Chapter 4: Imputation

Jae-Kwang Kim

Department of Statistics, Iowa State University

Outline

1. Introduction
2. Basic Theory for imputation
3. Variance estimation after imputation
4. Replication variance estimation
5. Multiple imputation
6. Fractional imputation

Introduction

Basic setup

- $\mathbf{Y}$: a vector of random variables with distribution $F(\mathbf{y};\theta)$.
- $\mathbf{y}_1, \cdots, \mathbf{y}_n$ are n independent realizations of $\mathbf{Y}$.
- We are interested in estimating ψ which is implicitly defined by $E\{U(\psi;\mathbf{Y})\} = 0$.
- Under complete observation, a consistent estimator $\hat{\psi}_n$ of ψ can be obtained by solving estimating equation for ψ:

$$\sum_{i=1}^{n} U(\psi;\mathbf{y}_i) = 0.$$

- A special case of estimating function is the score function. In this case, $\psi = \theta$.
- Sandwich variance estimator is often used to estimate the variance of $\hat{\psi}_n$:

$$\hat{V}(\hat{\psi}_n) = \hat{\tau}_u^{-1}\hat{V}(U)\hat{\tau}_u^{-1'}$$

where $\tau_u = E\{\partial U(\psi;\mathbf{y})/\partial\psi'\}$.

1. Introduction

Missing data setup

- Suppose that $\mathbf{y}_i$ is not fully observed.
- $\mathbf{y}_i = (\mathbf{y}_{obs,i}, \mathbf{y}_{mis,i})$: (observed, missing) part of $\mathbf{y}_i$
- $\boldsymbol{\delta}_i$: response indicator functions for $\mathbf{y}_i$.
- Under the existence of missing data, we can use the following estimators:

$$\hat{\psi}\text{: solution to } \sum_{i=1}^{n} E\left\{U\left(\psi; \mathbf{y}_i\right) \mid \mathbf{y}_{obs,i}, \boldsymbol{\delta}_i\right\} = 0. \tag{1}$$

- The equation in (1) is often called expected estimating equation.

1. Introduction

Motivation (for imputation)

Computing the conditional expectation in (1) can be a challenging problem.

1. The conditional expectation depends on unknown parameter values. That is,

$$E\left\{U\left(\psi;\mathbf{y}_i\right)\mid \mathbf{y}_{obs,i},\boldsymbol{\delta}_i\right\}=E\left\{U\left(\psi;\mathbf{y}_i\right)\mid \mathbf{y}_{obs,i},\boldsymbol{\delta}_i;\theta,\phi\right\},$$

   where θ is the parameter in $f(\mathbf{y};\theta)$ and ϕ is the parameter in $p(\boldsymbol{\delta}\mid\mathbf{y};\phi)$.
2. Even if we know $\eta=(\theta,\phi)$, computing the conditional expectation is numerically difficult.

- Imputation: Monte Carlo approximation of the conditional expectation (given the observed data).

$$E\left\{U\left(\psi;\mathbf{y}_i\right)\mid \mathbf{y}_{obs,i},\boldsymbol{\delta}_i\right\}\cong\frac{1}{m}\sum_{j=1}^{m}U\left(\psi;\mathbf{y}_{obs,i},\mathbf{y}_{mis,i}^{*(j)}\right)$$

1. Bayesian approach: generate $\mathbf{y}_{mis,i}^{*}$ from

$$f\left(\mathbf{y}_{mis,i}\mid \mathbf{y}_{obs},\boldsymbol{\delta}\right)=\int f\left(\mathbf{y}_{mis,i}\mid \mathbf{y}_{obs},\boldsymbol{\delta};\eta\right)p(\eta\mid \mathbf{y}_{obs},\boldsymbol{\delta})d\eta$$

2. Frequentist approach: generate $\mathbf{y}_{mis,i}^{*}$ from $f\left(\mathbf{y}_{mis,i}\mid \mathbf{y}_{obs,i},\boldsymbol{\delta};\hat{\eta}\right)$, where $\hat{\eta}$ is a consistent estimator.

Example 4.1

- Basic Setup

  Let $(x, y)'$ be a vector of bivariate random variables. Assume that x_i are always observed and y_i are subject to missingness in the sample, and the probability of missingness does not depend on the value of y_i. In this case, an imputed estimator of $\theta = E(Y)$ based on single imputation can be computed by

$$\hat{\theta}_I = \frac{1}{n} \sum_{i=1}^{n} \left\{ \delta_i y_i + (1 - \delta_i) y_i^* \right\} \tag{2}$$

  where y_i^* is an imputed value for y_i.

- Imputation model

$$y_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma_e^2\right),$$

  for some $(\beta_0, \beta_1, \sigma_e^2)$.

Example 4.1 (Cont'd)

- Deterministic imputation: Use $y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i$ where

$$\left(\hat{\beta}_0, \hat{\beta}_1 \right) = \left(\bar{y}_r - \hat{\beta}_1 \bar{x}_r, S_{xxr}^{-1} S_{xyr} \right).$$

  Note that

$$E \left(\hat{\theta}_I - \theta \right) = 0$$

  and

$$V \left(\hat{\theta}_I \right) = \frac{1}{n} \sigma_y^2 + \left(\frac{1}{r} - \frac{1}{n} \right) \sigma_e^2 = \frac{\sigma_y^2}{r} \left\{ 1 - \left(1 - \frac{r}{n} \right) \rho^2 \right\}.$$

- Stochastic imputation: Use $y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i^*$, where $e_i^* \sim (0, \hat{\sigma}_e^2)$. The imputed estimator under stochastic imputation satisfies

$$V \left(\hat{\theta}_I \right) = \frac{1}{n} \sigma_y^2 + \left(\frac{1}{r} - \frac{1}{n} \right) \sigma_e^2 + \frac{n - r}{n^2} \sigma_e^2$$

  where the third term represents the additional variance due to stochastic imputation.

