On and Off Semiparametric Models

Jon A. Wellner

University of Washington
• joint work with Takumi Saegusa,
  University of Washington, Biostatistics
• Talk at First IMS Asia-Pacific Rim Conference,
• Seoul, Korea
• June 29, 2009
• Email: jaw@stat.washington.edu
  http://www.stat.washington.edu/jaw/jaw.research.html
Outline

- Introduction: estimation theory on a model $\mathcal{P}$
Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts
Outline

• Introduction: estimation theory on a model $\mathcal{P}$
• Estimation theory off a model $\mathcal{P}$: notation and concepts
• Estimation theory off a model $\mathcal{P}$: three questions
Outline

• Introduction: estimation theory on a model $\mathcal{P}$
• Estimation theory off a model $\mathcal{P}$: notation and concepts
• Estimation theory off a model $\mathcal{P}$: three questions
• An answer to question 1
Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts
- Estimation theory off a model $\mathcal{P}$: three questions
- An answer to question 1
- An answer to questions 2 and 3
Outline

• Introduction: estimation theory on a model $\mathcal{P}$
• Estimation theory off a model $\mathcal{P}$: notation and concepts
• Estimation theory off a model $\mathcal{P}$: three questions
• An answer to question 1
• An answer to questions 2 and 3
• Examples
  ◦ Example 1. Symmetric location model
  ◦ Example 2. Paired exponential mixture model
• Summary; further problems
1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. $P_0$. 
1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. $P_0$.
- Let $\mathcal{P} \equiv \mathcal{P}(P_0)$ be the tangent space of $\mathcal{P}$ at $P_0$. 
1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. $P_0$.
- Let $\dot{\mathcal{P}} \equiv \dot{\mathcal{P}}(P_0)$ be the tangent space of $\mathcal{P}$ at $P_0$.
- Assume that $\dot{\mathcal{P}} \subsetneq L^0_2(P_0)$; thus $L^0_2(P_0) = \dot{\mathcal{P}} + \dot{\mathcal{P}}^\perp$ where $\dot{\mathcal{P}}^\perp \neq \emptyset$. 
1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. $P_0$.
- Let $\dot{\mathcal{P}} \equiv \dot{\mathcal{P}}(P_0)$ be the tangent space of $\mathcal{P}$ at $P_0$.
- Assume that $\dot{\mathcal{P}} \subsetneq L^0_2(P_0)$; thus $L^0_2(P_0) = \dot{\mathcal{P}} + \dot{\mathcal{P}}^\perp$ where $\dot{\mathcal{P}}^\perp \neq \emptyset$.
- Suppose that $\nu = \nu(P)$ is a (differentiable) parameter: i.e. $\nu : \mathcal{P} \rightarrow \mathbb{R}$. 
• Suppose that $\hat{\nu}_n$ is an inefficient estimator of $\nu(P)$ at $P_0$ with influence function $\psi$: thus

$$\sqrt{n}(\hat{\nu}_n - \nu(P_0)) = \sqrt{n} \mathbb{P}_n \psi + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) + o_p(1)$$

where $E_0 \psi(X_1) = 0$, $E_0 \psi^2 < \infty$, and $\psi \notin \mathcal{P}$.
• Suppose that $\hat{\nu}_n$ is an inefficient estimator of $\nu(P)$ at $P_0$ with influence function $\psi$: thus

$$\sqrt{n}(\hat{\nu}_n - \nu(P_0)) = \sqrt{n}P_n\psi + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) + o_p(1)$$

where $E_0\psi(X_1) = 0$, $E_0\psi^2 < \infty$, and $\psi \notin \hat{\mathcal{P}}$.

• It is well-known that the efficient influence function $\tilde{I}_\nu$ for estimation of $\nu$ on $\mathcal{P}$ is given by

$$\tilde{I}_\nu = \Pi(\psi|\hat{\mathcal{P}}) \equiv \text{orthogonal projection of } \psi \text{ onto } \hat{\mathcal{P}};$$

see e.g. Bickel, Klaassen, Ritov, and Wellner (1993, 1998), Proposition 1, page 65.
\[ \psi - \tilde{l}_\nu \]
• If $\hat{\nu}_n^{\text{eff}}$ is a locally regular and asymptotically efficient estimator of $\nu(P_0)$, then $\hat{\nu}_n^{\text{eff}}$ satisfies

$$\sqrt{n}(\hat{\nu}_n^{\text{eff}} - \nu(P_0)) = \sqrt{n}P_n \tilde{I}_\nu + o_p(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{I}_\nu(X_i) + o_p(1).$$
2. Estimation theory off a model $\mathcal{P}$

- George Box (1987): “Essentially, all models are wrong, but some are useful.”
2. Estimation theory off a model $\mathcal{P}$

- George Box (1987): “Essentially, all models are wrong, but some are useful.”
- Variant 1: “Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.”
2. Estimation theory off a model $\mathcal{P}$

- George Box (1987): “Essentially, all models are wrong, but some are useful.”

