An Empirical Likelihood Approach of The Estimation of Linear Functionals with Side Information

Shan Wang

Department of Mathematical Sciences
Purdue School of Science
Indiana University-Purdue University at Indianapolis

July 30, 2013
Outline

Introduction

Review of Empirical Likelihood Approach

Construction

Efficiency

Proofs

Conclusion
Problem we are interested in

Let

\[ Z_1, \ldots, Z_n \sim i.i.d. \quad Z \]
Problem we are interested in

- Let

\[ Z_1, \ldots, Z_n \sim i.i.d. \quad Z \]

- We are interested in estimating the expectation

\[ \theta = E(\psi(Z)) = \int \psi dP \]

for some known square-integrable function \( \psi \) from \( \mathcal{X} \) into \( \mathbb{R}^d \) with side information expressed by
Problem we are interested in

- Let

\[ Z_1, \ldots, Z_n \sim i.i.d. \quad Z \]

- We are interested in estimating the expectation

\[ \theta = E(\psi(Z)) = \int \psi dP \]

for some known square-integrable function \( \psi \) from \( \mathcal{Z} \) into \( \mathbb{R}^d \) with side information expressed by

\[ E[u_k(Z)] = 0, \quad k = 1, \ldots, m \]
Problem we are interested in

- Let

\[ Z_1, \ldots, Z_n \sim i.i.d. \quad Z \]

- We are interested in estimating the expectation

\[ \theta = E(\psi(Z)) = \int \psi dP \]

for some known square-integrable function \( \psi \) from \( \mathcal{Z} \) into \( \mathbb{R}^d \) with side information expressed by

\[ E[u_k(Z)] = 0, \quad k = 1, \ldots, m \]

where \( u_1, \ldots, u_m \) are measurable functions from \( \mathcal{Z} \) to \( \mathbb{R} \) such that

\[ \text{Var}(u_k(Z)) < \infty. \]
Examples

Example 1.1
Estimation of Distribution function: Let
\[ \psi_{x,y}(X, Y) = 1[X \leq x, Y \leq y] \]
for known \( x, y \), with side information e.g. known marginal distribution of \( X \). Then \( \theta = E(\psi(X, Y)) = F(x, y) \) is the distribution function.
Examples

Example 1.1
Estimation of Distribution function: Let

$$\psi_{x,y}(X, Y) = 1[X \leq x, Y \leq y]$$

for known $x, y$, with side information e.g. known marginal distribution of $X$. Then $\theta = E(\psi(X, Y) = F(x, y)$ is the distribution function.
Example 1.3
Example 1.3

- Estimation of depth function:

Let $\psi(x) = x - \|x - \mu\|$, then $D(x, F) = 1 - \|E[\psi(x - \mu)]\|$ is called the depth function, and $D(x, F)$ is the depth value of $x$ w.r.t distribution $F$.

With side information, for instance, $X$ is spherical symmetric about some center $\mu$, we can estimate the spatial outlyingness function $O(x, F) = E[\psi(x - \mu)]$, then get the plug-in estimator of $D(x, F)$. 

Wang
Example 1.3

- Estimation of depth function:
- Let $\psi_x(X) = \frac{x - X}{\|x - X\|}$, then

$$D(x, F) = 1 - \|E\psi(x - X)\|$$

is called the depth function, and $D(x, F)$ is the depth value of $x$ w.r.t distribution $F$. 
Example 1.3

- Estimation of depth function:
  - Let $\psi_x(X) = \frac{x - X}{\|x - X\|}$, then

  $$D(x, F) = 1 - \|E\psi(x - X)\|$$

  is called the depth function, and $D(x, F)$ is the depth value of $x$ w.r.t distribution $F$.

- With side information, for instance, $X$ is spherical symmetric about some center $\mu$, we can estimate the spatial outlyingness function $O(x, F) = E(\psi(x - X))$, then get the plug-in estimator of $D(x, F)$. 
Method

▶ The approach we are using is Empirical Likelihood introduced by Owen(1988, 1990, 2001).
Method

- The approach we are using is Empirical Likelihood introduced by Owen (1988, 1990, 2001).

- Incorporating side information:
  The usual empirical estimation
  \[
  \theta = \frac{1}{n} \sum_{j=1}^{n} \psi(Z_i),
  \]
  does not use the side information.
Method

- The approach we are using is Empirical Likelihood introduced by Owen (1988, 1990, 2001).
- Incorporating side information:
  The usual empirical estimation

$$\theta = \frac{1}{n} \sum_{j=1}^{n} \psi(Z_i),$$

does not use the side information.

- Improved estimation is in the sense of least dispersed regular estimates.
Nonparametric (Empirical) Likelihood Approach

Let $X_1, ..., X_n \sim \text{iid distribution function (DF)} ~ F_0$. Let $F$ be an arbitrary DF.
Nonparametric (Empirical) Likelihood Approach

- Let $X_1, \ldots, X_n \sim \text{iid distribution function (DF)} \ F_0$. Let $F$ be an arbitrary DF.