Remark

- Deterministic imputation is unbiased for the estimating the mean but may not be unbiased for estimating the proportion. For example, if $\theta = Pr(Y < c) = E\{I(Y < c)\}$, the imputed estimator

$$\hat{\theta} = n^{-1} \sum_{i=1}^{n} \{\delta_i I(y_i < c) + (1 - \delta_i) I(y_i^* < c)\}$$

  is unbiased if $E\{I(Y < c)\} = E\{I(Y^* < c)\}$, which holds only when the marginal distribution of y^* is the same as the marginal distribution of y. In general, under deterministic imputation, we have $E(y) = E(y^*)$ but $V(y) > V(y^*)$. For regression imputation, $V(y^*) = \sigma_y^2(1 - \rho^2) < \sigma_y^2 = V(y)$.
- Imputation increases the variance (of the imputed estimator) because the imputed values are positively correlated.
- Variance estimation is complicated because of the correlation between the imputed values.

§2 Basic Theory for Imputation

Lemma 4.1

Let $\hat{\theta}$ be the solution to $\hat{U}(\theta) = 0$, where $\hat{U}(\theta)$ is a function of complete observations $\mathbf{y}_1, \cdots, \mathbf{y}_n$ and parameter θ. Let θ_0 be the solution to $E\left\{\hat{U}(\theta)\right\} = 0$. Then, under some regularity conditions,

$$\hat{\theta} - \theta_0 \cong -\left[E\left\{\dot{U}(\theta_0)\right\}\right]^{-1} \hat{U}(\theta_0),$$

where $\dot{U}(\theta) = \partial \hat{U}(\theta)/\partial \theta'$ and notation $A_n \cong B_n$ means that $B_n^{-1} A_n = 1 + R_n$ for some R_n which converges to zero in probability.

Remark (about Lemma 4.1)

- Its proof is based on Taylor linearization:

$$\begin{aligned}\hat{U}(\hat{\theta}) &\cong \hat{U}(\theta_0) + \dot{U}(\theta_0)\left(\hat{\theta} - \theta_0\right) \\ &\cong \hat{U}(\theta_0) + E\{\dot{U}(\theta_0)\}\left(\hat{\theta} - \theta_0\right),\end{aligned}$$

  where the second (approximate) equality follows by

$$\dot{U}(\theta_0) = E\{\dot{U}(\theta_0)\} + o_p(1)$$

  and $\hat{\theta} = \theta_0 + o_p(1)$.

- Need to assume that $E\left\{\dot{U}(\theta_0)\right\}$ is nonsingular.
- Also, we need conditions for $\hat{\theta} \xrightarrow{p} \theta_0$.
- Lemma 4.1 can be used to establish the asymptotic normality of $\hat{\theta}$. (Use Slutsky Theorem).

2. Basic Theory for Imputation

Basic Setup (for Case 1: $\psi = \eta$)

- $\mathbf{y} = (y_1, \cdots, y_n) \sim f(\mathbf{y}; \theta)$
- $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_n) \sim P(\boldsymbol{\delta} | \mathbf{y}; \phi)$
- $\mathbf{y} = (\mathbf{y}_{obs}, \mathbf{y}_{mis})$: (observed, missing) part of $\mathbf{y}$.
- $\mathbf{y}_{mis}^{*(1)}, \cdots, \mathbf{y}_{mis}^{*(m)}$: m imputed values of $\mathbf{y}_{mis}$ generated from

$$f(\mathbf{y}_{\mathrm{mis}} \mid \mathbf{y}_{\mathrm{obs}}, \boldsymbol{\delta}; \hat{\eta}_p) = \frac{f(\mathbf{y}; \hat{\theta}_p) P(\boldsymbol{\delta} \mid \mathbf{y}; \hat{\phi}_p)}{\int f(\mathbf{y}; \hat{\theta}_p) P(\boldsymbol{\delta} \mid \mathbf{y}; \hat{\phi}_p) d\mu(\mathbf{y}_{\mathrm{mis}})},$$

  where $\hat{\eta}_p = (\hat{\theta}_p, \hat{\phi}_p)$ is a preliminary estimator of $\eta = (\theta, \phi)$.