- Variant 1: “Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.”

- Variant 2: “All models are false, but some models are useful.”
Consider enlarging $\mathcal{P}$ by allowing an additional parametric sub-model $\mathcal{Q} \equiv \{ Q_\eta : \eta \in \mathbb{R} \}$ parametrized by a real parameter $\eta$ which satisfies $Q_0 = P_0 \in \mathcal{P}$. 
• Consider enlarging $\mathcal{P}$ by allowing an additional parametric sub-model $\mathcal{Q} \equiv \{ Q_\eta : \eta \in \mathbb{R} \}$ parametrized by a real parameter $\eta$ which satisfies $Q_0 = P_0 \in \mathcal{P}$.

• Suppose that $\mathcal{Q}$ is regular with, for simplicity, densities $\{q_\eta\}$ with respect to a dominating measure $\mu$ and score function $a \in L^0_2(P_0)$ but $a \notin \mathcal{P}$. 
• Consider enlarging $\mathcal{P}$ by allowing an additional parametric sub-model $\mathcal{Q} \equiv \{ Q_\eta : \eta \in \mathbb{R} \}$ parametrized by a real parameter $\eta$ which satisfies $Q_0 = P_0 \in \mathcal{P}$.

• Suppose that $\mathcal{Q}$ is regular with, for simplicity, densities $\{q_\eta\}$ with respect to a dominating measure $\mu$ and score function $a \in L_2^0(P_0)$ but $a \notin \dot{\mathcal{P}}$.

• Thus for the sequence of densities $q_n \equiv q_{n-1/2}$,

$$\sqrt{n} \left\{ \sqrt{q_n} - \sqrt{p_0} \right\} \rightarrow \frac{1}{2} a \sqrt{p_0} \quad \text{in} \quad L_2(\mu).$$
Given $a \in L_2^0(P_0) \setminus \mathcal{P}$, define a one-dimensional parametric submodel

$$Q = \{Q_\eta : \frac{dQ_\eta}{d\mu} \equiv q_\eta, \eta \in \mathbb{R}\}$$

where, with $m(x) = 2/(1 + e^{-x})$,

$$q_\eta(x) = p_0(x) \frac{m(\eta a(x))}{\int m(\eta a)dP_0}.$$

Then $Q_\eta$ has score function $a$ and satisfies $Q_0 = P_0$. 
Now asymptotic linearity of $\hat{\nu}_n$ and Le Cam’s second lemma yields

\[
\left( \frac{\sqrt{n}(\hat{\nu}_n - \nu(P_0))}{\log \frac{dQ^n}{dP_0^n}} \right) \xrightarrow{P^n_0} dN_2 \left( \begin{pmatrix} 0 \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} E\psi^2 & E(\psi a) \\ E(\psi a) & \sigma^2 \end{pmatrix} \right)
\]

where $\sigma^2 = E\alpha^2(X)$. 
• Hence by Le Cam’s third lemma we find that

\[
\sqrt{n}(\hat{\nu}_n - \nu(P_0)) \rightarrow_d N_2 \left( \left( \begin{array}{c} E(\psi a) \\ +\sigma^2/2 \end{array} \right), \left( \begin{array}{cc} E\psi^2 & E(\psi a) \\ E(\psi a) & \sigma^2 \end{array} \right) \right)
\]

where \( \sigma^2 = Ea^2(X) \).
• Hence by Le Cam’s third lemma we find that

\[
\begin{align*}
\left( \sqrt{n}(\hat{\nu}_n - \nu(P_0)) \right) \\
\text{log} \frac{dQ^n}{dP^n}
\end{align*}
\]

\[
\xrightarrow{Q^n_{\rightarrow d}} N_2 \left( \left( \begin{array}{c} E(\psi a) \\ +\sigma^2/2 \end{array} \right), \left( \begin{array}{cc} E\psi^2 & E(\psi a) \\ E(\psi a) & \sigma^2 \end{array} \right) \right)
\]

where \( \sigma^2 = E\alpha^2(X) \).