- Nonparametric likelihood of $F$:

$$
\mathcal{L}_n(F) = \prod_{i=1}^n \{F(X_i) - F(X_i^-)\}.
$$
Nonparametric (Empirical) Likelihood Approach

- Let $X_1, \ldots, X_n \sim \text{iid distribution function (DF)} \ F_0$. Let $F$ be an arbitrary DF.
- Nonparametric likelihood of $F$:
  \[
  \mathcal{L}_n(F) = \prod_{i=1}^{n} \{F(X_i) - F(X_i^-)\}.
  \]
- Nonparametric Likelihood Ratio Function:
  \[
  \mathcal{R}_n(\mu) = \frac{\max \{\mathcal{L}_n(F) : E_F(X) = \mu\}}{\max \{\mathcal{L}_n(F) : F\}}.
  \]
Nonparametric (Empirical) Likelihood Approach

- Let $X_1, \ldots, X_n \sim \text{iid distribution function (DF) } F_0$. Let $F$ be an arbitrary DF.
- Nonparametric likelihood of $F$:
  \[ L_n(F) = \prod_{i=1}^{n} \{F(X_i) - F(X_i^-)\}. \]
- Nonparametric Likelihood Ratio Function:
  \[ R_n(\mu) = \frac{\max \{L_n(F) : E_F(X) = \mu\}}{\max \{L_n(F) : F\}}. \]

- Note $L_n(F) = 0$ if $F$ is continuous. Choose $F$ to be supported on \{X_1, \ldots, X_n\} and place point mass $\pi_i$ at $X_i$. Then
  \[ \pi \in \mathcal{P}_n, \quad E_F(X) = \pi_1X_1 + \ldots + \pi_nX_n. \]
  where $\mathcal{P}_n = \{\pi = (\pi_1, \ldots, \pi_n) : \pi_i \geq 0, \pi_1 + \ldots + \pi_n = 1\}$. 
Nonparametric (Empirical) Likelihood Approach

- Let $X_1, \ldots, X_n \sim \text{iid distribution function (DF)} F_0$. Let $F$ be an arbitrary DF.
- Nonparametric likelihood of $F$:
  \[
  L_n(F) = \prod_{i=1}^{n} \{F(X_i) - F(X_i^-)\}.
  \]
- Nonparametric Likelihood Ratio Function:
  \[
  R_n(\mu) = \frac{\max \{L_n(F) : E_F(X) = \mu\}}{\max \{L_n(F) : F\}}.
  \]
- Note $L_n(F) = 0$ if $F$ is continuous. Choose $F$ to be supported on \{\$X_1, \ldots, X_n\}$ and place point mass $\pi_i$ at $X_i$. Then
  \[
  \pi \in \mathcal{P}_n, \quad E_F(X) = \pi_1 X_1 + \ldots + \pi_n X_n.
  \]
  where $\mathcal{P}_n = \{\pi = (\pi_1, \ldots, \pi_n) : \pi_i \geq 0, \pi_1 + \ldots + \pi_n = 1\}$. 

Wang
The Wilks Theorem of EL for Univariate Mean

Since $\max \{ L_n(F) : F \} = (1/n)^n$, it follows

$$R_n(\mu) = \sup \left\{ \prod_{i=1}^n \pi_i : \pi \in \mathcal{P}_n, \sum_{i=1}^n \pi_i X_i = \mu \right\}$$

$$= \sup \left\{ \prod_{i=1}^n n\pi_i : \pi \in \mathcal{P}_n, \sum_{i=1}^n \pi_i X_i = \mu \right\}$$
The Wilks Theorem of EL for Univariate Mean

- Since \( \max \{ \mathcal{L}_n(F) : F \} = (1/n)^n \), it follows

\[
R_n(\mu) = \sup \left\{ \prod_{i=1}^{n} \pi_i : \pi \in \mathcal{P}_n, \sum_{i=1}^{n} \pi_i X_i = \mu \right\} \frac{(1/n)^n}{(1/n)^n}
\]

\[
= \sup \left\{ \prod_{i=1}^{n} n\pi_i : \pi \in \mathcal{P}_n, \sum_{i=1}^{n} \pi_i X_i = \mu \right\}
\]

- Suppose \( X_1, \ldots, X_n \) are i.i.d with \( 0 < \text{Var}(X_1) < \infty \). Then

\[
-2 \log R_n(\mu_0) \Rightarrow \chi^2_1
\]
The Wilks Theorem of EL for Univariate Mean

Since \( \max \{ L_n(F) : F \} = (1/n)^n \), it follows

\[
R_n(\mu) = \sup \left\{ \prod_{i=1}^{n} \pi_i : \pi \in P_n, \sum_{i=1}^{n} \pi_i X_i = \mu \right\} \frac{(1/n)^n}{(1/n)^n}
\]

\[
= \sup \left\{ \prod_{i=1}^{n} n\pi_i : \pi \in P_n, \sum_{i=1}^{n} \pi_i X_i = \mu \right\}
\]

Suppose \( X_1, \ldots, X_n \) are iid with \( 0 < \text{Var}(X_1) < \infty \). Then

\[
-2 \log R_n(\mu_0) \Rightarrow \chi_1^2
\]
Construction

Available side information expressed via

\[ E(u_k(Z)) = 0, \ k = 1, \ldots, m \]

Following Owen's method,

\[ R_n = \sup \left\{ n \prod_{j=1}^n \pi_j : \pi \in P_n, \ n \sum_{j=1}^n \pi_j u_k(Z_j) = 0, \ k = 1, 2, \ldots, m \right\} \]

where

\[ P_n = \{ \pi = (\pi_1, \ldots, \pi_n)^\top \in [0, 1]^n : n \sum_{i=1}^{n} \pi_i = 1 \} \]
Construction

- Available side information expressed via

\[ E(u_k(Z)) = 0, \quad k = 1, \ldots, m \]
Construction

- Available side information expressed via

\[ E(u_k(Z)) = 0, \quad k = 1, \ldots, m \]

