The behavior of one-step estimators in misspecified location and regression models

Priscilla (Cindy) Greenwood
School of Mathematics and Statistics
and
Mathematical and Computational Sciences Modeling Center
Arizona State University

jointly with
Anton Schick (Binghamton University)
and Wolfgang Wefelmeyer (University of Cologne)

Outline.

What do one-step-estimators for semiparametric models estimate when the model is wrong? We look first at parametric models; they show already most features. For comparison we also look at the maximum likelihood estimator.

Parametric models:
- Maximum likelihood estimator.
- One-step estimator.
- One-step estimator as nonparametric estimator.
- One-step estimator when the model is only partially wrong.
- Robust one-step estimator.

Semiparametric model:
- Linear regression.

Maximum likelihood estimator when the model is correct.

Let $X_1, \ldots, X_n$ be observations from a parametric model P_ϑ with density f_ϑ and one-dimensional parameter ϑ (for notational simplicity). The *maximum likelihood estimator* $\hat{\vartheta}$ maximizes $\sum_{j=1}^n \log f_\vartheta(X_j)$. If the model is correct and ϑ is the true parameter, the m.l.e. estimates ϑ, which is the maximum of $P_\vartheta \log f_\tau$ in τ, and has the stochastic expansion

$$\hat{\vartheta} = \vartheta + J_\vartheta^{-1} \frac{1}{n} \sum_{j=1}^n \ell_\vartheta(X_j) + o_p(n^{-1/2}),$$

where $\ell_\vartheta = \partial_\vartheta \log f_\vartheta = \dot{f}_\vartheta / f_\vartheta$ is the score function and $J_\vartheta = P_\vartheta \ell_\vartheta^2$ is the Fisher information for ϑ.

Maximum likelihood estimator when the model is wrong.

Suppose the parametric model P_ϑ is *not* correct, and the true distribution is Q. Then the maximum likelihood estimator estimates the maximizer of $Q \log f_\tau$ in τ, which is the *Kullback–Leibler information* $K(Q)$. The score function at $K(Q)$ is still centered, $Q\ell_{K(Q)} = 0$, and the m.l.e. has the stochastic expansion

$$\hat{\vartheta} = K(Q) + (Q\dot{\ell}_{K(Q)})^{-1}\frac{1}{n}\sum_{j=1}^{n} \ell_{K(Q)}(X_j) + o_p(n^{-1/2}).$$

This is similar to the expansion when the true $Q = P_\vartheta$ is in the model (previous slide):

$$\hat{\vartheta} = \vartheta + J_\vartheta^{-1}\frac{1}{n}\sum_{j=1}^{n} \ell_\vartheta(X_j) + o_p(n^{-1/2}).$$

One-step estimator when the model is correct.

If $\hat{t}$ is a $n^{1/2}$-consistent *initial estimator*, a *one-step* or *Newton–Raphson* estimator is

$$\hat{\vartheta} = \hat{t} + J_{\hat{t}}^{-1} \frac{1}{n} \sum_{j=1}^{n} \ell_{\hat{t}}(X_j)$$

Suppose the model is correct. Then the initial estimator cancels,

$$\hat{\vartheta} = \hat{t} + J_{\vartheta}^{-1} \frac{1}{n} \sum_{j=1}^{n} \ell_{\vartheta}(X_j) + J_{\vartheta}^{-1} \frac{1}{n} \sum_{j=1}^{n} \dot{\ell}_{\vartheta}(X_j)(\hat{t} - \vartheta) + o_p(n^{-1/2}),$$

and

$$\frac{1}{n} \sum_{j=1}^{n} \dot{\ell}_{\vartheta}(X_j) \to P_{\vartheta}\dot{\ell}_{\vartheta} = -P_{\vartheta}\ell_{\vartheta}^2 = -J_{\vartheta}.$$

Hence the one-step estimator is asymptotically equivalent to the maximum likelihood estimator.

One-step estimator when the model is wrong.