- Using m imputed values, imputed score function is computed as

$$\bar{S}_{imp,m}^{*}(\eta \mid \hat{\eta}_p) \equiv m^{-1} \sum_{j=1}^{m} S_{\mathrm{com}}\left(\eta; \mathbf{y}_{obs}, \mathbf{y}_{mis}^{*(j)}, \boldsymbol{\delta}\right)$$

  where $S_{\mathrm{com}}(\eta; \mathbf{y})$ is the score function of $\eta = (\theta, \phi)$ under complete response

Lemma 4.2 (Asymptotic results for $m = \infty$)

Assume that $\hat{\eta}_p$ converges in probability to η. Let $\hat{\eta}^*_{I,m}$ be the solution to

$$\frac{1}{m}\sum_{j=1}^{m} S_{\mathrm{com}}\left(\eta; \mathbf{y}_{obs}, \mathbf{y}^{*(j)}_{mis}, \boldsymbol{\delta}\right) = 0,$$

where $\mathbf{y}^{*(1)}_{mis}, \cdots, \mathbf{y}^{*(m)}_{mis}$ are the imputed values generated from $f(\mathbf{y}_{\mathrm{mis}} \mid \mathbf{y}_{\mathrm{obs}}, \boldsymbol{\delta}; \hat{\eta}_p)$. Then, under some regularity conditions, for $m \to \infty$,

$$\hat{\eta}^*_{imp,\infty} \cong \hat{\eta}_{\mathrm{MLE}} + \mathcal{J}_{\mathrm{mis}}\left(\hat{\eta}_p - \hat{\eta}_{\mathrm{MLE}}\right) \tag{3}$$

and

$$V\left(\hat{\eta}^*_{imp,\infty}\right) \doteq \mathcal{I}^{-1}_{obs} + \mathcal{J}_{mis}\left\{V\left(\hat{\eta}_p\right) - V\left(\hat{\eta}_{\mathrm{MLE}}\right)\right\}\mathcal{J}'_{mis},$$

where $\mathcal{J}_{mis} = \mathcal{I}^{-1}_{com}\mathcal{I}_{mis}$ is the fraction of missing information.

Remark

- Equation (3) means that
$$\hat{\eta}^*_{imp,\infty} = (I - \mathcal{J}_{\rm mis})\,\hat{\eta}_{MLE} + \mathcal{J}_{\rm mis}\hat{\eta}_p. \tag{4}$$
That is, $\hat{\eta}^*_{imp,\infty}$ is a convex combination of $\hat{\eta}_{MLE}$ and $\hat{\eta}_p$.
- Note that $\hat{\eta}^*_{imp,\infty}$ is one-step EM update with initial estimate $\hat{\eta}_p$. Let $\hat{\eta}^{(t)}$ be the t-th EM update of η that is computed by solving
$$\bar{S}\left(\eta \mid \hat{\eta}^{(t-1)}\right) = 0$$
with $\hat{\eta}^{(0)} = \hat{\eta}_p$. Equation (4) implies that
$$\hat{\eta}^{(t)} = (I - \mathcal{J}_{\rm mis})\,\hat{\eta}_{MLE} + \mathcal{J}_{\rm mis}\hat{\eta}^{(t-1)}.$$
- Thus, we can obtain
$$\hat{\eta}^{(t)} = \hat{\eta}_{MLE} + (\mathcal{J}_{\rm mis})^{t-1}\left(\hat{\eta}^{(0)} - \hat{\eta}_{MLE}\right),$$
which justifies $\lim_{t\to\infty}\hat{\eta}^{(t)} = \hat{\eta}_{MLE}$.

Proof for Lemma 4.2

Wang and Robins (1998)

Theorem 4.1 (Asymptotic results for $m < \infty$)

Let $\hat{\eta}_p$ be a preliminary $\sqrt{n}$-consistent estimator of η with variance V_p. Under some regularity conditions, the solution $\hat{\eta}^*_{imp,m}$ to

$$\bar{S}^*_m\left(\eta \mid \hat{\eta}_p\right) \equiv \frac{1}{m}\sum_{j=1}^{m} S_{\mathrm{com}}\left(\eta; \mathbf{y}_{obs}, \mathbf{y}^{*(j)}_{mis}, \boldsymbol{\delta}\right) = 0$$

has mean η_0 and asymptotic variance equal to

$$V\left(\hat{\eta}^*_{imp,m}\right) \doteq \mathcal{I}^{-1}_{\mathrm{obs}} + \mathcal{J}_{\mathrm{mis}}\left\{V_p - \mathcal{I}^{-1}_{\mathrm{obs}}\right\}\mathcal{J}'_{\mathrm{mis}} + m^{-1}\mathcal{I}^{-1}_{\mathrm{com}}\mathcal{I}_{\mathrm{mis}}\mathcal{I}^{-1}_{\mathrm{com}} \qquad (5)$$

where $\mathcal{J}_{\mathrm{mis}} = \mathcal{I}^{-1}_{\mathrm{com}}\mathcal{I}_{\mathrm{mis}}$.

Remark

- If we use $\hat{\eta}_p = \hat{\eta}_{MLE}$, then the asymptotic variance in (5) is

$$V\left(\hat{\eta}^*_{imp,m}\right) \doteq \mathcal{I}_{\mathrm{obs}}^{-1} + m^{-1}\mathcal{I}_{\mathrm{com}}^{-1}\mathcal{I}_{\mathrm{mis}}\mathcal{I}_{\mathrm{com}}^{-1}.$$

- In Bayesian imputation (or multiple imputation), the posterior values of η are independently generated from $\eta \sim N(\hat{\eta}_{MLE}, \mathcal{I}_{obs}^{-1})$, which implies that we can use $V_p = \mathcal{I}_{\mathrm{obs}}^{-1} + m^{-1}\mathcal{I}_{\mathrm{obs}}^{-1}$. Thus, the asymptotic variance in (5) for multiple imputation is

$$V\left(\hat{\eta}^*_{imp,m}\right) \doteq \mathcal{I}_{\mathrm{obs}}^{-1} + m^{-1}\mathcal{J}_{\mathrm{mis}}\mathcal{I}_{\mathrm{obs}}^{-1}\mathcal{J}_{\mathrm{mis}}' + m^{-1}\mathcal{I}_{\mathrm{com}}^{-1}\mathcal{I}_{\mathrm{mis}}\mathcal{I}_{\mathrm{com}}^{-1}.$$

The second term is the additional price we pay when generating the posterior values, rather than using $\hat{\eta}_{MLE}$ directly.