- Following Owen’s method,

\[ R_n = \sup \left\{ \prod_{j=1}^{n} n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^{n} \pi_j u_k(Z_j) = 0, k = 1, 2, \ldots, m \right\}, \]

where

\[ \mathcal{P}_n = \{ \pi = (\pi_1, \ldots, \pi_n)^\top \in [0, 1]^n : \sum_{i=1}^{n} \pi_i = 1 \}. \]
Construction

Using Lagrange multipliers one derives that

\[ \tilde{\pi}_j = \frac{1}{n} \frac{1}{1 + \tilde{\zeta}^\top \mathbf{u}(Z_j)}, \quad j = 1, \ldots, n, \]
Using Lagrange multipliers one derives that

\[ \tilde{\pi}_j = \frac{1}{n} \frac{1}{1 + \tilde{\zeta}^\top u(Z_j)}, \quad j = 1, \ldots, n, \]

where \( u = (u_1, \ldots, u_m)^\top \) and \( \tilde{\zeta} \) solves the equation

\[ \sum_{j=1}^{n} \frac{u(Z_j)}{1 + \tilde{\zeta}^\top u(Z_j)} = 0. \]

under certain conditions, and the solution is unique.
A natural estimate \( \tilde{\theta} \) of \( \theta \) is given by

\[
\tilde{\theta} = \sum_{j=1}^{n} \tilde{\pi}_j \psi(Z_j).
\]
A natural estimate $\tilde{\theta}$ of $\theta$ is given by

$$\tilde{\theta} = \sum_{j=1}^{n} \tilde{\pi}_j \psi(Z_j).$$

This estimator of $\theta$ improves the efficiency of the empirical estimator $\bar{\psi} = \frac{1}{n} \sum_{j=1}^{n} \psi(Z_j)$ as is shown later. As a matter of fact, it is the maximum empirical likelihood estimator (MELE) of $\theta$. 
Efficiency

Example 4.1: Estimate distribution function when mean is zero

Let $X_1, \ldots, X_n \sim i.i.d.$ with side information $\mathbb{E}(X) = 0$.

Let $\psi_x(X) = 1[\{X \leq x\}$, we're interested in estimating $\theta(x) = \mathbb{E}(\psi_x(X)) = \Phi(x)$ where $x$ is a known number.

By our approach, $\tilde{F}(x) = \tilde{\theta}(x) = \frac{1}{n} \sum_{j=1}^{n} 1[\{X_j \leq x\] + \tilde{\zeta} X_j$. 

Wang
Efficiency

- Finitely many known $u_1, \ldots, u_m$
Efficiency

- Finitely many known $u_1, \ldots, u_m$

**Example 4.1**: Estimate distribution function when mean is zero
Efficiency

- Finitely many known $u_1, \ldots, u_m$

**Example 4.1**: Estimate distribution function when mean is zero
- Let $X_1, \ldots, X_n \sim i.i.d.$. $X$ with side information $E(X) = 0$
Efficiency

- Finitely many known $u_1, \ldots, u_m$

**Example 4.1:** Estimate distribution function when mean is zero

- Let $X_1, \ldots, X_n \sim i.i.d. \ X$ with side information $E(X) = 0$
- Let $\psi_x(X) = 1[X \leq x]$, we’re interested in estimating

$$\theta(x) = E(\psi(X)) = F(x)$$

where $x$ is a known number.
Efficiency

- Finitely many known $u_1, \ldots, u_m$

**Example 4.1**: Estimate distribution function when mean is zero
- Let $X_1, \ldots, X_n \sim i.i.d. \ X$ with side information $E(X) = 0$
- Let $\psi_x(X) = 1[X \leq x]$, we’re interested in estimating
  \[
  \theta(x) = E(\psi(X)) = F(x)
  \]
  where $x$ is a known number.
- By our approach,
  \[
  R_n = \sup \left\{ \prod_{j=1}^{n} n \pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^{n} \pi_j(X_j) = 0 \right\}
  \]
Efficiency

- Finitely many known $u_1, \ldots, u_m$

**Example 4.1:** Estimate distribution function when mean is zero

- Let $X_1, \ldots, X_n \sim i.i.d. \ X$ with side information $E(X) = 0$

- Let $\psi_x(X) = 1[X \leq x]$, we’re interested in estimating

$$\theta(x) = E(\psi(X)) = F(x)$$

where $x$ is a known number.

- By our approach,

$$\mathcal{R}_n = \sup \left\{ \prod_{j=1}^{n} n\pi_j : \pi \in \mathcal{P}_n, \sum_{j=1}^{n} \pi_j(X_j) = 0 \right\}$$

- A nature estimator is given by

$$\tilde{F}(x) = \tilde{\theta}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1[X_j \leq x]}{1 + \zeta X_j}.$$
THEOREM 1.1

Suppose $u_1, \ldots, u_m$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. If $W = E(u(Z)u(Z)^	op)$ is positive definite, then $\tilde{\theta}$ satisfies the stochastic expansion,

$$\tilde{\theta} = \bar{\psi} - \bar{\phi}_0 + o_p(n^{-1/2}),$$

where $\phi_0 = \Pi(\psi|\mathbf{u}) = E(\psi \otimes \mathbf{u}^\top)W^{-1}\mathbf{u}$. 
THEOREM 1.1

Suppose \( u_1, \ldots, u_m \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. If \( W = E(u(Z)u(Z)\top) \) is positive definite, then \( \tilde{\theta} \) satisfies the stochastic expansion,

\[
\tilde{\theta} = \bar{\psi} - \bar{\phi}_0 + o_p(n^{-1/2}),
\]

where \( \phi_0 = \Pi(\psi|[u]) = E(\psi \otimes u\top)W^{-1}u \).

