Efficient Estimation In Constrained Models

Hanxiang Peng

Department of Math Sciences, IUPUI

This talk is based on a joint work with Anton Schick, Binghamton University.

October 18, 2008

Outline

The Problem

- Let $(X, Y) \sim Q$, $X \sim F$ and $Y \sim G$. Assume F and G are continuous. We want to estimate

$$Qh = \mathbb{E}h(X, Y) = \int h \, dQ$$

for some Q-square integrable h based on n independent observations: $(X_1, Y_1), ..., (X_n, Y_n)$.

The Problem

- Let $(X, Y) \sim Q$, $X \sim F$ and $Y \sim G$. Assume F and G are continuous. We want to estimate

$$Qh = \mathbb{E}h(X, Y) = \int h\, dQ$$

for some Q-square integrable h based on n independent observations: $(X_1, Y_1), ..., (X_n, Y_n)$.

- Examples:

$$P(X > Y), \qquad P(X \le s, Y \le t)$$

$$P(\min(X, Y) > s), \qquad P(\max(X, Y) < t).$$

The Problem

- The usual estimator of Qh is the **empirical estimator**:

$$\widehat{Qh} = \frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j)$$

It is **efficient** if no additional knowledge is available.

The Problem

- The usual estimator of Qh is the **empirical estimator**:

$$\widehat{Qh} = \frac{1}{n} \sum_{j=1}^{n} h(X_j, Y_j)$$

It is **efficient** if no additional knowledge is available.
- Better estimators exist if we know more about Q.

Additional knowledge: Symmetry

- **Exchangeability**: $(X, Y) \stackrel{d}{=} (Y, X)$

$$\frac{1}{2n}\sum_{j=1}^{n} h(X_j, Y_j) + h(Y_j, X_j)$$

Additional knowledge: Symmetry

- **Exchangeability**: $(X, Y) \overset{d}{=} (Y, X)$

$$\frac{1}{2n} \sum_{j=1}^{n} h(X_j, Y_j) + h(Y_j, X_j)$$

- **Central Symmetry**: $-(X, Y) \overset{d}{=} (X, Y)$

$$\frac{1}{2n} \sum_{j=1}^{n} h(X_j, Y_j) + h(-X_j, -Y_j)$$

Additional knowledge: Dependence

- **Independence**: $X \perp\!\!\!\perp Y$.

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i, Y_j)$$

Additional knowledge: Dependence

- **Independence**: $X \perp\!\!\!\perp Y$.

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i, Y_j)$$

- **IID**: $X \perp\!\!\!\perp Y$ and $X \overset{d}{=} Y$.

$$\frac{1}{2n(2n-1)}\sum_{1\le i\ne j\le n} h(Z_i, Z_j)$$

where $Z_i = X_i$ and $Z_{n+i} = Y_i$ for $i = 1, ..., n$.

Additional knowledge: Dependence

- **Uncorrelated**: $\mathbb{C}\mathrm{ov}(X, Y) = 0$. Consider unbiased estimators

$$\hat{H}(c) = \frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j) - c\hat{C}, \quad c \in \mathbb{R}$$

where $\hat{C} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})(Y_j - \bar{Y})$ is the sample covariance.

Additional knowledge: Dependence

- **Uncorrelated**: $\mathbb{C}\mathrm{ov}(X, Y) = 0$. Consider unbiased estimators

$$\hat{H}(c) = \frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j) - c\hat{C}, \quad c \in \mathbb{R}$$

  where $\hat{C} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})(Y_j - \bar{Y})$ is the sample covariance.
- With $\mu = \mathbb{E}X$ and $\nu = \mathbb{E}Y$,

$$\begin{aligned} \hat{C} &= \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)(Y_j - \nu) - (\bar{X} - \mu)(\bar{Y} - \nu) \\ &= \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)(Y_j - \nu) + O_p(n^{-1}). \end{aligned}$$

- Thus the asymptotic variance of $\hat{H}(c)$ is

$$\sigma^2(c) = \mathbb{V}\mathrm{ar}(h(X,Y)) - 2cA + c^2 B,$$

where

$$A = \mathbb{C}\mathrm{ov}[h(X,Y), (X-\mu)(Y-\nu)], \; B = \mathbb{E}(X-\mu)^2(Y-\nu)^2$$