Remark (about Theorem 4.1)

- Variance term (5) has three components. Writing

$$\hat{\eta}^*_{\text{imp,m}} = \hat{\eta}_{MLE} + \left(\hat{\eta}^*_{\text{imp},\infty} - \hat{\eta}_{MLE}\right) + \left(\hat{\eta}^*_{\text{imp,m}} - \hat{\eta}^*_{\text{imp},\infty}\right),$$

  we can establish that the three terms are independent and satisfies

$$\begin{aligned} V\left(\hat{\eta}_{MLE}\right) &= \mathcal{I}_{\text{obs}}^{-1} \\ V\left(\hat{\eta}^*_{\text{imp},\infty} - \hat{\eta}_{MLE}\right) &= \mathcal{J}_{\text{mis}}\left\{V_p - \mathcal{I}_{\text{obs}}^{-1}\right\}\mathcal{J}_{\text{mis}}' \\ V\left(\hat{\eta}^*_{\text{imp,m}} - \hat{\eta}^*_{\text{imp},\infty}\right) &= m^{-1}\mathcal{I}_{\text{com}}^{-1}\mathcal{I}_{\text{mis}}\mathcal{I}_{\text{com}}^{-1} \end{aligned}$$

Sketched proof for Theorem 4.1

- By Lemma 4.1 applied to $\bar{S}(\eta \mid \hat{\eta}_p) = 0$, we have
$$\hat{\eta}^*_{\text{imp},\infty} - \eta_0 \cong \mathcal{I}_{\text{com}}^{-1} \bar{S}(\eta \mid \hat{\eta}_p).$$
Similarly, we can write
$$\hat{\eta}^*_{\text{imp},m} - \eta_0 \cong \mathcal{I}_{\text{com}}^{-1} \bar{S}^*_m(\eta \mid \hat{\eta}_p).$$
- Thus,
$$V\left(\hat{\eta}^*_{\text{imp,m}} - \hat{\eta}^*_{\text{imp},\infty}\right) = \mathcal{I}_{\text{com}}^{-1} V\left\{\bar{S}^*_m(\eta \mid \hat{\eta}_p) - \bar{S}(\eta \mid \hat{\eta}_p)\right\} \mathcal{I}_{\text{com}}^{-1}$$
and
$$\begin{aligned} V\left\{\bar{S}^*_m(\eta \mid \hat{\eta}_p) - \bar{S}(\eta \mid \hat{\eta}_p)\right\} &= \frac{1}{m} V\left\{S_{\text{com}}(\eta) \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \hat{\eta}_p\right\} \\ &= \frac{1}{m} V\left\{S_{\text{mis}}(\eta) \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \hat{\eta}_p\right\} \\ &= \frac{1}{m} E\left\{S_{\text{mis}}(\eta)^{\otimes 2} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \hat{\eta}_p\right\} \\ &\cong \frac{1}{m} E\left\{S_{\text{mis}}(\eta)^{\otimes 2} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \eta\right\}. \end{aligned}$$

2. Basic Theory for Imputation

Basic Setup (for Case 2: $\psi \neq \eta$)

- Parameter ψ defined by $E\{U(\psi; \mathbf{y})\} = 0$.
- Under complete response, a consistent estimator of ψ can be obtained by solving $U(\psi; \mathbf{y}) = 0$.
- Assume that some part of $\mathbf{y}$, denoted by $\mathbf{y}_{\mathrm{mis}}$, is not observed and m imputed values, say $\mathbf{y}_{\mathrm{mis}}^{*(1)}, \cdots, \mathbf{y}_{\mathrm{mis}}^{*(m)}$, are generated from $f(\mathbf{y}_{\mathrm{mis}} \mid \mathbf{y}_{\mathrm{obs}}, \boldsymbol{\delta}; \hat{\eta}_{MLE})$, where $\hat{\eta}_{MLE}$ is the MLE of η_0.
- The imputed estimating function using m imputed values is computed as

$$\bar{U}_{imp,m}^{*}(\psi \mid \hat{\eta}_{MLE}) = \frac{1}{m}\sum_{j=1}^{m} U(\psi; \mathbf{y}^{*(j)}), \tag{6}$$

  where $\mathbf{y}^{*(j)} = (\mathbf{y}_{\mathrm{obs}}, \mathbf{y}_{\mathrm{mis}}^{*(j)})$.
- Let $\hat{\psi}_{imp,m}^{*}$ be the solution to $\bar{U}_{imp,m}^{*}(\psi \mid \hat{\eta}_{MLE}) = 0$. We are interested in the asymptotic properties of $\hat{\psi}_{imp,m}^{*}$.