Thus,

\[
\sqrt{n}(\tilde{\theta} - \theta) \Rightarrow \mathcal{N}(0, V_1)
\]
THEOREM 1.1

Suppose $u_1, \ldots, u_m$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. If $W = E(u(Z)u(Z)^	op)$ is positive definite, then $\tilde{\theta}$ satisfies the stochastic expansion,

$$\tilde{\theta} = \bar{\psi} - \bar{\phi}_0 + o_p(n^{-1/2}),$$

where $\phi_0 = \Pi(\psi|[u]) = E(\psi \otimes u^\top)W^{-1}u$.

Thus,

$$\sqrt{n}(\tilde{\theta} - \theta) \Rightarrow \mathcal{N}(0, V_1)$$

where $V_1 = \text{Var}(\psi(Z)) - \text{Var}(\phi_0(Z))$. 

By the above theorem, the estimator $\tilde{\theta}$ has smaller variance than the variance of the empirical estimator $\bar{\psi}$. 

Wang

An Empirical Likelihood Approach of The Estimation of Linear Functionals with Side Information
THEOREM 1.1

Suppose $u_1, \ldots, u_m$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. If $W = E(u(Z)u(Z)^	op)$ is positive definite, then $\tilde{\theta}$ satisfies the stochastic expansion,

$$\tilde{\theta} = \bar{\psi} - \bar{\phi}_0 + o_p(n^{-1/2}),$$

where $\phi_0 = \Pi(\psi|[u]) = E(\psi \otimes u^\top)W^{-1}u$.

Thus,

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V_1)$$

where $V_1 = Var(\psi(Z)) - Var(\phi_0(Z))$.

By the above theorem, the estimator $\tilde{\theta}$ has smaller variance than the variance of the empirical estimator $\bar{\psi}$. 

By the above theorem, the estimator $\tilde{\theta}$ has smaller variance than the variance of the empirical estimator $\bar{\psi}$. 

Efficiency
Example 4.1 continued:

In this case, $u(X) = X$, $W = E(X^2) = Var(X) = \sigma^2 > 0$.

Recall that $\tilde{F}(x) = \tilde{\theta}(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}[X_j \leq x] + \tilde{\zeta} X_j = F_n(x) - \bar{\phi}_0 + o_p(n^{-1/2})$,

where $\phi_0(X) = E(\psi \otimes u^\top W^{-1} u) = E[\mathbb{1}[X \leq x] X] X \sigma^2$.

Then $V_1 = F(x)(1 - F(x)) - E[\mathbb{1}[X \leq x] X] X^2$.
Example 4.1 continued:

- In this case, $u(X) = X$, $W = E(X^2_1) = Var(X_j) = \sigma^2 > 0$
Example 4.1 continued:

- In this case, \( u(X) = X \), \( W = E(X^2) = Var(X_j) = \sigma^2 > 0 \)
- Recall that

\[
\tilde{F}(x) = \tilde{\theta}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1[X_j \leq x]}{1 + \zeta X_j} = F_n(x) - \bar{\phi}_0 + o_p(n^{-1/2}),
\]
Example 4.1 continued:

- In this case, $u(X) = X$, $W = E(X_i^2) = \text{Var}(X_j) = \sigma^2 > 0$
- Recall that

$$
\tilde{F}(x) = \tilde{\theta}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1[X_j \leq x]}{1 + \zeta X_j} = F_n(x) - \phi_0 + o_p(n^{-1/2}),
$$

where

$$
\phi_0(X) = E(\psi \otimes u^\top)W^{-1}u = \frac{E[1[X \leq x]X]X}{\sigma^2}.
$$

Then

$$
V_1 = F(x)(1 - F(x)) - E[1[X \leq x]X]^2
$$
THEOREM 1.2

Suppose \( u_1, \ldots, u_m \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. Let \( \hat{u}_1, \ldots, \hat{u}_m \) be their estimators respectively such that
THEOREM 1.2

Suppose \( u_1, \ldots, u_m \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. Let \( \hat{u}_1, \ldots, \hat{u}_m \) be their estimators respectively such that

\[
\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p(n^{1/2}),
\]
THEOREM 1.2

Suppose $u_1, \ldots, u_m$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. Let $\hat{u}_1, \ldots, \hat{u}_m$ be their estimators respectively such that

$$\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p(n^{1/2}),$$

$$|\hat{W}_n - W|_o = o_p(1),$$
THEOREM 1.2

Suppose \( u_1, \ldots, u_m \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. Let \( \hat{u}_1, \ldots, \hat{u}_m \) be their estimators respectively such that

\[
\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p(n^{1/2}),
\]

\[
|\hat{W}_n - W|_o = o_p(1),
\]

where \( \hat{W}_n = \frac{1}{n} \sum_{j=1}^{n} \hat{u}_n(Z_j) \hat{u}_n(Z_j)^\top \), and
THEOREM 1.2

- Suppose \( u_1, \ldots, u_m \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. Let \( \hat{u}_1, \ldots, \hat{u}_m \) be their estimators respectively such that

\[
\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p(n^{1/2}),
\]

\[
|\hat{W}_n - W|_o = o_p(1),
\]

where \( \hat{W}_n = \frac{1}{n} \sum_{j=1}^{n} \hat{u}_n(Z_j)\hat{u}_n(Z_j)^\top \), and

\[
\frac{1}{n} \sum_{j=1}^{n} (\psi(Z_j) \otimes \hat{u}(Z_j)^\top - E(\psi(Z_j) \otimes \hat{u}(Z_j)^\top)) = o_p(1),
\]
THEOREM 1.2 continued