• This implies that

\[
\liminf_n E_{Q^n} \left\{ \sqrt{n}(\hat{\nu}_n - \nu(P_0)) \right\}^2 \geq E\psi^2 + \{E(\psi a)\}^2
\]

\[
\equiv \text{AMSE}_{\hat{\nu}}(\alpha).
\]

where \( \text{AMSE}_T(\alpha) \) stands for the “Asymptotic Mean Squared Error of the estimator \( T \) in the direction \( \alpha \)”. 

• Repeating this argument for the efficient estimator $\hat{\nu}_n^{\text{eff}}$ with efficient influence function $\tilde{l}_\nu$ yields

$$
\left( \sqrt{n}(\hat{\nu}_n^{\text{eff}} - \nu(P_0)) \right) \\
\log \frac{dQ_n}{dP_0^n}
$$

\[ Q_n \xrightarrow{d} N_2 \left( \left( \begin{array}{c} E(\tilde{l}_\nu a) \\ + \sigma^2 / 2 \end{array} \right), \left( \begin{array}{cc} E\tilde{l}_\nu^2 & E(\tilde{l}_\nu a) \\ E(\tilde{l}_\nu a) & \sigma^2 \end{array} \right) \right) \]

where $\sigma^2 = E\alpha^2(X)$. 
Repeating this argument for the efficient estimator $\hat{\nu}_n^{\text{eff}}$ with efficient influence function $\tilde{l}_\nu$ yields

\[
\left( \sqrt{n}(\hat{\nu}_n^{\text{eff}} - \nu(P_0)) \right) \log \frac{dQ_n}{dP_n} \rightarrow_d N_2 \left( \left( \begin{array}{c} E(\tilde{l}_\nu a) \\ +\sigma^2/2 \end{array} \right), \left( \begin{array}{cc} E\tilde{l}_\nu^2 & E(\tilde{l}_\nu a) \\ E(\tilde{l}_\nu a) & \sigma^2 \end{array} \right) \right)
\]

where $\sigma^2 = Ea^2(X)$.

This implies that

\[
\liminf_n E_{Q_n} \left\{ \sqrt{n}(\hat{\nu}_n^{\text{eff}} - \nu(P_0)) \right\}^2 \geq E\tilde{l}_\nu^2 + \{E(\tilde{l}_\nu a)\}^2 \\
\equiv \text{AMSE}_{\hat{\nu}_n^{\text{eff}}(a)}.
\]
3. Estimation theory off a model $\mathcal{P}$: two questions

- **Question 1:** When does it hold that

$$\text{AMSE}_{\tilde{\nu}}(a) = E(\psi^2) + \{E(\psi a)\}^2$$

$$< E(\tilde{I}_\nu^2) + \{E(\tilde{I}_\nu a)\}^2$$

$$= \text{AMSE}_{\tilde{\nu}}^{\text{eff}}(a)?$$
3. Estimation theory off a model \( \mathcal{P} \): two questions

- **Question 1:** When does it hold that
  \[
  \text{AMSE}_{\hat{\nu}}(a) = E(\psi^2) + \{E(\psi a)\}^2 \\
  < E(\tilde{I}_{\nu}^2) + \{E(\tilde{I}_{\nu} a)\}^2 \\
  = \text{AMSE}_{\hat{\nu} \text{eff}}(a) ?
  \]

- **Question 2:** How large can the difference
  \[
  \text{AMSE}_{\nu_n}(a) - \text{AMSE}_{\hat{\nu} \text{eff}}(a)
  \]
  be, subject to a bound on \( E \alpha^2 \)? That is, can we bound
  \[
  \sup \left\{ E\psi^2 + \{E(\psi a)\}^2 - \left( E\tilde{I}_{\nu}^2 + \{E(\tilde{I}_{\nu} a)\}^2 \right) : a \notin \hat{\mathcal{P}}, \ E\alpha^2 \leq \kappa^2 \right\} ?
  \]
• Question 3: How large can the difference

\[ \text{AMSE}_{\tilde{\nu}}(a) - \text{AMSE}_{\tilde{\nu}_n}(a) \]

be, subject to a bound on \( E_a^2 \)? That is, can we bound

\[
\sup \left\{ E\tilde{1}_\nu^2 + \{ E(\tilde{1}_\nu a) \}^2 - (E\psi^2 + \{ E(\psi a) \}^2) : a \notin \hat{\mathcal{P}}, E_a^2 \leq \kappa^2 \right\}
\]
4. An answer to question 1