Theorem 4.2

Theorem 4.2

Suppose that the parameter of interest ψ_0 is estimated by solving $U(\psi) = 0$ under complete response. Then, under some regularity conditions, the solution to

$$E\left\{U(\psi) \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}; \hat{\eta}_{MLE}\right\} = 0 \tag{7}$$

has mean ψ_0 and the asymptotic variance $\tau^{-1}\Omega\tau'^{-1}$, where

$$\begin{aligned} \tau &= -E\left\{\partial U(\psi_0)/\partial\psi'\right\} \\ \Omega &= V\left\{\bar{U}(\psi_0 \mid \eta_0) + \kappa S_{\rm obs}(\eta_0)\right\} \end{aligned}$$

and

$$\kappa = E\left\{U(\psi_0)\, S_{\rm mis}(\eta_0)\right\} \mathcal{I}_{\rm obs}^{-1}.$$

Sketched Proof

- Writing

$$\bar{U}(\psi \mid \eta) \equiv E\{U(\psi) \mid \mathbf{y}_{obs}, \boldsymbol{\delta}; \eta\},$$

  the solution to (7) can be treated as the solution to the joint estimating equation

$$\mathbf{U}(\psi, \eta) \equiv \begin{bmatrix} U_1(\psi, \eta) \\ U_2(\eta) \end{bmatrix} = \mathbf{0},$$

  where $U_1(\psi, \eta) = \bar{U}(\psi \mid \eta)$ and $U_2(\eta) = S_{\mathrm{obs}}(\eta)$.

- We can apply the Taylor expansion to get

$$\begin{pmatrix} \hat{\psi} \\ \hat{\eta} \end{pmatrix} \cong \begin{pmatrix} \psi_0 \\ \eta_0 \end{pmatrix} - \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}^{-1} \begin{bmatrix} U_1(\psi_0, \eta_0) \\ U_2(\eta_0) \end{bmatrix}$$

  where

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{bmatrix} E(\partial U_1/\partial \psi') & E(\partial U_1/\partial \eta') \\ E(\partial U_2/\partial \psi') & E(\partial U_2/\partial \eta') \end{bmatrix}.$$

Sketched Proof (Cont'd)

- Note that

$$\begin{array}{rcl} B_{11} & = & E\{\partial U(\psi)/\partial \psi'\} \\ B_{21} & = & 0 \\ B_{12} & = & E\{U(\psi)S_{mis}(\eta_0)\} \\ B_{22} & = & -\mathcal{I}_{\rm obs} \end{array}$$

- Thus,

$$\hat{\psi} \cong \psi_0 - B_{11}^{-1}\left\{U_1(\psi_0,\eta_0) - B_{12}B_{22}^{-1}U_2(\eta_0)\right\}.$$

Alternative approach

Use Randles (1982) theorem: An estimator $\hat{\theta}(\hat{\beta})$ is asymptotically equivalent to $\hat{\theta}(\beta)$ if

$$E\left\{\frac{\partial}{\partial\beta}\hat{\theta}(\beta)\right\}=0.$$

Thus, writing $\hat{\theta}_k(\beta)=\hat{\theta}(\beta)+kS_{obs}(\beta)$, we have

1. $\hat{\theta}(\hat{\beta})=\hat{\theta}_k(\hat{\beta})$ holds for any k.
2. If we can find $k=k^*$ such that $\hat{\theta}_{k^*}(\hat{\beta})$ satisfies Randles' condition, then we can safely ignore the effect of the sampling variability of $\hat{\beta}$ and assume that $\hat{\theta}_{k^*}(\hat{\beta})\cong\hat{\theta}_{k^*}(\beta)$.

Example 4.2

- Under the setup of Example 4.1, we are interested in obtaining the asymptotic variance of the regression imputation estimator

$$\hat{\theta}_{Id} = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i y_i + (1-\delta_i)\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\},$$

where $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1\right)$ is the solution to

$$S_{\text{obs}}(\beta) = \frac{1}{\sigma_e^2}\sum_{i=1}^{n}\delta_i(y_i - \beta_0 - \beta_1 x_i)(1, x_i)'.$$

- The imputed estimator is the solution to

$$E\left\{U(\theta) \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \hat{\beta}\right\} = 0$$

where

$$U(\theta) = \sum_{i=1}^{n}\left(y_i - \theta\right)/\sigma_e^2.$$

Example 4.2 (Cont'd)

- Since

$$
\begin{aligned}
E\left\{U(\theta) \mid \mathbf{y}_{\mathrm{obs}}, \boldsymbol{\delta}; \beta\right\} &\equiv \bar{U}(\theta \mid \beta) \\
&= \sum_{i=1}^{n}\left\{\delta_i y_i + (1-\delta_i)\left(\beta_0 + \beta_1 x_i\right) - \theta\right\}/\sigma_e^2 .
\end{aligned}
$$

- Thus, using the linearization formula in Theorem 4.2, we have

$$\bar{U}_l(\theta \mid \beta) = \bar{U}(\theta \mid \beta) + (\kappa_1, \kappa_2) S_{\mathrm{obs}}(\beta) \tag{8}$$

where

$$(\kappa_1, \kappa_2)' = \mathcal{I}_{\mathrm{obs}}^{-1} E\left\{S_{\mathrm{mis}}(\beta) U(\theta)\right\}. \tag{9}$$

Example 4.2 (Cont'd)

