 E(Z+\sqrt{n}\theta)^2 1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]} \\
&\quad - 2(1-a)E\{Z(Z+\sqrt{n}\theta)1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]}\} \\
&= 1 + \{(1-a)^2 - 2(1-a)\}E(Z+\sqrt{n}\theta)^2 1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]} \\
&\quad + 2(1-a)\sqrt{n}\theta E\{(Z+\sqrt{n}\theta)1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]}\} \\
&= 1 - (1-a^2)E(Z+\sqrt{n}\theta)^2 1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]} \\
&\quad + 2(1-a)\sqrt{n}\theta E\{(Z+\sqrt{n}\theta)1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]}\}
\end{aligned}$$

(This confirms the first identity in Lehmann's example 4.7, page 442.) Squaring out the expectation in the second term and writing the third term as the sum of two terms yields, with $\alpha_n \equiv n^{1/4} - \sqrt{n}\theta$, $\beta_n \equiv -n^{1/4} - \sqrt{n}\theta$,

$$\begin{aligned}
R_n(\theta) &= 1 - (1-a^2)EZ^2 1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]} \\
&\quad - 2(1-a^2)\sqrt{n}\theta EZ1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]} \\
&\quad - (1-a^2)n\theta^2(\Phi(\beta_n) - \Phi(\alpha_n)) \\
&\quad + 2(1-a)n\theta^2(\Phi(\beta_n) - \Phi(\alpha_n)) \\
&\quad + 2(1-a)\sqrt{n}\theta E\{Z1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]}\} \\
&= 1 - (1-a^2)EZ^2 1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]} \\
&\quad +(1-a)^2 n\theta^2(\Phi(\beta_n) - \Phi(\alpha_n)) \\
&\quad -2a(1-a)\sqrt{n}\theta E(Z1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]})
\end{aligned}$$

where

$$\begin{aligned}
E(Z1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]}) &= \int_{\alpha_n}^{\beta_n} z\phi(z)dz \\
&= -\int_{\alpha_n}^{\beta_n} \phi'(z)dz \qquad \text{since} \quad \phi'(z) = -z\phi(z) \\
&= -(\phi(\beta_n) - \phi(\alpha_n)).
\end{aligned}$$

Thus it follows that

$$\begin{aligned}
R_n(\theta) &= 1 - (1-a^2)EZ^2 1_{[|Z+\theta\sqrt{n}| \leq n^{1/4}]} \\
&\quad + (1-a)^2 n\theta^2(\Phi(\beta_n) - \Phi(\alpha_n)) \\
&\quad + 2a(1-a)\sqrt{n}\theta(\phi(\beta_n) - \phi(\alpha_n)).
\end{aligned}$$

(This confirms the second identity in Lehmann's problem 4.7, page 442.) Now we take $\theta = \theta_n = n^{-1/4}$, and note that $\alpha_n = -2n^{1/4}$, $\beta_n = 0$. Since the expectation of in the second term in the last display is bounded below by zero and above by 1 we find that

$$\begin{aligned} R_n(n^{-1/4}) &\geq a^2 + (1-a)^2 n^{1/2}(1/2 - \Phi(-2n^{1/4}) \\ &\qquad +2a(1-a)n^{1/4}(\phi(0) - \phi(-2n^{1/4})) \\ &\to a^2 + \infty + \infty = \infty \end{aligned}$$

since $n^{1/2}\Phi(-2n^{1/4}) \to 0$ and $n^{1/4}\phi(-2n^{1/4}) \to 0$.