- Proposition 1. Let

\[
\tilde{a} = \Pi \left( a \mid \begin{bmatrix} \psi, \tilde{1}_\nu \end{bmatrix} \right) = c\frac{\psi - \tilde{1}_\nu}{E(\psi - \tilde{1}_\nu)^2} + d\frac{\tilde{1}_\nu}{E(\tilde{1}_\nu^2)}.
\]

Then

\[
\text{AMSE}_{\tilde{\nu}}(a) = E(\psi^2) + \{E(\psi a)\}^2 < E(\tilde{1}_\nu^2) + \{E(\tilde{1}_\nu a)\}^2 = \text{AMSE}_{\tilde{\nu}}^{\text{eff}}(a)
\]

(1)

if \(d^2 > E(\psi - \tilde{1}_\nu)^2 \equiv B^2\) and

\[-d - \sqrt{d^2 - B^2} < c < -d + \sqrt{d^2 - B^2}.\]
Proof: Since

\[ E(\psi^2) = E(\psi - \tilde{l}_\nu)^2 + E(\tilde{l}_\nu^2), \]

the inequality (1) is equivalent to

\[ E(\psi - \tilde{l}_\nu)^2 + \{E(\psi a)\}^2 < \{E(\tilde{l}_\nu a)\}^2. \] (2)

But since

\[ E(\tilde{l}_\nu a) = E(\tilde{l}_\nu \tilde{a}) = d, \quad \text{and} \]

\[ E(\psi a) = E(\psi \tilde{a}) = E \left( \psi - \tilde{l}_\nu + \tilde{l}_\nu \right) \left( c \frac{\psi - \tilde{l}_\nu}{E(\psi - \tilde{l}_\nu)^2} + d \frac{\tilde{l}_\nu}{E(\tilde{l}_\nu^2)} \right) \]

\[ = c + d, \]

we can rewrite (2) as:
\[ B^2 \equiv E(\psi - \tilde{\nu})^2 < d^2 - (d + c)^2 = -2cd - c^2, \]

or, equivalently

\[ c^2 + 2dc + B^2 < 0. \]

This holds only if \( d^2 > B^2 \), and

\[ -d - \sqrt{d^2 - B^2} < c < -d + \sqrt{d^2 - B^2}. \]
\[
d^2_0 \equiv B^2 = E(\psi - \tilde{l}_\nu)^2
\]
\[
\tilde{a}_0 = -d_0 \frac{\psi - \tilde{l}_\nu}{E(\psi - \tilde{l}_\nu)^2} + d_0 \frac{\tilde{l}_\nu}{E(\tilde{l}_\nu)^2}
\]
\[
\tilde{a}_0 \perp \psi
\]
\[ \tilde{a}_L, \quad c = -d - \sqrt{d^2 - B^2} \]

\[ \tilde{a}_U, \quad c = -d + \sqrt{d^2 - B^2} \]
\[ \tilde{a}_L, \quad c = -d - \sqrt{d^2 - B^2} \]

\[ \tilde{a}_U, \quad c = -d + \sqrt{d^2 - B^2} \]
5. An answer to question 2

- “Proposition 2.F” (F for false! or first attempt)

\[
\sup \left\{ E\psi^2 + \{E(\psi a)\}^2 - \left( E\tilde{l}_\nu^2 + \{E(\tilde{l}_\nu a)\}^2 \right) : a \notin \hat{P}, \ E a^2 \leq \kappa^2 \right\} \\
\leq \frac{1}{I_\nu} r \left\{ 1 + \kappa^2 \left( 1 + \frac{1}{r + 1} \right) \right\} \\
\leq B^2 \{1 + 2\kappa^2\}, \quad r \equiv E(\psi - \tilde{l}_\nu)^2 / E\tilde{l}_\nu^2 = I_\nu B^2.
\]
5. An answer to question 2

- “Proposition 2.F” (F for false! or first attempt)

\[
\sup \left\{ E\psi^2 + \{E(\psi a)\}^2 - \left( E\tilde{l}_\nu^2 + \{E(\tilde{l}_\nu a)\}^2 \right) : a \notin \hat{\mathcal{P}}, \ Ea^2 \leq \kappa^2 \right\}
\]

\[
\leq \frac{1}{I_\nu} r \left\{ 1 + \kappa^2 \left( 1 + \frac{1}{r + 1} \right) \right\}
\]

\[
\leq B^2 \{1 + 2\kappa^2\}, \quad r \equiv E(\psi - \tilde{l}_\nu)^2 / E\tilde{l}_\nu^2 = I_\nu B^2.
\]