Suppose the parametric model P_ϑ is *not* correct, and the true distribution is Q. The initial estimator $\hat{t}$ converges to $t(Q)$, say. But the score function at $t(Q)$ will, in general, *not* be centered, $Q\ell_{t(Q)} \neq 0$. To expand the one-step estimator $\hat{\vartheta} = \hat{t} + J_{\hat{t}}^{-1}\frac{1}{n}\sum_{j=1}^{n} \ell_{\hat{t}}(X_j)$, we must add and subtract $Q\ell_{\hat{t}}$. This creates an additional *bias term*: The one-step estimator estimates not $t(Q)$ but

$$t(Q) + J_{t(Q)}^{-1} Q\ell_{t(Q)}.$$

This also creates two additional *variance terms* from the expansions of $J_{\hat{t}}$ and $Q\ell_{\hat{t}}$ in the stochastic expansion of the one-step estimator:

$$\begin{aligned}\hat{\vartheta} = \hat{t} + J_{t(Q)}^{-1} Q\ell_{t(Q)} + J_{t(Q)}^{-1}\frac{1}{n}\sum_{j=1}^{n}(\ell_{t(Q)}(X_j) - Q\ell_{t(Q)}) \\ + \Big(J_{t(Q)}^{-1} Q\dot{\ell}_{t(Q)} - J_{t(Q)}^{-2}\dot{J}_{t(Q)} Q\ell_{t(Q)}\Big)(\hat{t} - t(Q)) + o_p(n^{-1/2}).\end{aligned}$$

One-step estimator as nonparametric estimator.

The initial estimator $\widehat{t}$ is a nonparametric estimator of some functional $t(Q)$. We note that the one-step estimator also is a nonparametric estimator, of a different functional,

$$t(Q) + J_{t(Q)}^{-1} Q \ell_{t(Q)}.$$

The one-step estimator is obtained by plugging in nonparametric estimators $\widehat{t}(= t(\widehat{Q}))$ and empirical estimators

$$\widehat{Q} = \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j} \quad \text{and} \quad \hat{J} = \frac{1}{n}\sum_{j=1}^{n} \ell_{\widehat{t}}^{2}(X_j).$$

The Hájek–Le Cam theory of efficient estimation tells us that the one-step estimator (if it is regular and asymptotically linear) has an influence function which equals the gradient of the above functional, and is therefore efficient.

One-step estimator when the model is only partially wrong.

Let $X_1, \ldots, X_n$ be real observations with density $f(x - \vartheta)$, where f is a known density and symmetric about 0. Let $\ell = -f'/f$ and $J = \int \ell^2(x)\,dx$. Let $\hat{t}$ be an initial estimator (e.g. the sample mean). The one-step estimator is

$$\hat{\vartheta} = \hat{t} + J^{-1}\frac{1}{n}\sum_{j=1}^{n} \ell(X_j - \hat{t}).$$

Suppose that the density f is not correctly specified, but the true distribution Q is symmetric, about $t(Q)$, say. Then the score function remains centered, $E_Q\ell(X - t(Q)) = 0$. Suppose the initial estimator $\hat{t}$ estimates $t(Q)$. Then the one-step estimator also estimates $t(Q)$ and has a simple stochastic expansion

$$\begin{aligned}\hat{\vartheta} = t(Q) &+ J^{-1}\frac{1}{n}\sum_{j=1}^{n} \ell(X_j - t(Q)) \\ &+ \Big(1 - J^{-1}E_Q\ell'(X - t(Q))\Big)(\hat{t} - t(Q)) + o_p(n^{-1/2}).\end{aligned}$$

Robust one-step estimator.

When the model is correct, the initial estimator cancels out of the one-step estimator. This also happens under misspecification (at least partially) when we choose a "robust" estimator for the variance. Consider again observations $X_1, \dots, X_n$ with density $f(x - \vartheta)$, where f is a known density and symmetric about 0. Let $\ell = -f'/f$ and $J = \int \ell^2(x)\, dx$. Let $\hat{t}$ be an initial estimator (e.g. the sample mean). Replace J by $\hat{J} = \frac{1}{n}\sum_{j=1}^n \ell'(X_j - \hat{t})$ in the one-step estimator:

$$\hat{\vartheta} = \hat{t} + \hat{J}^{-1}\frac{1}{n}\sum_{j=1}^n \ell(X_j - \hat{t}).$$

Then

$$\hat{\vartheta} = t(Q) + J^{-1}\frac{1}{n}\sum_{j=1}^n \ell(X_j - t(Q)) + o_p(n^{-1/2}).$$

Linear regression.