- In this example, we have

$$
\begin{aligned}
\begin{pmatrix} \kappa_0 \\ \kappa_1 \end{pmatrix} &= \left[E\left\{ \sum_{i=1}^{n} \delta_i (1, x_i)(1, x_i)' \right\} \right]^{-1} E\left\{ \sum_{i=1}^{n} (1-\delta_i)(1, x_i)' \right\} \\
&\cong E\left\{ (-1 + (n/r)(1 - g\bar{x}_r), (n/r)g)' \right\},
\end{aligned}
$$

where $g = (\bar{x}_n - \bar{x}_r) / \sum_{i=1}^{n} \delta_i (x_i - \bar{x}_r)^2 / r$. Thus,

$$
\begin{aligned}
\bar{U}_l(\theta \mid \beta)\,\sigma_e^2 &= \sum_{i=1}^{n} \delta_i (y_i - \theta) + \sum_{i=1}^{n} (1-\delta_i)(\beta_0 + \beta_1 x_i - \theta) \\
&\quad + \sum_{i=1}^{n} \delta_i (y_i - \beta_0 - \beta_1 x_i)(\kappa_0 + \kappa_1 x_i).
\end{aligned}
$$

Example 4.2 (Cont'd)

- Note that the solution to $\bar{U}_I(\theta \mid \beta) = 0$ leads to

$$\begin{aligned} \hat{\theta}_{Id,I} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\beta_0 + \beta_1 x_i + \delta_i(1+\kappa_0+\kappa_1 x_i)\left(y_i - \beta_0 - \beta_1 x_i\right)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n} d_i, \end{aligned}$$

  where $1+\kappa_0+\kappa_1 x_i = (n/r)\{1+g(x_i-\bar{x}_r)\}$.

- Thus, $\hat{\theta}_{Id}$ is asymptotically equivalent to $\hat{\theta}_{Id,I}$, which is the sample mean of d_i, the influence function of unit i to $\hat{\theta}_{Id}$.
- Under uniform response mechanism, $1+\kappa_0+\kappa_1 x_i \cong n/r$ and the asymptotic variance of $\hat{\theta}_I$ is equal to

$$\frac{1}{n}\beta_1^2\sigma_x^2 + \frac{1}{r}\sigma_e^2 = \frac{1}{n}\sigma_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\sigma_e^2$$

Remark

- Let $X_1, \cdots, X_n$ be IID sample from $f(x;\theta_0), \theta_0 \in \Theta$ and we are interested in estimating $\gamma_0 = \gamma(\theta_0)$, where $\gamma(\cdot) : \Theta \to R^k$. An estimator $\hat{\gamma} = \hat{\gamma}_n$ is called asymptotically linear if there exist a random vector $\psi(x)$ such that

$$\sqrt{n}\left(\hat{\gamma}_n - \gamma_0\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) + o_p(1) \tag{10}$$

  with $E_{\theta_0}\{\psi(X)\} = 0$ and $E_{\theta_0}\{\psi(X)\psi(X)'\}$ is finite and non-singular. Here, $Z_n = o_p(1)$ means that Z_n converges to zero in probability.

Remark

- The function $\psi(x)$ is referred to as an influence function. The phrase influence function was used by Hampel (JASA, 1974) and is motivated by the fact that to the first order $\psi(x)$ is the influence of a single observation on the estimator $\hat{\gamma} = \hat{\gamma}(X_1, \cdots, X_n)$.
- The asymptotic properties of an asymptotically linear estimator, $\hat{\gamma}_n$ can be summarized by considering only its influence function.
- Since $\psi(X)$ has zero mean, the CLT tells us that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(X_i) \xrightarrow{L} N\left[0, E_{\theta_0}\{\psi(X)\psi(X)'\}\right]. \tag{11}$$

Thus, combining (10) with (11) and applying Slutsky's theorem, we have

$$\sqrt{n}\,(\hat{\gamma}_n - \gamma_0) \xrightarrow{L} N\left[0, E_{\theta_0}\{\psi(X)\psi(X)'\}\right].$$

§3 Variance estimation after imputation

Deterministic Imputation

- In Example 4.2, the imputed estimator can be written as $\hat{\theta}_{Id}(\hat{\beta})$. Note that we can write the deterministic imputed estimator as

$$\hat{\theta}_{Id} = n^{-1} \sum_{i=1}^{n} \hat{y}_i(\hat{\beta}),$$

  where $\hat{y}_i(\hat{\beta}) = \hat{\beta}_0 + \hat{\beta}_1 x_i$.
- In general, the asymptotic variance of $\hat{\theta}_{Id} = \hat{\theta}(\hat{\beta})$ is different from the asymptotic variance of $\hat{\theta}(\beta)$

- As in Example 4.2, if we can find $d_i = d_i(\beta)$ such that

$$\hat{\theta}_{Id}(\hat{\beta}) = n^{-1}\sum_{i=1}^{n} d_i(\hat{\beta}) \cong n^{-1}\sum_{i=1}^{n} d_i(\beta),$$

  then the asymptotic variance of $\hat{\theta}_{Id}$ is equal to the asymptotic variance of $\bar{d}_n = n^{-1}\sum_{i=1}^{n} d_i(\beta)$.
- Note that, if (x_i, y_i, δ_i) are IID, then $d_i = d(x_i, y_i, \delta_i)$ are also IID. Thus, the variance of $\bar{d}_n = n^{-1}\sum_{i=1}^{n} d_i$ is unbiasedly estimated by

$$\hat{V}(\bar{d}_n) = \frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}_n\right)^2. \tag{12}$$

  Unfortunately, we cannot compute $\hat{V}(\bar{d}_n)$ in (12) since $d_i = d_i(\beta)$ is a function of unknown parameters. Thus, we use $\hat{d}_i = d_i(\hat{\beta})$ in (12) to get a consistent variance estimator of the imputed estimator.