- “Proof.” Since \( E\psi^2 = E\tilde{l}_\nu^2 + E(\psi - \tilde{l}_\nu)^2 \),

\[
E\psi^2 + \{E(\psi a)\}^2 - \left( E\tilde{l}_\nu^2 + \{E(\tilde{l}_\nu a)\}^2 \right)
\]

\[
= E(\psi - \tilde{l}_\nu)^2 + \{E(\psi a)\}^2 - \{E(\tilde{l}_\nu a)\}^2
\]

\[
\leq E(\psi - \tilde{l}_\nu)^2 + E(\psi^2) E(a^2) - \{E(\tilde{l}_\nu a)\}^2
\]

where the inequality follows by Cauchy-Schwarz,
"Proof", cont’d: and equality holds if \( a = C \psi \) for some \( C \). Taking \( a = C \psi \), the right side in the last display becomes

\[
E(\psi - \tilde{1}_\nu)^2 + E(\psi^2)E(a^2) - \{E(\tilde{1}_\nu a)\}^2
\]

\[
= E(\psi - \tilde{1}_\nu)^2 + C^2 \left\{ \{E(\psi^2)\}^2 - \{E(\tilde{1}_\nu \psi)\}^2 \right\}
\]

\[
= E(\psi - \tilde{1}_\nu)^2 + C^2 \left\{ \{E(\psi^2)\}^2 - \{E(\tilde{1}_\nu^2)\}^2 \right\}
\]

since \( \psi - \tilde{1}_\nu \perp \tilde{1}_\nu \)

\[
\leq E(\psi - \tilde{1}_\nu)^2 + \frac{\kappa^2}{E\psi^2} \left\{ \{E(\psi^2)\}^2 - \{E(\tilde{1}_\nu^2)\}^2 \right\}
\]

since \( Ea^2 = C^2 E\psi^2 \leq \kappa^2 \) if \( C^2 \leq \kappa^2 / E\psi^2 \)

\[
= E(\psi - \tilde{1}_\nu)^2 + \frac{\kappa^2}{E\psi^2} \left\{ E(\psi^2) - E(\tilde{1}_\nu^2) \right\} \left\{ E(\psi^2) + E(\tilde{1}_\nu^2) \right\}
\]

\[
= E(\psi - \tilde{1}_\nu)^2 \left\{ 1 + \kappa^2 \left\{ 1 + \frac{E(\tilde{1}_\nu^2)}{E(\psi^2)} \right\} \right\}.
\]
• **Proposition 2.** An answer to question 2

\[
\sup \left\{ E\psi^2 + \{E(\psi a)\}^2 - \left( E\tilde{l}_\nu^2 + \{E(\tilde{l}_\nu a)\}^2 \right) : a \notin \hat{P}, \ Ea^2 \leq \kappa^2 \right\}
\]

\[
= \frac{1}{I_\nu} r \left\{ 1 + \kappa^2 \frac{1 + (2 + r)s_+}{2 + rs_+} \right\}
\]

\[
\sim \frac{1}{I_\nu} r \left\{ 1 + \kappa^2 \frac{2 + r}{1 + r} \right\}, \quad \text{as } r \to \infty
\]

where

\[
s_+ \equiv s_+(r) \equiv \sqrt{\frac{1}{4} + \frac{1}{r} + \frac{1}{2}}.
\]
• Proposition 3. An answer to question 3

\[
\sup \left\{ E \hat{\nu}^2 + \{ E(\hat{1}_\nu a) \}^2 - (E \psi^2 + \{ E(\psi a) \}^2) : a \notin \hat{\mathcal{P}}, E a^2 \leq \kappa^2 \right\}
\]

\[
= \frac{r}{I_\nu} \left\{ \kappa^2 \frac{(2 + r)s_- - 1}{2 - rs_-} - 1 \right\} > 0
\]

if

\[
\kappa^2 > \frac{2 - rs_-}{(2 + r)s_- - 1} > 0
\]

where

\[
s_- \equiv s_-(r) \equiv \sqrt{\frac{1}{4} + \frac{1}{r} - \frac{1}{2}}.
\]
• Sketch of Proofs for Propositions 2 and 3.