Let (X_j, Y_j) be observations of the linear regression model $Y = \vartheta X + \varepsilon$ with X and ε independent and ε having density f with mean 0. (We take ϑ and X one-dimensional for notational simplicity.) If the model is correct, an efficient estimator $\hat{\vartheta}$ of ϑ is characterized by

$$\hat{\vartheta} = \vartheta + \Lambda^{-1} \frac{1}{n} \sum_{j=1}^{n} g(X_j, \varepsilon_j) + o_p(n^{-1/2})$$

with "efficient score function" $g(X, \varepsilon) = (X - \mu)\ell(\varepsilon) + \sigma^{-2}\mu\varepsilon$ and $\mu = EX$, $\ell = -f'/f$, $\sigma^2 = E\varepsilon^2$, and with variance Λ^{-1}, where

$$\Lambda = Eg^2(X, \varepsilon) = J(M - \mu^2) + \sigma^{-2}\mu^2$$

with $J = E\ell^2(\varepsilon)$ and $M = EX^2$. The one-step estimator replaces ϑ by an initial estimator $\hat{t}$ (e.g. the least squares estimator), the density f by a kernel estimator based on residuals $\hat{\varepsilon}_j = Y_j - \hat{t}X_j$, and the expectations empirically.

One-step estimator for linear regression.

The one-step estimator replaces ϑ by an initial estimator $\hat{t}$ (e.g. the least squares estimator), the density f by a kernel estimator $\hat{f}$ based on residuals $\hat{\varepsilon}_j = Y_j - \hat{t}X_j$, and the expectations empirically:

$$\hat{\vartheta} = \hat{t} + \hat{L}^{-1}\frac{1}{n}\sum_{j=1}^{n}\hat{g}(X_j, \hat{\varepsilon}_j)$$

with $\hat{g}(X,\varepsilon) = (X - \hat{\mu})\hat{\ell}(\varepsilon) + \hat{\sigma}^{-2}\hat{\mu}\varepsilon$ and $\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j$, $\hat{\ell} = -\hat{f}'/\hat{f}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}\hat{\varepsilon}_j^2$, and with

$$\hat{L} = \hat{J}(\hat{M} - \hat{\mu}^2) + \hat{\sigma}^{-2}\hat{\mu}^2$$

with $\hat{J} = \frac{1}{n}\sum_{j=1}^{n}\hat{\ell}^2(\hat{\varepsilon}_j)$ and $\hat{M} = \frac{1}{n}\sum_{j=1}^{n} X_j^2$.

Misspecified linear regression.

Suppose the linear model is not correctly specified. Let Q be the true distribution of (X, Y). Suppose the initial estimator $\hat{t}$ estimates $t(Q)$. Independence of X and ε might be lost under Q. Write $s(X, y)$ for the conditional density of ε given X. The conditional density of $\varepsilon_Q = Y - t(Q)X$ is $s(X, y + t(Q)X)$, and the unconditional density is $f_Q(y) = E_Q s(X, y + t(Q)X)$. Set $\ell_Q = -f'_Q/f_Q$. Then

$$\frac{1}{n}\sum_{j=1}^{n}(X_j - \hat{\mu})\hat{\ell}(\hat{\varepsilon}_j) \to E_Q(X - \mu_Q)\ell_Q(Y - t(Q)X) = I_Q,$$

$$\hat{J} = \frac{1}{n}\sum_{j=1}^{n}\hat{\ell}^2(\hat{\varepsilon}_j) \to E_Q\ell_Q^2(Y - t(Q)X) = J_Q.$$

Hence the one-step estimator estimates $t(Q) + L_Q^{-1}(I_Q + \sigma_Q^{-2}\mu_Q E\varepsilon_Q)$ with $L_Q = J_Q(M_Q - \mu_Q^2) + \sigma_Q^{-2}\mu_Q^2$, where $\mu_Q = E_Q X$, $M_Q = E_Q X^2$ and $\sigma_Q^2 = E_Q\varepsilon_Q^2$.