Also, for some measurable function $v_n$ from $\mathcal{Z}$ into $\mathbb{R}^m$ such that $\int v_n(Z) dP = 0$ and $|v_n|$ is Lindeberg, it satisfies
THEOREM 1.2 continued

Also, for some measurable function \( v_n \) from \( \mathcal{L} \) into \( \mathbb{R}^m \) such that \( \int v_n(Z) dP = 0 \) and \( |v_n| \) is Lindeberg, it satisfies

\[
\frac{1}{n} \sum_{j=1}^{n} E \left( |\hat{u}_n(Z_j) - v_n(Z_j)|^2 \right) = o_p(1),
\]

where \( \hat{u}_n \) is the empirical estimator of \( u(Z) \).
THEOREM 1.2 continued

Also, for some measurable function \( v_n \) from \( \mathcal{L} \) into \( \mathbb{R}^m \) such that 
\[
\int v_n(Z) dP = 0 \quad \text{and} \quad |v_n| \text{ is Lindeberg, it satisfies}
\]
\[
\frac{1}{n} \sum_{j=1}^{n} E \left( |\hat{u}_n(Z_j) - v_n(Z_j)|^2 \right) = o_p(1),
\]
\[
\frac{1}{n} \sum_{j=1}^{n} \hat{u}_n(Z_j) = \frac{1}{n} \sum_{j=1}^{n} v_n(Z_j) + o_p(n^{-1/2}),
\]
THEOREM 1.2 continued

Then $\hat{\theta}$ satisfies the stochastic expansion,

$$\hat{\theta} = \bar{\psi} - \bar{\phi} + o_p(n^{-1/2}),$$

where $\phi = E(\psi \otimes u^\top)W^{-1}u$. 
THEOREM 1.2 continued

- Then $\hat{\theta}$ satisfies the stochastic expansion,

$$\hat{\theta} = \bar{\psi} - \bar{\phi} + o_p(n^{-1/2}),$$

where $\phi = E(\psi \otimes u^\top) W^{-1} u$.

Thus

$$\sqrt{n}(\hat{\theta} - \theta) \overset{d}{\rightarrow} \mathcal{N}(0, V_2),$$

where $V_2 = \text{Var}(\psi(Z)) - \text{Var}(\phi(Z))$.
Example 4.3

Estimate distribution function when the marginal distribution is known

Suppose $X, Y$ have joint distribution $H$. Suppose the marginal distribution function of $X$ is known and equal to $F_0$. Suppose $F_0$ is continuous, then the above information is equivalent to

$$\int a(x) \, dH(x, y) = \int a(x) \, dF_0(x) = 0, \quad \forall a \in L^2_0(F_0).$$

This can be reduced to countably many constraints

$$\int a_k(x) \, dH(x, y) = \int a_k(x) \, dF_0(x) = 0, \quad k = 1, 2, ...$$

where $a_k(x)$ is an orthonormal basis of $L^2_0(F_0)$. 

▶ Infinitely many constraints
Efficiency

- Infinitely many constraints

**Example 4.3** Estimate distribution function when the marginal distribution is known
Example 4.3 Estimate distribution function when the marginal distribution is known

Suppose $X, Y$ have joint distribution $H$. Suppose the marginal distribution function of $X$ is known and equal to $F_0$. This can be reduced to countably many constraints

$\int a_k(x) \, dH(x, y) = \int a_k(x) \, dF_0(x) = 0, \quad k = 1, 2, \ldots$

where $a_k(x)$ is an orthonormal basis of $L^2(F_0)$. 

Infinitely many constraints
Efficiency

- Infinitely many constraints

**Example 4.3** Estimate distribution function when the marginal distribution is known

- Suppose $X, Y$ have joint distribution $H$. Suppose the marginal distribution function of $X$ is known and equal to $F_0$.

- Suppose $F_0$ is continuous, then the above information is equivalent to

\[
\int a(x) dH(x, y) = \int a(x) dF_0(x) = 0, \quad \forall a \in L_{2,0}(F_0)
\]
Infinitely many constraints

**Example 4.3** Estimate distribution function when the marginal distribution is known

Suppose \( X, Y \) have joint distribution \( H \). Suppose the marginal distribution function of \( X \) is known and equal to \( F_0 \).

Suppose \( F_0 \) is continuous, then the above information is equivalent to

\[
\int a(x) dH(x, y) = \int a(x) dF_0(x) = 0, \quad \forall a \in L_{2,0}(F_0)
\]

This can be reduced to countably many constraints

\[
\int a_k(x) dH(x, y) = \int a_k(x) dF_0(x) = 0, \quad k = 1, 2, \ldots
\]

where \( a_k(x) \) is an orthonormal basis of \( L_{2,0}(F_0) \).
Rewrite conditions to be $E(a_k(X)) = 0$
Efficiency

- Rewrite conditions to be $E(a_k(X)) = 0$
- Trigonometric basis $\phi_1, \phi_2, \ldots$ defined by
  \[ \phi_k(t) = \sqrt{2} \cos(k\pi t), \quad t \in [0, 1], \quad k = 1, 2, \ldots \]
Rewrite conditions to be $E(a_k(X)) = 0$

Trigonometric basis $\phi_1, \phi_2, \ldots$ defined by

$$\phi_k(t) = \sqrt{2} \cos(k\pi t), \quad t \in [0, 1], \quad k = 1, 2, \ldots.$$ 