First consider the case that

\[ a = \tilde{a} = c \frac{\psi - \tilde{l}_\nu}{E(\psi - \tilde{l}_\nu)^2} + d \frac{\tilde{l}_\nu}{E(\tilde{l}_\nu^2)}. \]

In this case we seek to maximize (Proposition 2) or minimize (Proposition 3)

\[ c^2 + 2cd \]

subject to the constraint

\[ \kappa^2 \geq E(\tilde{\alpha}^2) = \frac{c^2}{E(\psi - \tilde{l}_\nu)^2} + \frac{d^2}{E(\tilde{l}_\nu^2)} \]

\[ = \frac{c^2}{B^2} + I_\nu d^2 = \frac{1}{B^2} (c^2 + rd^2). \]
Carry this out by maximizing (or minimizing)

\[ f(c, d, \lambda) = c^2 + 2dc + \lambda \left( \kappa^2 - \frac{c^2}{B^2} - I_\nu d^2 \right). \]

Then show that the solution for general \( a = \tilde{a} + (a - \tilde{a}) \) is achieved when \( a = \tilde{a} \in [\psi, \tilde{l}_\nu] \).

Geometric view?
\[-(1 + c^2 + 2cd)\]
Positive part of $-(1 + c^2 + 2cd)$
Negative part of $-(1 + c^2 + 2cd)$
Positive part of $-(1 + c^2 + 2cd)$ restricted to $\kappa^2 \geq (c^2 + rd^2)/B^2$; $r = 3$, $\kappa^2 = 6$
Negative part of \(-(1 + c^2 + 2cd)\) restricted to \(\kappa^2 \geq (c^2 + rd^2)/B^2; \quad r = 3, \quad \kappa^2 = 6\)
Example 1. Symmetric location model

- Model:

\[ P = \{ P_{\nu, f} : \frac{dP_{\nu, f}}{d\lambda}(x) = f(x-\nu), \, \nu \in \mathbb{R}, \, f \text{ symmetric at } 0, I_f < \infty \} \]
6. Examples

Example 1. Symmetric location model

- **Model:**

  \[ \mathcal{P} = \{ P_{\nu,f} : \frac{dP_{\nu,f}}{d\lambda}(x) = f(x-\nu), \ \nu \in \mathbb{R}, \ f \text{ symmetric at } 0, I_f < \infty \}. \]

- **Tangent space:** At \( P_0 = P_{0,f_0} \) the tangent space is

  \[ \dot{\mathcal{P}} = \{-f_0'/f_0 + \text{all even functions } h \in L_2(P_0)\}. \]
6. Examples

Example 1. Symmetric location model

- **Model:**

\[
\mathcal{P} = \{ P_{\nu,f} : \frac{dP_{\nu,f}}{d\lambda}(x) = f(x-\nu), \, \nu \in \mathbb{R}, \, f \text{ symmetric at } 0, I_f < \infty \}.
\]

- **Tangent space:** At \( P_0 = P_{0, f_0} \) the tangent space is

\[
\dot{\mathcal{P}} = \{ [-f'_0/f_0] + \text{all even functions } h \in L_2(P_0) \}.
\]

- **Inefficient estimator:** Let \( \hat{\nu}_n \) be the M-estimator corresponding to the logistic density; i.e. the solution of

\[
\mathbb{P}_n \tilde{\psi}(X - \nu) = 0,
\]

where \( g \) is the logistic density, \( g(x) = e^{-x}/(1 + e^{-x})^2 \), and

\[
\tilde{\psi}(x) = -\frac{g'(x)}{g(x)} = \frac{1 - e^{-x}}{1 + e^{-x}}.
\]
- **Influence function** $\psi$: $\hat{\nu}_n$ is asymptotically linear with influence function

$$
\psi(x) = \frac{\tilde{\psi}(x)}{\tilde{\Psi}'(0)}, \quad \text{where} \quad \tilde{\Psi}(\nu) = P_0 \tilde{\psi}(X - \nu),
$$
• **Influence function** $\psi$: $\hat{\nu}_n$ is asymptotically linear with influence function

$$
\psi(x) = \frac{\tilde{\psi}(x)}{\tilde{\Psi}'(0)},
$$

where

$$
\tilde{\Psi}(\nu) = P_0 \tilde{\psi}(X - \nu),
$$

• **Efficient influence function:**

$$
\tilde{I}_\nu(x) = -\frac{1}{I_{f_0}} \cdot \frac{f'_0(x)}{f_0(x)}
$$

where $I_f = \int (f'(x)/f(x))^2 f(x) dx$. 