- Thus the asymptotic variance of $\hat{H}(c)$ is

$$\sigma^2(c) = \mathbb{V}\mathrm{ar}(h(X,Y)) - 2cA + c^2 B,$$

where

$$A = \mathbb{C}\mathrm{ov}[h(X,Y), (X-\mu)(Y-\nu)], \; B = \mathbb{E}(X-\mu)^2(Y-\nu)^2$$

- The asymptotic variance $\sigma^2(c)$ is minimized at $c = c_* = A/B$, suggesting a plut-in estimator:

$$\hat{H} = \hat{H}(\hat{c}) = \frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j) - \hat{c}\hat{C}$$

where

$$\hat{c} = \frac{\hat{A}}{\hat{B}} = \frac{\frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j)(X_j - \bar{X})(Y_j - \bar{Y})}{\frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2(Y_j - \bar{Y})^2}$$

- Easy to show

$$\sqrt{n}(\hat{H}(\hat{c}) - \hat{H}(c_*)) = o_p(1).$$

Thus $\hat{H}$ is asymptotically equivalent to the best estimator among $\hat{H}(c) : c \in \mathbb{R}$. Consequently it will be at least as good as the empirical estimator and **better** if $c_* \neq 0$.

- Easy to show

$$\sqrt{n}(\hat{H}(\hat{c}) - \hat{H}(c_*)) = o_p(1).$$

Thus $\hat{H}$ is asymptotically equivalent to the best estimator among $\hat{H}(c) : c \in \mathbb{R}$. Consequently it will be at least as good as the empirical estimator and **better** if $c_* \neq 0$.

- The **variance reduction** is

$$\frac{A^2}{B}.$$

- Easy to show

$$\sqrt{n}(\hat{H}(\hat{c}) - \hat{H}(c_*)) = o_p(1).$$

Thus $\hat{H}$ is asymptotically equivalent to the best estimator among $\hat{H}(c) : c \in \mathbb{R}$. Consequently it will be at least as good as the empirical estimator and **better** if $c_* \neq 0$.

- The **variance reduction** is

$$\frac{A^2}{B}.$$

- Indeed one can show that $\hat{H}$ is **efficient**.

Additional knowledge: Marginals

- **Equal marginal means**: $\mathbb{E}X = \mathbb{E}Y$.

$$\frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j) - \hat{c}(\bar{X} - \bar{Y})$$

where

$$\hat{c} = \frac{\frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j)(X_j - Y_j)}{\frac{1}{n}\sum_{j=1}^{n}(X_j - Y_j)^2}$$

One can show this estimator is **efficient**.

Additional knowledge: Marginals

- **Equal marginal means**: $\mathbb{E}X = \mathbb{E}Y$.

$$\frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j) - \hat{c}(\bar{X} - \bar{Y})$$

where

$$\hat{c} = \frac{\frac{1}{n}\sum_{j=1}^{n} h(X_j, Y_j)(X_j - Y_j)}{\frac{1}{n}\sum_{j=1}^{n}(X_j - Y_j)^2}$$

One can show this estimator is **efficient**.

- **Known Marginals**: Bickel, Ritov & Wellner (1991): $F = F_0, G = G_0$. Partition plane into rectangles:

$$C_{i,j} = A_i \times B_j : \quad i = 1, ..., I, j = 1, ..., J.$$

Let $N_{i,j}$ be # of observations falling into $C_{i,j}$:

$$N_{i,j} = \sum_{j=1}^{n} 1[X_j \in A_i, Y_j \in B_j]$$

Additional Knowledge: Known Marginals

- and $\hat{h}_{i,j}$ be the average of h over cell $C_{i,j}$:

$$\bar{h}_{i,j} = \frac{1}{N_{i,j}} \sum_{l=1}^{n} \mathbf{1}[X_l \in A_i, Y_l \in B_j] h(X_l, Y_l)$$

Additional Knowledge: Known Marginals

- and $\hat{h}_{i,j}$ be the average of h over cell $C_{i,j}$:

$$\bar{h}_{i,j} = \frac{1}{N_{i,j}} \sum_{l=1}^{n} \mathbf{1}[X_l \in A_i, Y_l \in B_j] h(X_l, Y_l)$$

- BRW propose the esttimator

$$\sum_{i=1}^{I} \sum_{j=1}^{J} \hat{p}_{i,j} \bar{h}_{i,j}$$

where $\hat{p}_{i,j}$ are chosen to minimize the χ^2:

$$\sum_{i,j} \frac{(N_{i,j} - np_{i,j})^2}{N_{i,j}}$$

subject to constraints:

J I

Additional Knowledge: Known Marginals

- subject to constraints:

$$p_{i\cdot} = \sum_{j=1}^{J} p_{i,j} = F_0(X \in A_i), \quad p_{\cdot j} = \sum_{i=1}^{I} p_{i,j} = G_0(X \in B_j)$$

Additional Knowledge: Known Marginals

- subject to constraints:

$$p_{i\cdot} = \sum_{j=1}^{J} p_{i,j} = F_0(X \in A_i), \quad p_{\cdot j} = \sum_{i=1}^{I} p_{i,j} = G_0(X \in B_j)$$

- BRW show that under mild assumptions on the dependence structure and choice of partitions (nested, $I, J \to \infty$):

$$\sum_{i,j} \hat{p}_{i,j} \bar{h}_{i,j} = \frac{1}{n} \sum_{l=1}^{n} h(X_l, Y_l) - a_*(X_l) - b_*(Y_l) + o_p(n^{-1/2})$$

where a_* and b_* minimize

$$\mathbb{E}(h(X, Y) - a(X) - b(Y))^2, \quad a \in L_{2,0}(F_0), b \in L_{2,0}(G_0).$$

Additional Knowledge: Known Marginals

- subject to constraints:

$$p_{i\cdot} = \sum_{j=1}^{J} p_{i,j} = F_0(X \in A_i), \quad p_{\cdot j} = \sum_{i=1}^{I} p_{i,j} = G_0(X \in B_j)$$

- BRW show that under mild assumptions on the dependence structure and choice of partitions (nested, $I, J \to \infty$):

$$\sum_{i,j} \hat{p}_{i,j} \bar{h}_{i,j} = \frac{1}{n} \sum_{l=1}^{n} h(X_l, Y_l) - a_*(X_l) - b_*(Y_l) + o_p(n^{-1/2})$$

  where a_* and b_* minimize

$$\mathbb{E}(h(X, Y) - a(X) - b(Y))^2, \quad a \in L_{2,0}(F_0), b \in L_{2,0}(G_0).$$

- Further, they show their estimator is **efficient**.

Alternative Estimate: Known Marginals

- Peng and Schick (2002): For $a \in L_{2,0}(F_0)$, $b \in L_{2,0}(G)$,

$$\hat{H}(a, b) = \frac{1}{n} \sum_{l=1}^{n} h(X_l, Y_l) - a(X_l) - b(Y_l)$$

is an unbiased estimator with possible smallest variance for $(a, b) = (a_*, b_*)$.

Alternative Estimate: Known Marginals

- Peng and Schick (2002): For $a \in L_{2,0}(F_0)$, $b \in L_{2,0}(G)$,

$$\hat{H}(a,b) = \frac{1}{n}\sum_{l=1}^{n} h(X_l, Y_l) - a(X_l) - b(Y_l)$$

  is an unbiased estimator with possible smallest variance for $(a,b) = (a_*, b_*)$.

- This motivates to use the estimator

$$\frac{1}{n}\sum_{l=1}^{n}\left[h(X_l, Y_l) - \sum_{i=1}^{I}\hat{\alpha}_i v_i(X_l) - \sum_{j=1}^{J}\hat{\beta}_j w_j(Y_l)\right]$$

  where $\hat{\alpha}_i$ and $\hat{\beta}_j$ minimize

$$\frac{1}{n}\sum_{l=1}^{n}\left[h(X_l, Y_l) - \sum_{i=1}^{I}\alpha_i v_i(X_l) - \sum_{j=1}^{J}\beta_j w_j(Y_l)\right]^2$$

Alternative estimates: Known Marginals

- where $v_1, v_2, \ldots$ is an ONS for $L_{2,0}(F_0)$ and $w_1, w_2, \ldots$ an ONS for $L_{2,0}(G_0)$.

Alternative estimates: Known Marginals

- where $v_1, v_2, \ldots$ is an ONS for $L_{2,0}(F_0)$ and $w_1, w_2, \ldots$ an ONS for $L_{2,0}(G_0)$.
- PS ('02) show their estimator is also **asymptotically efficient** under mild assumptions on the ONS, $I, J \to \infty$ at certain rates and the dependence conditions of BRW.