Stochastic Imputation

- Instead of the deterministic imputation, suppose that a stochastic imputation is used such that

$$\hat{\theta}_I = n^{-1} \sum_{i=1}^{n} \left\{ \delta_i y_i + (1-\delta_i)\left(\hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{e}_i^* \right) \right\},$$

  where $\hat{e}_i^*$ are the additional noise terms in the stochastic imputation. Often $\hat{e}_i^*$ are randomly selected from the empirical distribution of the sample residuals in the respondents.

- The variance of the imputed estimator can be decomposed into two parts:

$$V\left(\hat{\theta}_I\right) = V\left(\hat{\theta}_{Id}\right) + V\left(\hat{\theta}_I - \hat{\theta}_{Id}\right) \quad (13)$$

  where the first part is the deterministic part and the second part is the additional variance due to stochastic imputation. The first part can be estimated by the linearization method discussed above. The second part is called the imputation variance.

- If we require the imputation mechanism to satisfy

$$\sum_{i=1}^{n} (1-\delta_i)\,\hat{e}_i^* = 0$$

then the imputation variance is equal to zero.

- Often the variance of $\hat{\theta}_I - \hat{\theta}_{Id} = n^{-1}\sum_{i=1}^{n}(1-\delta_i)\,\hat{e}_i^*$ can be computed under the known imputation mechanism. For example, if simple random sampling without replacement is used then

$$V\left(\hat{\theta}_I - \hat{\theta}_{Id}\right) = E\left\{V\left(\hat{\theta}_I \mid \mathbf{y}_{obs}, \boldsymbol{\delta}\right)\right\} = n^{-2}(1-m/r)\,(r-1)^{-1}\sum_{i=1}^{n}\delta_i\hat{e}_i^2$$

where $m = n - r$.

Imputed estimator for general parameters

- We now discuss a general case of parameter estimation when the parameter of interest ψ is estimated by the solution $\hat{\psi}_n$ to

$$\sum_{i=1}^{n} U(\psi; \mathbf{y}_i) = 0 \tag{14}$$

  under complete response of $\mathbf{y}_1, \cdots, \mathbf{y}_n$.

- Under the existence of missing data, we can use the imputed estimating equation

$$\bar{U}_m^*(\psi) \equiv m^{-1} \sum_{i=1}^{n} \sum_{j=1}^{m} U(\psi; \mathbf{y}_i^{*(j)}) = 0, \tag{15}$$

  where $\mathbf{y}_i^{*(j)} = (\mathbf{y}_{i,\mathrm{obs}}, \mathbf{y}_{i,mis}^{*(j)})$ and $\mathbf{y}_{i,mis}^{*(j)}$ are randomly generated from the conditional distribution $h(\mathbf{y}_{i,\mathrm{mis}} \mid \mathbf{y}_{i,\mathrm{obs}}, \boldsymbol{\delta}_i; \hat{\eta}_p)$ where $\hat{\eta}_p$ is estimated by solving

$$\hat{U}_p(\eta) \equiv \sum^{n} U_p(\eta; \mathbf{y}_{i,\mathrm{obs}}) = 0. \tag{16}$$

- To apply the linearization method, we first compute the conditional expectation of $U(\psi; \mathbf{y}_i)$ given $(\mathbf{y}_{i,\mathrm{obs}}, \boldsymbol{\delta}_i)$ evaluated at $\hat{\eta}_p$. That is, compute

$$\bar{U}\left(\psi \mid \hat{\eta}_p\right) = \sum_{i=1}^{n} \bar{U}_i\left(\psi \mid \hat{\eta}_p\right) = \sum_{i=1}^{n} E\left\{U(\psi; \mathbf{y}_i) \mid \mathbf{y}_{i,\mathrm{obs}}, \boldsymbol{\delta}_i; \hat{\eta}_p\right\}. \quad (17)$$

- Let $\hat{\psi}_R$ be the solution to $\bar{U}\left(\psi \mid \hat{\eta}_p\right) = 0$. Using the linearization technique, we have

$$\bar{U}\left(\psi \mid \hat{\eta}_p\right) \cong \bar{U}\left(\psi \mid \eta_0\right) + E\left\{\frac{\partial}{\partial \eta'} \bar{U}\left(\psi \mid \eta_0\right)\right\}\left(\hat{\eta}_p - \eta_0\right) \quad (18)$$

and

$$0 = \hat{U}_p\left(\hat{\eta}_p\right) = \hat{U}_p\left(\eta_0\right) + E\left\{\frac{\partial}{\partial \eta'} \hat{U}_p\left(\eta_0\right)\right\}\left(\hat{\eta}_p - \eta_0\right). \quad (19)$$

- Thus, combining (18) and (19), we have

$$\bar{U}\left(\psi \mid \hat{\eta}_p\right) \cong \bar{U}\left(\psi \mid \eta_0\right) + \kappa(\psi)\hat{U}_p\left(\eta_0\right) \tag{20}$$

  where

$$\kappa(\psi) = -E\left\{\frac{\partial}{\partial \eta'}\bar{U}\left(\psi \mid \eta_0\right)\right\}\left[E\left\{\frac{\partial}{\partial \eta'}\hat{U}_p\left(\eta_0\right)\right\}\right]^{-1}.$$