(b), Second (more elegant) solution: Julia Palacios (and others): from the lecture notes, 3.3 (3), it follows that

$$R_n(\theta) = E[n(T_n - \theta)^2] = nVar[T_n] + nb_n(\theta)^2 \geq a^2 + nb_n(\theta)^2.$$

Using the formula for $b_n(\theta)$ from part (a) above, it follows that it is enough to show that

$$\left| \int_{-n^{1/4}}^{n^{1/4}} x\phi(x - n^{1/4})dx \right| \to \infty.$$

But we have, with $Z \sim N(0,1)$ (and hence $E|Z| < \infty$),

$$\begin{aligned} \left| \int_{-n^{1/4}}^{n^{1/4}} x\phi(x - n^{1/4})dx \right| &= \left| \int_{-2n^{1/4}}^{0} (y + n^{1/4})\phi(y)dy \right| \\ &\geq \left| n^{1/4} \int_{-2n^{1/4}}^{0} \phi(y)dy \right| - \left| \int_{-2n^{1/4}}^{0} y\phi(y)dy \right| \\ &\geq n^{1/4}(\Phi(0) - \Phi(-2n^{1/4})) - E|Z| \\ &\to \infty. \end{aligned}$$

4. Suppose that $Z \sim N(0,1)$ and, for $\mu \in R$ and $\sigma > 0$, that $X = \mu + \sigma Z \sim P_{\mu,\sigma} = N(\mu, \sigma^2)$.
(a) Compute the likelihood ratio

$$\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) = \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} \qquad \text{and} \qquad Y \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X)\,.$$

What is the distribution of Y under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?
(b) Plot the function $l(\mu; X) \equiv \log(dP_{\mu,\sigma}/dP_{0,\sigma})(X)$ as a function of μ.
(c) Find the maximum value of the function $l(\mu; X)$ in B (as a function of μ) and the value of $\mu \equiv \hat{\mu}$ which achieves the maximum.
(d) What is the distribution of $\hat{\mu}$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$? What is the distribution of $l(\hat{\mu}; X)$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?

Solution: A. The likelihood ratio

$$\begin{aligned} \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) &= \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} = \frac{\exp(-(x-\mu)^2/(2\sigma^2))}{\exp(-x^2/(2\sigma^2))} \\ &= \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2}\frac{\mu^2}{\sigma^2}\right). \end{aligned}$$

Hence

$$Y \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{1}{2}\frac{\mu^2}{\sigma^2}.$$

Under $P_{0,\sigma}$ we find that $E(Y) = 0 - \frac{\mu^2}{2\sigma^2}$ and $Var(Y) = \mu^2/\sigma^2 \equiv V^2$ so that

$$Y \sim N(-\frac{1}{2}V^2, V^2) \qquad \text{under } \ P_{0,\sigma}.$$

Under $P_{\mu,\sigma}$ a similar computation gives $E(Y) = \mu^2/\sigma^2 - \mu^2/(2\sigma^2) = V^2/2$ and $Var(Y) = V^2$, so

$$Y \sim N(\frac{1}{2}V^2, V^2) \qquad \text{under } \ P_{\mu,\sigma}.$$

B and C. The function

$$l(\mu,\sigma;X) \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) = \frac{\mu}{\sigma}\frac{X}{\sigma} - \frac{\mu^2}{2\sigma^2} = \frac{X^2}{2\sigma^2} - \frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}$$

is quadratic in μ with maximum value $X^2/(2\sigma^2)$ which is achieved at $\mu = \hat{\mu} \equiv X$.
D. Under $P_{0,\sigma}$, $\hat{\mu} = X \sim N(0,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2/2$. Under $P_{\mu,\sigma}$, $\hat{\mu} = X \sim N(\mu,\sigma^2)$ and $l(\hat{\mu},\sigma;X) = X^2/(2\sigma^2) \sim \chi_1^2(\delta)/2$ with $\delta = \mu^2/\sigma^2$.

5. Read Note 8.5, Lehmann and Casella, page 145. Explore the identity in the second display in this note and see if it makes sense as written. If not, rewrite the identity in a way that makes sense to you. [Compare with Efron and Johnstone (1990) and/or Bickel, Klaassen, Ritov, and Wellner (1993), pages 420-424.]