Alternative estimates: Known Marginals

- where $v_1, v_2, \dots$ is an ONS for $L_{2,0}(F_0)$ and $w_1, w_2, \dots$ an ONS for $L_{2,0}(G_0)$.
- PS ('02) show their estimator is also **asymptotically efficient** under mild assumptions on the ONS, $I, J \to \infty$ at certain rates and the dependence conditions of BRW.
- One can take the usual trigo-bases:

$$v_k(x) = u_k(F_0(x)), \quad w_k(y) = u_k(G_0(y)),$$

where $u_k(x) = \sqrt{2}\cos(i\pi x), \quad 0 \le x \le 1, k = 1, 2, \dots$

Additional Knowledge: Equal Marginals

- Peng and Schick (2005): $\mathcal{L}(X) = \mathcal{L}(Y)$, so for $a \in L_{2,0}(F)$, we have $\mathbb{E}[a(X) - a(Y)] = 0$. Thus for each such a,

$$\hat{H}(a) = \frac{1}{n} \sum_{l=1}^{n} h(X_l, Y_l) - a(X_l) + a(Y_l)$$

is an unbiased estimator with smallest possible variance for $a = a_*$, which minimizes

$$\mathbb{E}[h(X, Y) - a(X) + a(Y)]^2.$$

Additional Knowledge: Equal Marginals

- Peng and Schick (2005): $\mathcal{L}(X) = \mathcal{L}(Y)$, so for $a \in L_{2,0}(F)$, we have $\mathbb{E}[a(X) - a(Y)] = 0$. Thus for each such a,

$$\hat{H}(a) = \frac{1}{n} \sum_{l=1}^{n} h(X_l, Y_l) - a(X_l) + a(Y_l)$$

  is an unbiased estimator with smallest possible variance for $a = a_*$, which minimizes

$$\mathbb{E}[h(X, Y) - a(X) + a(Y)]^2.$$

- This motivates to use the estimator

$$\frac{1}{n} \sum_{l=1}^{n} \left(h(X_l, Y_l) - \sum_{i=1}^{I} \hat{\alpha}_i [u_i(\hat{F}(X_l)) - u_i(\hat{F}(Y_l))] \right)$$

Additional Knowledge: Equal Marginals

- where $\hat{\alpha}_i$ is the unique minizer to

$$\frac{1}{n}\sum_{l=1}^{n}\left(h(X_l, Y_l) - \sum_{i=1}^{I}\alpha_i[u_i(\hat{F}(X_l)) - u_i(\hat{F}(Y_l))]\right)^2$$

Here $\hat{F}$ is the pooled empirical distribution function:

$$\hat{F}(x) = \frac{1}{2n}\sum_{l=1}^{n}\mathbf{1}[X_l \leq x] + \mathbf{1}[Y_l \leq x].$$

Additional Knowledge: Equal Marginals

- where $\hat{\alpha}_i$ is the unique minizer to

$$\frac{1}{n}\sum_{l=1}^{n}\left(h(X_l, Y_l) - \sum_{i=1}^{I}\alpha_i[u_i(\hat{F}(X_l)) - u_i(\hat{F}(Y_l))]\right)^2$$

Here $\hat{F}$ is the pooled empirical distribution function:

$$\hat{F}(x) = \frac{1}{2n}\sum_{l=1}^{n}\mathbf{1}[X_l \leq x] + \mathbf{1}[Y_l \leq x].$$

- PS ('05) show that their estimator is asymptotically equivalent to the best unbiased estimator $\hat{H}(a_*)$ and is **efficient** as $I \to \infty$ slowly with n and under the depedence structure of BRW.

Additional Knowledge: Parametric Marginals

- Peng and Schick (2004): assume $F = F_{\vartheta_1}$ and $G = G_{\vartheta_2}$ for unknown parameters ϑ_1 and ϑ_2. Suppose we can estimate ϑ_1 and ϑ_2 by $\hat{\vartheta}_1$ and $\hat{\vartheta}_2$ at the square root rate. Then an estimator of Qh is

$$\frac{1}{n}\sum_{l=1}^{n} h(X_l, Y_l) - \sum_{i=1}^{I} \hat{\alpha}_i v_i(F_{\hat{\vartheta}_1}(X_l)) - \sum_{j=1}^{J} \hat{\beta}_j w_j(G_{\hat{\vartheta}_2}(Y_l))$$

where $\hat{\alpha}_i$ and $\hat{\beta}_j$ minimize

$$\frac{1}{n}\sum_{l=1}^{n}\Big(h(X_l, Y_l) - \sum_{i=1}^{I} \alpha_i v_i(F_{\hat{\vartheta}_1}(X_l)) - \sum_{j=1}^{J} \beta_j w_j G_{\hat{\vartheta}_2}(Y_l))\Big)^2.$$

Additional Knowledge: Parametric Marginals

- Peng and Schick ('04) show that their estimator is asymptotically equivalent to

$$\frac{1}{n}\sum_{l=1}^{n} h(X_l, Y_l) - a_*(X_l) - b_*(Y_l) + D_1(\hat{\vartheta}_1 - \vartheta_1) + D_2(\hat{\vartheta}_2 - \vartheta_2)$$

for some matrices D_1, D_2 under mild regularity conditions.