Then $u_k(X_j) = \phi_k(F_0(X_j)), \quad k = 1, 2, \ldots, m,$
Rewrite conditions to be $E(a_k(X)) = 0$

Trigonometric basis $\phi_1, \phi_2, \ldots$ defined by

$$\phi_k(t) = \sqrt{2} \cos(k\pi t), \quad t \in [0, 1], \quad k = 1, 2, \ldots.$$ 

Then $u_k(X_j) = \phi_k(F_0(X_j)), \quad k = 1, 2, \ldots, m$, 

Estimate $\theta = F(x, y) = E(\psi(x, y))$, where

$$\psi(x, y) = 1[X \leq x, Y \leq y],$$
Rewrite conditions to be $E(a_k(X)) = 0$

Trigonometric basis $\phi_1, \phi_2, \ldots$ defined by

$$\phi_k(t) = \sqrt{2} \cos(k\pi t), \quad t \in [0, 1], \quad k = 1, 2, \ldots.$$

Then $u_k(X_j) = \phi_k(F_0(X_j)), \quad k = 1, 2, \ldots, m$,

Estimate $\theta = F(x, y) = E(\psi(x, y))$, where

$$\psi(x, y) = 1[X \leq x, Y \leq y],$$

then a nature estimator would be

$$\tilde{\theta} = \frac{1}{n} \sum_{j=1}^{n} \frac{1[X \leq x, Y \leq y]}{1 + \tilde{\zeta}^\top \Phi(F_0(X_j))},$$

where $\Phi(t) = (\phi_1(t), \ldots, \phi_m(t))^\top$. 
THEOREM 1.3

Suppose that \( u_1, u_2, \ldots \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. If
THEOREM 1.3

Suppose that \( u_1, u_2, \ldots \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. If

\[
\max_{1 \leq j \leq n} |u_n(Z_j)| = o_p(m_n^{-1} n^{1/2}),
\]

where \( u_n(Z_j) = \frac{1}{n} \sum_{j=1}^{n} (\psi(Z_j) \otimes u(Z_j))^\top - E[\psi(Z_j) \otimes u(Z_j)]^\top \) and \( W_n \) is the dispersion matrix.
THEOREM 1.3

Suppose that $u_1, u_2, \ldots$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. If

$$\max_{1 \leq j \leq n} |u_n(Z_j)| = o_p(m_n^{-1} n^{1/2}),$$

the $m_n \times m_n$ dispersion matrix $W_n$ is regular and satisfies

$$|S_n - W_n|_o = o_p(m_n^{-1})$$
THEOREM 1.3

- Suppose that $u_1, u_2, \ldots$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. If

$$\max_{1 \leq j \leq n} |u_n(Z_j)| = o_p(m_n^{-1} n^{1/2}),$$

the $m_n \times m_n$ dispersion matrix $W_n$ is regular and satisfies

$$|S_n - W_n|_o = o_p(m_n^{-1})$$

and

$$\left| \frac{1}{n} \sum_{j=1}^{n} \left( \psi(Z_j) \otimes u(Z_j)^\top - E(\psi(Z_j) \otimes u(Z_j)^\top) \right) \right|_o = o_p(m_n^{-1/2}).$$
THEOREM 1.3 (continued)

Then \( \hat{\theta}_n \) satisfies, as \( m_n \) tends to infinity, the stochastic expansion,

\[
\hat{\theta}_n = \bar{\psi} - \bar{\varphi}_0 + o_p(n^{-1/2}),
\]

where \( \varphi_0 = \Pi(\psi|\mathbf{u}_\infty) \) is the projection of \( \psi \) onto the closed linear span \( [\mathbf{u}_\infty] \).
THEOREM 1.3 (continued)

Then \( \tilde{\theta}_n \) satisfies, as \( m_n \) tends to infinity, the stochastic expansion,

\[
\tilde{\theta}_n = \bar{\psi} - \bar{\varphi}_0 + o_p(n^{-1/2}),
\]

where \( \varphi_0 = \prod(\psi|[u_{\infty}]) \) is the projection of \( \psi \) onto the closed linear span \([u_{\infty}]\).

Thus

\[
\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{\text{d}} N(0, V_3).
\]

where \( V_3 = \text{Var}(\psi(Z)) - \text{Var}(\varphi_0(Z)) \).
Example 4.3 continued

- Apply theorem 1.3, check conditions:

  For $u_k(Z_j) = \phi_k(F_0(X_j)) = \sqrt{2} \cos(k\pi F_0(X_j)),$
Example 4.3 continued

- Apply theorem 1.3, check conditions:
  
  For $u_k(Z_j) = \phi_k(F_0(X_j)) = \sqrt{2} \cos(k\pi F_0(X_j))$,

- (1) $u_k$ has mean zero and finite variances
Example 4.3 continued

- Apply theorem 1.3, check conditions:
  - For $u_k(Z_j) = \phi_k(F_0(X_j)) = \sqrt{2} \cos(k\pi F_0(X_j))$,
- (1) $u_k$ has mean zero and finite variances
- (2) $\max_{1 \leq j \leq n} |u_n(Z_j)| = o_p(m_n^{-1}n^{1/2}) \iff m_n^3 = o_p(n)$
Example 4.3 continued

- Apply theorem 1.3, check conditions:
  For \( u_k(Z_j) = \phi_k(F_0(X_j)) = \sqrt{2} \cos(k\pi F_0(X_j)) \),

- (1) \( u_k \) has mean zero and finite variances

- (2) \( \max_{1 \leq j \leq n} |u_n(Z_j)| = o_p(m_n^{-1}n^{1/2}) \iff m_n^3 = o_p(n) \)