On and Off Semiparametric Models – p. 41/50
For example, suppose \( f_0 = \phi \), the standard Normal density given by \( \phi(x) = (2\pi)^{-1/2} \exp(-x^2/2) \). Then

- \( \tilde{I}_\nu(x) = x \)
- \( E(\tilde{I}_\nu^2) = 1 \)
- \( E(\psi - \tilde{I}_\nu)^2 = B^2 = .01609... \)
- \( c = -d_0 = -B = -\sqrt{.01609...} = -.126846 \) yields \( \tilde{a}_0 \) given by

\[
\tilde{a}_0(x) = -d_0 \frac{\psi(x) - \tilde{I}_\nu(x)}{E(\psi - \tilde{I}_\nu)^2} + d_0 \frac{\tilde{I}_\nu(x)}{E(\tilde{I}_\nu^2)} \\
= -(\psi(x) - x) + d_0 x.
\]

\[
AMSE_{\tilde{\nu}}(a_0) = E(\psi^2) + \{E(\psi a_0)\}^2 = E\psi^2 = 1.01609...
\]

\[
AMSE_{\tilde{\nu} \text{ eff}}(a_0) = E(\tilde{I}_\nu^2) + \{E(\tilde{I}_\nu a_0)\}^2 = 1 + d_0^2 = 1.01609....
\]
\( \tilde{a}_0 \), symmetric location model, \( f_0 = \phi \)
• For example, suppose $f_0 = \phi$, the standard Normal density given by $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$. Then

- $\tilde{l}_\nu(x) = x$, $E(\tilde{l}_\nu^2) = 1$, $E(\psi - \tilde{l}_\nu)^2 \equiv B^2 = .01609…$
- If $d^2 > d_0^2 \equiv B^2$, and $c = -d$, then $\tilde{a}_{-d,d}$ is given by

\[
\tilde{a}_{-d,d}(x) = -d \frac{\psi(x) - \tilde{l}_\nu(x)}{E(\psi - \tilde{l}_\nu)^2} + d \frac{\tilde{l}_\nu(x)}{E(\tilde{l}_\nu^2)} = -(\psi(x) - x) + d_0 x.
\]

- $AMSE_{\hat{\nu}}(a_{-d,d}) = E(\hat{\psi}^2) + \{0\}^2 = E(\psi^2) = 1 + d_0^2 = 1.01609…$
- $AMSE_{\hat{\nu}}^{eff}(a_{-d,d}) = E(\tilde{l}_\nu^2) + \{d\}^2 = 1 + d^2 > 1 + d_0^2 = 1.01609…$. 

On and Off Semiparametric Models – p. 44/50
\( \tilde{a}_0 \) (red), \( \tilde{a}_{d,d} \) (blue), \( d^2 = 2B^2 \), symmetric location model, \( f_0 = \phi \)
Example 2. Paired exponential mixture model

- Model:

\[ \mathcal{P} = \{ P_{\nu,G} : \frac{dP_{(\nu,G)}}{d\lambda} = p_{(\nu,G)} : \nu > 0, G \text{ a d.f.} \}, \]

where

\[ p_{(\nu,G)}(x, y) = \int_0^\infty \lambda \exp(-\lambda x) \cdot \nu \lambda \exp(-\nu \lambda y) dG(\lambda) \]
Example 2. Paired exponential mixture model

• Model:

\[ \mathcal{P} = \{ P_{\nu,G} : \frac{dP_{(\nu,G)}}{d\lambda} = p_{(\nu,G)} : \nu > 0, \ G \text{ a d.f.} \}, \]

where

\[ p_{(\nu,G)}(x, y) = \int_{0}^{\infty} \lambda \exp(-\lambda x) \cdot \nu \lambda \exp(-\nu \lambda y) dG(\lambda) \]

• Inefficient estimator: Let \( \hat{\nu}_n \) to be the MLE based on the (pairwise) maximal invariants, \( (X_i/Y_i), i = 1, \ldots, n. \)
Example 2. Paired exponential mixture model

- **Model:**

\[ \mathcal{P} = \{ P_{\nu,G} : \frac{dP_{(\nu,G)}}{d\lambda} = p_{(\nu,G)} : \nu > 0, G \text{ a d.f.} \}, \]

where

\[ p_{(\nu,G)}(x, y) = \int_0^\infty \lambda \exp(-\lambda x) \cdot \nu \lambda \exp(-\nu \lambda y) dG(\lambda) \]

- **Inefficient estimator:** Let \( \hat{\nu}_n \) to be the MLE based on the (pairwise) maximal invariants, \( (X_i/Y_i), i = 1, \ldots, n \).