Additional Knowledge: Parametric Marginals

- Peng and Schick ('04) show that their estimator is asymptotically equivalent to

$$\frac{1}{n}\sum_{l=1}^{n} h(X_l, Y_l) - a_*(X_l) - b_*(Y_l) + D_1(\hat{\vartheta}_1 - \vartheta_1) + D_2(\hat{\vartheta}_2 - \vartheta_2)$$

  for some matrices D_1, D_2 under mild regularity conditions.
- By the general plug-in principle (Klasseen and Putter (2005)) this estimator will be **efficient** if $\hat{\vartheta}_1, \hat{\vartheta}_2$ are **efficient**.

Estimation of Maringal Parameters

- PS ('08): Suppose bivariate Q has two smooth marginals $X \sim F_\alpha$ with score $\dot{\kappa}_1(\cdot, \alpha)$ and $Y \sim G_\beta$ with score $\dot{\kappa}_2(\cdot, \beta)$.

Estimation of Maringal Parameters

- PS ('08): Suppose bivariate Q has two smooth marginals $X \sim F_\alpha$ with score $\dot\kappa_1(\cdot, \alpha)$ and $Y \sim G_\beta$ with score $\dot\kappa_2(\cdot, \beta)$.
- Efficient estimate $\hat\alpha_n$ of α_0 based only the X observations may not be efficient in the bivariate model because the information from Y is not used. Similarly for $\hat\beta_n$ of β_0.

Estimation of Maringal Parameters

- PS ('08): Suppose bivariate Q has two smooth marginals $X \sim F_\alpha$ with score $\dot\kappa_1(\cdot,\alpha)$ and $Y \sim G_\beta$ with score $\dot\kappa_2(\cdot,\beta)$.
- Efficient estimate $\hat\alpha_n$ of α_0 based only the X observations may not be efficient in the bivariate model because the information from Y is not used. Similarly for $\hat\beta_n$ of β_0.
- Choose $W(y,\beta_0) \in L_{2,0}(G_{\beta_0})$ and $W \perp \dot\kappa_2$. Then

$$\hat\alpha_n(D) = \hat\alpha_n - \frac{1}{n}\sum_{j=1}^{n} DW(Y_j,\hat\beta_n)$$

is an unbiased estimator of α_0 for any matrix D with possible smallest variance

$$\Psi(D) = J_1^{-1} - DD^\top$$

where J_1 is the X-marginal information.

Estimation of Maringal Parameters

- The dispersion matrix is minimized at

$$D = D_* = \mathbb{E}[\psi_1(X, \alpha_0) W^\top (Y, \beta_0)]$$

where $\psi_1(x, \alpha) = J_1^{-1} \dot{\kappa}_1(x, \alpha)$.

- The gain in efficiency is $D_* D_*^\top$.

Estimation of Maringal Parameters

- The dispersion matrix is minimized at

$$D = D_* = \mathbb{E}[\psi_1(X, \alpha_0) W^\top (Y, \beta_0)]$$

where $\psi_1(x, \alpha) = J_1^{-1} \dot{\kappa}_1(x, \alpha)$.

- The gain in efficiency is $D_* D_*^\top$.
- Since D_* is unknown, we estimate by

$$\hat{D}_* = \frac{1}{n} \sum_{j=1}^{n} \psi_1(X_j, \hat{\alpha}_n) W^\top (Y_j, \hat{\beta}_n).$$

Estimation of Maringal Parameters

- The dispersion matrix is minimized at

$$D = D_* = \mathbb{E}[\psi_1(X, \alpha_0) W^\top (Y, \beta_0)]$$

where $\psi_1(x, \alpha) = J_1^{-1} \dot{\kappa}_1(x, \alpha)$.
- The gain in efficiency is $D_* D_*^\top$.
- Since D_* is unknown, we estimate by

$$\hat{D}_* = \frac{1}{n} \sum_{j=1}^{n} \psi_1(X_j, \hat{\alpha}_n) W^\top (Y_j, \hat{\beta}_n).$$