- (3) \( W_n \) is regular and satisfies \( |S_n - W_n|_o = o_p(m_n^{-1}) \)

\[ \iff W_n = I, m_n^4 = o_p(n) \]
Example 4.3 continued

- Apply theorem 1.3, check conditions:
  \[ u_k(Z_j) = \phi_k(F_0(X_j)) = \sqrt{2} \cos(k \pi F_0(X_j)), \]
  \[ u_k \text{ has mean zero and finite variances} \]
  \[ \max_{1 \leq j \leq n} |u_n(Z_j)| = o_p(m_n^{-1}n^{1/2}) \iff m_n^3 = o_p(n) \]
- \[ W_n \text{ is regular and satisfies } |S_n - W_n|_o = o_p(m_n^{-1}) \]
  \[ \iff W_n = I, m_n^4 = o_p(n) \]

- \[ \frac{1}{n} \sum_{j=1}^{n} (\psi(Z_j) \otimes u(Z_j)^\top - E(\psi(Z_j) \otimes u(Z_j)^\top)) = o_p(m_n^{-1/2}) \]
  \[ \iff m_n^2 = o_p(n) \]
Example 4.3 continued

- After satisfying all the conditions, we can apply theorem 1.3 that the estimator

\[ \tilde{\theta}_n = \bar{\psi} - \bar{\phi}_0 + o_p(n^{-1/2}), \]

and when

\[ m_4 = o_p(n^{1/2}) \],

\[ \sqrt{n}(\tilde{\theta}_n - \theta) \Rightarrow N(0, V_3) \]
Example 4.3 continued

- After satisfying all the conditions, we can apply theorem 1.3 that the estimator

\[ \hat{\theta}_n = \bar{\psi} - \bar{\varphi}_0 + o_p(n^{-1/2}), \]
Example 4.3 continued

After satisfying all the conditions, we can apply theorem 1.3 that the estimator

$$
\tilde{\theta}_n = \bar{\Psi} - \bar{\varphi}_0 + o_p(n^{-1/2}),
$$

and when $m_n^4 = o_p(n)$,

$$
\sqrt{n}(\tilde{\theta}_n - \theta) \implies \mathcal{N}(0, V_3)
$$
Example 4.3 continued

- Calculation of \( \varphi_0 = \Pi(\psi|\mathbf{u}_\infty) \):
Example 4.3 continued

• Calculation of \( \varphi_0 = \Pi(\psi|\mathbf{u}_\infty) \):

\[
\varphi_0 = \Pi(\psi|L_{2,0}(F_0))
\]
Example 4.3 continued

- Calculation of $\varphi_0 = \Pi(\psi|u_\infty)$:

\[
\varphi_0 = \Pi(\psi|L_{2,0}(F_0))
\]

$\Leftrightarrow (\psi - \varphi_0) \perp L_{2,0}(F_0)$
Example 4.3 continued

- Calculation of $\varphi_0 = \Pi(\psi|[u_\infty])$:

  $$\varphi_0 = \Pi(\psi|L_{2,0}(F_0))$$

  $$\Leftrightarrow (\psi - \varphi_0) \perp L_{2,0}(F_0)$$

  $$\Leftrightarrow E((\psi - \varphi_0)a) = 0, \forall a \in L_{2,0}(F_0)$$
Example 4.3 continued

- Calculation of \( \varphi_0 = \Pi(\psi|[u_\infty]) \):

\[
\varphi_0 = \Pi(\psi|L_{2,0}(F_0))
\]

\[\iff (\psi - \varphi_0) \perp L_{2,0}(F_0)\]

\[\iff E((\psi - \varphi_0)a) = 0, \forall a \in L_{2,0}(F_0)\]

\[\iff E((\psi(X, Y) - \varphi_0(X))a(X)) = 0, \forall a \in L_{2,0}(F_0)\]
Example 4.3 continued

- Calculation of \( \varphi_0 = \Pi(\psi|[u_\infty]) \):

\[
\varphi_0 = \Pi(\psi|L_{2,0}(F_0))
\]

\( \iff (\psi - \varphi_0) \perp L_{2,0}(F_0) \)

\( \iff E((\psi - \varphi_0)a) = 0, \forall a \in L_{2,0}(F_0) \)

\( \iff E((\psi(X, Y) - \varphi_0(X))a(X)) = 0, \forall a \in L_{2,0}(F_0) \)

Change it to the conditional expectation given \( X = x \), we can solve

\[
\varphi_0(x) = E(\psi(X, Y)|X = x)
\]
Example 4.3 continued

**Theorem:**
Let \((X_1, Y_1), \ldots, (X_n, Y_n)\) be i.i.d. from \(H\). Suppose \(X_1\) has known distribution \(F_0\), which is continuous. Then \(\hat{\theta}\) is an efficient estimator of \(\theta\) such that

\[
\hat{\theta} = \frac{1}{n} \sum_{j=1}^{n} \psi(X_j, Y_j) - \frac{1}{n} \sum_{j=1}^{n} \varphi_0(X_j) + o_p(n^{-1/2})
\]

where \(\varphi_0\) is the projector of \(\psi\) onto \(L_{2,0}(F_0)\), with formular

\[
\varphi_0(x) = E(\psi(X, Y) \mid X = x),
\]

Thus

\[
\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow \mathcal{N}(0, \sigma^2).
\]

where \(\sigma^2 = \text{Var}(\psi) - \text{Var}(\varphi_0)\).
THEOREM 1.4

Suppose that \( u_1, u_2, \ldots \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. Let \( \hat{u}_1, \hat{u}_2, \ldots \) be their estimators respectively such that
THEOREM 1.4