- **Influence function** \( \psi \): \( \hat{\nu}_n \) is asymptotically linear with influence function \( \psi \) given by

\[ \psi(x, y) = 3\nu \left\{ \frac{x - \nu y}{x + \nu y} \right\} \equiv 3\nu^2 \hat{i}_{\nu}^{\text{inv}}(x/y) \]
\[ E\psi^2(X, Y) = 3\nu^2 \]
• $E\psi^2(X, Y) = 3\nu^2$

• **Efficient score function for $\nu$:**

$$l^*_\nu(x, y) = \frac{1}{2\nu} \left( \frac{x - \nu y}{x + \nu y} \right) \left\{ 2 - 1 - s \frac{p'_S}{p_S}(s) \right\}$$

where $p_S$ is the density of $S \equiv (X + \nu Y)$ given by

$$p_S(s, G) = s \int_{0}^{\infty} \lambda^2 e^{-\lambda s} dG(\lambda).$$
• $E \psi^2(X, Y) = 3\nu^2$

• **Efficient score function for $\nu$:**

$$I^*_\nu(x, y) = \frac{1}{2\nu} \left( \frac{x - \nu y}{x + \nu y} \right) \left\{ 2 - 1 - s \frac{p_S'}{p_S}(s) \right\}$$

where $p_S$ is the density of $S \equiv (X + \nu Y)$ given by

$$p_S(s, G) = s \int_0^\infty \lambda^2 e^{-\lambda s} dG(\lambda).$$

• **Information for $\nu$ in $\mathcal{P}$:**

$$E(I^*_\nu(X, Y)^2) = \frac{1}{3\nu^2} + \frac{1}{12\nu^2} I_{\text{scale}}(p_S).$$
- \(E\psi^2(X, Y) = 3\nu^2\)

- **Efficient score function for \(\nu\):**

\[
1^*_\nu(x, y) = \frac{1}{2\nu} \left( \frac{x - \nu y}{x + \nu y} \right) \left\{ 2 - 1 - s \frac{p'_S}{p_S}(s) \right\}
\]

where \(p_S\) is the density of \(S \equiv (X + \nu Y)\) given by

\[
p_S(s, G) = s \int_0^\infty \lambda^2 e^{-\lambda s} dG(\lambda).
\]

- **Information for \(\nu\) in \(\mathcal{P}\):**

\[
E(1^*_\nu(X, Y)^2) = \frac{1}{3\nu^2} + \frac{1}{12\nu^2} I_{scale}(p_S).
\]

- **Variance bound:**

\[
E(\tilde{1}_\nu)^2 = \left( \frac{1}{3\nu^2} + \frac{1}{12\nu^2} I_{scale}(p_S) \right)^{-1} = 3\nu^2 \left( 1 + \frac{1}{4} I_{scale}(p_S) \right)^{-1}.
\]
Inefficiency of $\hat{\nu}$:

\[
B^2 = E(\psi - \tilde{I}_\nu)^2 = E\psi^2 - E\tilde{I}_\nu^2
\]

\[
= 3\nu^2 \left\{ 1 - \frac{1}{1 + I_{\text{scale}}(p_S)/4} \right\}
\]

\[
= 3\nu^2 \left\{ \frac{I_{\text{scale}}(p_S)/4}{1 + I_{\text{scale}}(p_S)/4} \right\}
\]

\[
\equiv 3\nu^2 \left\{ \frac{b}{1 + b} \right\}.
\]
• **Inefficiency of $\hat{\nu}$:**

\[
B^2 = E(\psi - \tilde{I}_\nu)^2 = E\psi^2 - E\tilde{I}_\nu^2
\]

\[
= 3\nu^2 \left\{ 1 - \frac{1}{1 + I_{\text{scale}}(p_S)/4} \right\}
\]

\[
= 3\nu^2 \left\{ \frac{I_{\text{scale}}(p_S)/4}{1 + I_{\text{scale}}(p_S)/4} \right\}
\]

\[
\equiv 3\nu^2 \left\{ \frac{b}{1 + b} \right\}.
\]

• $\hat{\nu}$ is asymptotically more efficient than $\hat{\nu}^{\text{eff}}$ along $\{Q_n\}$ when $d > d_0 = B$ and $a = a_{-d,d}$:

\[
a_{-d,d}(x, y) = \frac{d}{2\nu} \frac{1 + b}{b} \left( \frac{x - \nu y}{x + \nu y} \right) i_{\text{scale}}(s).
\]