- Our proposed estimate of α is

$$\hat{\alpha}_n^* = \hat{\alpha}_n - \hat{D}_* \frac{1}{n} \sum_{j=1}^{n} W(Y_j, \hat{\beta}_n).$$

Examples

- Choose Q the Farlie-Gumbel-Morgenstern copula density,

  $$q(x,y) = [1+\rho u(x-\alpha)v(y-\beta)]f(x-\alpha)g(y-\beta), \quad x,y \in \mathbb{R},$$

  where f, g are the densities of the location models with finite Fisher informations, and $\rho \in [-1,1]$.
- For standard normal marginals, we have

  $$\mathrm{ARE} = 1 - \rho^2 \frac{2}{\pi}\Big(\int_0^\infty \mathrm{x} \exp(-\mathrm{x}^2/2)\, \mathrm{dx}\Big)^2 = 1 - \frac{2\rho^2}{\pi}.$$

  It can be as large as 36 perent.

Examples

- Choose Q the Farlie-Gumbel-Morgenstern copula density,

  $$q(x,y) = [1+\rho u(x-\alpha)v(y-\beta)]f(x-\alpha)g(y-\beta), \quad x,y \in \mathbb{R},$$

  where f, g are the densities of the location models with finite Fisher informations, and $\rho \in [-1,1]$.
- For standard normal marginals, we have

  $$\mathrm{ARE} = 1 - \rho^2 \frac{2}{\pi}\Big(\int_0^\infty \mathrm{x}\exp(-\mathrm{x}^2/2)\,\mathrm{dx}\Big)^2 = 1 - \frac{2\rho^2}{\pi}.$$

  It can be as large as 36 perent.
- For the double exponential, we have

  $$\mathrm{ARE} = 1 - \rho^2\Big(\int_0^\infty \mathrm{e}^{-\mathrm{x}}\,\mathrm{dx}\Big)^2 = 1 - \rho^2.$$

  It varies from 0 to 1.

MELE: Finitely Many Constraints: $m < \infty$

- Put $\Psi_m(z,\vartheta) = (\psi_1(z,\vartheta), \dots, \psi_m(z,\vartheta))^\top$. Then

$$\int \Psi_m(z,\vartheta)\, dQ(z) = 0$$

MELE: Finitely Many Constraints: $m < \infty$

- Put $\Psi_m(z,\vartheta) = (\psi_1(z,\vartheta), \dots, \psi_m(z,\vartheta))^\top$. Then

$$\int \Psi_m(z,\vartheta)\, dQ(z) = 0$$

- Fix integer m, Qin and Lawless (1994) showed that the MELE $\hat{\vartheta}_n$ of ϑ_0 is *asymptotically efficient*:

$$\hat{\vartheta}_n = \arg\max_{\vartheta} \prod_{j=1}^{n} \omega_j(\vartheta), \quad \omega_j(\vartheta) = \frac{1}{n} \frac{1}{1 + \zeta(\vartheta)^\top \Psi_m(Z_j,\vartheta)}$$

where $\zeta = \zeta(\vartheta)$ is the solution to:

$$\frac{1}{n} \sum_{j=1}^{n} \frac{\Psi_m(Z_j,\vartheta)}{1 + \zeta^\top \Psi_m(Z_j,\vartheta)} = 0.$$

Q with parametric marginals as Infinitely Many Constraints

$$\mathcal{Q} = \left\{ Q << \mu : \int u \, dQ = \int u \, dF_\alpha, u \in \mathcal{U}, \int v \, dQ = \int v \, dG_\beta, v \in \mathcal{V} \right\}$$

where $\mathcal{U} = \{u_k \circ F_\alpha : k = 0, 1, 2, ...\}$ and $\mathcal{V} = \{u_k \circ G_\beta : k = 0, 1, 2, ...\}$.

MELE: Infinitely Many Constraints: $m = \infty$

Theorem 1 Under regularity condistions,

$$\hat{\vartheta}_n = \vartheta_0 + \frac{1}{n}\sum_{i=1}^{n} J_{\#}^{-1}\dot{\vartheta}_{\#}(Z_i, \vartheta_0) + o_P(n^{-1/2})$$

provided $m = m_n$ tends *slowly* to infinity with the sample size n. Here $\dot{\vartheta}_{\#} = \Pi(\dot{\vartheta}_{\natural}|\mathcal{V})$ and $J_{\#} = \mathbb{E}\dot{\vartheta}_{\#}(Z_1, \vartheta_0)^{\otimes 2}$. Hence

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \Rightarrow \mathcal{N}(0, J_{\#}^{-1}).$$

THANKS