Suppose that $u_1, u_2, \ldots$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. Let $\hat{u}_1, \hat{u}_2, \ldots$ be their estimators respectively such that

$$
\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p\left(m_n^{-1}n^{1/2}\right),
$$
THEOREM 1.4

Suppose that $u_1, u_2, \ldots$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. Let $\hat{u}_1, \hat{u}_2, \ldots$ be their estimators respectively such that

$$\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p(m_n^{-1} n^{1/2}),$$

$$|\hat{W}_n - W_n|_o = o_p(m_n^{-1}),$$
THEOREM 1.4

Suppose that $u_1, u_2, \ldots$ are known functions s.t. $E[u_k(Z)] = 0$ and second moment finite. Let $\hat{u}_1, \hat{u}_2, \ldots$ be their estimators respectively such that

$$\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p(m_n^{-1}n^{1/2}),$$

$$|\hat{W}_n - W_n|_o = o_p(m_n^{-1}),$$

$$\frac{1}{n} \sum_{j=1}^{n} \left( \psi(Z_j) \otimes \hat{u}_n(Z_j)^\top - E(\psi(Z_j) \otimes \hat{u}_n(Z_j)^\top) \right) = o_p(m_n^{-1/2}),$$
THEOREM 1.4

Suppose that \( u_1, u_2, \ldots \) are known functions s.t. \( E[u_k(Z)] = 0 \) and second moment finite. Let \( \hat{u}_1, \hat{u}_2, \ldots \) be their estimators respectively such that

\[
\max_{1 \leq j \leq n} |\hat{u}_n(Z_j)| = o_p(m_n^{-1/2}),
\]

\[
|\hat{W}_n - W_n|_o = o_p(m_n^{-1}),
\]

\[
\frac{1}{n} \sum_{j=1}^{n} \left( \psi(Z_j) \otimes \hat{u}_n(Z_j)^\top - E(\psi(Z_j) \otimes \hat{u}_n(Z_j)^\top) \right) = o_p(m_n^{-1/2}),
\]

\[
\frac{1}{n} \sum_{j=1}^{n} E \left( |\hat{u}_n(Z_j) - v_n(Z_j)|^2 \right) = o_p(1),
\]
THEOREM 1.4 continued

- for some measurable function $v_n$ from $\mathcal{L}$ into $\mathbb{R}^{m_n}$ such that
  \[ \int v_n(Z) \, dP = 0 \]
  and $|v_n|$ is Lindeberg.

\[
\frac{1}{n} \sum_{j=1}^{n} \hat{u}_n(Z_j) = \frac{1}{n} \sum_{j=1}^{n} v_n(Z_j) + o_p(n^{-1/2}),
\]
THEOREM 1.4 continued

- for some measurable function $v_n$ from $\mathcal{L}$ into $\mathbb{R}^{m_n}$ such that $\int v_n(Z) \, dP = 0$ and $|v_n|$ is Lindeberg.

\[
\frac{1}{n} \sum_{j=1}^{n} \hat{u}_n(Z_j) = \frac{1}{n} \sum_{j=1}^{n} v_n(Z_j) + o_p(n^{-1/2}),
\]

- Then $\hat{\theta}$ satisfies, as $m_n$ tends to infinity, the stochastic expansion,

\[
\hat{\theta} = \bar{\psi} - \bar{\varphi} + o_p(n^{-1/2}),
\]
THEOREM 1.4 continued

- for some measurable function $v_n$ from $\mathcal{L}$ into $\mathbb{R}^{m_n}$ such that 
  $\int v_n(Z) \, dP = 0$ and $|v_n|$ is Lindeberg.

\[
\frac{1}{n} \sum_{j=1}^{n} \hat{u}_n(Z_j) = \frac{1}{n} \sum_{j=1}^{n} v_n(Z_j) + o_p(n^{-1/2}),
\]

- Then $\hat{\theta}$ satisfies, as $m_n$ tends to infinity, the stochastic expansion,

\[
\hat{\theta} = \bar{\psi} - \bar{\varphi} + o_p(n^{-1/2}),
\]

where $\varphi = \Pi(\psi | [v_\infty])$ is the projection of $\psi$ onto the closed linear span $[v_\infty]$. 
THEOREM 1.4 continued

1. For some measurable function $v_n$ from $\mathcal{L}$ into $\mathbb{R}^{m_n}$ such that
   \[ \int v_n(Z) \, dP = 0 \text{ and } |v_n| \text{ is Lindeberg.} \]

2. Then $\hat{\theta}$ satisfies, as $m_n$ tends to infinity, the stochastic expansion,
   \[ \hat{\theta} = \bar{\psi} - \bar{\varphi} + o_p(n^{-1/2}), \]
   where $\varphi = \Pi(\psi|[v_\infty])$ is the projection of $\psi$ onto the closed linear span $[v_\infty]$.
   Thus
   \[ \sqrt{n}(\hat{\theta}_n - \theta) \rightsquigarrow \mathcal{N}(0, V_4). \]
   where
   \[ V_4 = \text{Var}(\psi(Z)) - \text{Var}(\varphi(Z)). \]
Conclusion and Further Topics

Based on previous discussion, when we obtain the side information, we can give a more efficiency estimator to the linear functional we are interested in.
Conclusion and Further Topics

- Based on previous discussion, when we obtain the side information, we can give a more efficiency estimator to the linear functional we are interested in.
- The same method can apply to many topics, Especially interests in Missing Data.
Thank you!