Solution: The identity as written in Lehmann and Casella, page 145 is

$$I(\theta) = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} \log[f_\theta(x)]^2 f_\theta(x)dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} \log[h_\theta(x)]^2 f_\theta(x)dx.$$

This is correct *only if* it is interpreted as

$$I(\theta) = \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial\theta}\log[f_\theta(x)]\right\}^2 f_\theta(x)dx = \int_{-\infty}^{\infty} \left\{\frac{\partial}{\partial\theta}\log[h_\theta(x)]\right\}^2 f_\theta(x)dx; \qquad (0.2)$$

compare with BKRW (1993), page 424, last line. The identity holds since

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log h_\theta(x) &= \frac{\partial}{\partial\theta}\log\left(\frac{f_\theta(x)}{1-F_\theta(x)}\right) \\
&= \frac{\partial}{\partial\theta}\log f_\theta(x) - \frac{\partial}{\partial\theta}\log(1-F_\theta(x)) \\
&= \dot{l}_\theta(x) - \frac{1}{1-F_\theta(x)}\int_x^\infty \dot{l}_\theta(y) f_\theta(y)dy \\
&\equiv (R\dot{l}_\theta)(x)
\end{aligned}$$

where R is the operator defined for $a \in L_2^0(F_\theta)$ by

$$\begin{aligned}
Ra(x) &= a(x) - \frac{1}{1-F_\theta(x)}\int_x^\infty a(y) f_\theta(y)dy \\
&= a(x) - E_\theta(a(X)|X > x).
\end{aligned}$$

As shown in BKRW (1993), Proposition A.1.8 (page 421) and in Efron and Johnstone (1990), R is an isometry on $L_2^0(F_\theta)$: if $X \sim F$ and $Ea(X) = 0$, then

$$\begin{aligned} Var[a(X)] = Ea^2(X) &= E[Ra(X)]^2 \\ &= E\{a(X) - \frac{1}{1-F(X)} \int_X^\infty a(y)dF(y)\}^2. \end{aligned} \tag{0.3}$$

Applying this identity to $a = \dot{l}_\theta$ with $X \sim F_\theta$ yields (0.2). This is sufficient for solving the problem as stated.

To see that (0.3) holds, we write

$$\begin{aligned} E[Ra(X)]^2 &= Ea^2(X) - 2E\left\{\frac{a(X)}{1-F(X)} \int_{(X,\infty)} a(y)dF(y)\right\} \\ &\quad + E\left\{\left(\frac{1}{(1-F(X))} \int_{(X,\infty)} a(y)dF(y)\right)^2\right\} \\ &= Ea^2(X) \end{aligned}$$

since, by repeated use of Fubini's theorem,

$$\begin{aligned} & E\left\{\frac{1}{(1-F(X))^2}\left(\int_{(X,\infty)} a(y)dF(y)\right)^2\right\} \\ &= E\left\{\frac{1}{1-F(X))^2} \int_{(X,\infty)} a(y)dF(y) \int_{(X,\infty)} a(y')dF(y')\right\} \\ &= \int_{-\infty}^\infty \frac{1}{(1-F(x))^2} \int_{(x,\infty)} a(y)dF(y) \int_{(x,\infty)} a(y')dF(y')dF(x) \\ &= \int_{-\infty}^\infty\int_{-\infty}^\infty\left(\int_{-\infty}^\infty 1_{[x<y]}1_{[x<y']}(1-F(x))^{-2}dF(x)\right) a(y)a(y')dF(y)dF(y') \\ &= \int_{-\infty}^\infty\int_{-\infty}^\infty\left(\frac{1}{1-F(y\wedge y')} - 1\right) a(y)a(y')dF(y)dF(y') \\ &= \int_{-\infty}^\infty\int_{-\infty}^\infty \frac{1}{1-F(y\wedge y')} a(y)a(y')dF(y)dF(y') \text{ since } \int_{-\infty}^\infty a(y)dF(y) = 0 \\ &= 2\int_{-\infty}^\infty \frac{a(y)}{1-F(y)}\left(\int_{(y,\infty)} a(y')dF(y')\right) dF(y) \text{ by symmetry} \\ &= 2E\left\{\frac{a(X)}{1-F(X)} \int_{(X,\infty)} a(y')dF(y')\right\}. \end{aligned}$$

A neat way to see the same identity is via martingale methods as in BKRW, pages 421-422.