- Thus, writing

$$\bar{U}_l\left(\psi \mid \eta_0\right) = \sum_{i=1}^{n}\left\{\bar{U}_i\left(\psi \mid \eta_0\right) + \kappa(\psi)\hat{U}_p\left(\eta_0; \mathbf{y}_{i,\mathrm{obs}}\right)\right\} = \sum_{i=1}^{n} q_i\left(\psi \mid \eta_0\right),$$

  and $q_i\left(\psi \mid \eta_0\right) = \bar{U}_i\left(\psi \mid \eta_0\right) + \kappa(\psi)\hat{U}_p\left(\eta_0; \mathbf{y}_{i,\mathrm{obs}}\right)$, the variance of $\bar{U}\left(\psi \mid \hat{\eta}_p\right)$ is asymptotically equal to the variance of $\bar{U}_l\left(\psi \mid \eta_0\right)$.

- Thus, the sandwich-type variance estimator for $\hat{\psi}_R$ is

$$\hat{V}\left(\hat{\psi}_R\right) = \hat{\tau}_q^{-1}\hat{\Omega}_q\hat{\tau}_q^{-1'} \tag{21}$$

  where

$$\begin{aligned} \hat{\tau}_q &= n^{-1}\sum_{i=1}^{n} \dot{q}_i\left(\hat{\psi}_R \mid \hat{\eta}_p\right) \\ \hat{\Omega}_q &= n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\left(\hat{q}_i - \bar{q}_n\right)^{\otimes 2}, \end{aligned}$$

  $\dot{q}_i(\psi \mid \eta) = \partial q_i(\psi \mid \eta)/\partial\psi$, $\bar{q}_n = n^{-1}\sum_{i=1}^{n}\hat{q}_i$, and $\hat{q}_i = q_i(\hat{\psi}_R \mid \hat{\eta}_p)$.

- Note that

$$\begin{aligned} \hat{\tau}_q &= n^{-1}\sum_{i=1}^{n} \dot{q}_i\left(\hat{\psi}_R \mid \hat{\eta}_p\right) \\ &= n^{-1}\sum_{i=1}^{n} E\left\{\dot{U}(\hat{\psi}_R; \mathbf{y}_i) \mid \mathbf{y}_{i,\mathrm{obs}}, \boldsymbol{\delta}_i; \hat{\eta}_p\right\} \end{aligned}$$

  because $\hat{\eta}_p$ is the solution to (16).

Example 4.3

- Assume that the original sample is decomposed into G disjoint groups (often called imputation cells) and the sample observations are IID within the same cell. That is,

$$y_i \mid i \in S_g \overset{i.i.d.}{\sim} \left(\mu_g, \sigma_g^2\right) \tag{22}$$

  where S_g is the set of sample indices in cell g. Assume that n_g sample elements in cell g and r_g elements are observed in the cell.

- For deterministic imputation, let $\hat{\mu}_g = r_g^{-1} \sum_{i \in S_g} \delta_i y_i$ be the g-th cell mean of y among the respondents. The (deterministically) imputed estimator of $\theta = E(Y)$ is, under MAR,

$$\hat{\theta}_{Id} = n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} \{\delta_i y_i + (1 - \delta_i)\hat{\mu}_g\} = n^{-1} \sum_{g=1}^{G} n_g \hat{\mu}_g . \tag{23}$$

Example 4.3 (Cont'd)

- Using the linearization technique in (20), the imputed estimator can be expressed as

$$\hat{\theta}_{Id} \cong n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} \left\{ \mu_g + \frac{n_g}{r_g} \delta_i (y_i - \mu_g) \right\} \tag{24}$$

and the plug-in variance estimator can be expressed as

$$\hat{V}(\hat{\theta}_{Id}) = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{d}_i - \bar{d}_n \right)^2 \tag{25}$$

where $\hat{d}_i = \hat{\mu}_g + (n_g / r_g) \delta_i (y_i - \hat{\mu}_g)$ and $\bar{d}_n = n^{-1} \sum_{i=1}^{n} \hat{d}_i$.

Example 4.3 (Cont'd)

- If a stochastic imputation is used where an imputed value is randomly selected from the set of respondents in the same cell, then we can write
$$\hat{\theta}_{Is} = n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} \left\{ \delta_i y_i + (1 - \delta_i) y_i^* \right\}. \tag{26}$$
Writing $\hat{\theta}_{Is} = \hat{\theta}_{Id} + n^{-1} \sum_{g=1}^{G} \sum_{i \in S_g} (1 - \delta_i) \left(y_i^* - \hat{\mu}_g \right)$, the variance of the second term can be estimated by $n^{-2} \sum_{g=1}^{G} \sum_{i \in S_g} (1 - \delta_i) \left(y_i^* - \hat{\mu}_g \right)^2$ if the imputed values are generated independently, conditional on the respondents.

§4.4 Replication variance estimation

Replication variance estimation (under complete response)

Let $\hat{\theta}_n$ be the complete-sample estimator of θ. The replication variance estimator of $\hat{\theta}_n$ takes the form of

$$\hat{V}_{rep}(\hat{\theta}_n) = \sum_{k=1}^{L} c_k \left(\hat{\theta}_n^{(k)} - \hat{\theta}_n \right)^2 \tag{27}$$

where L is the number of replicates, c_k is the replication factor associated with replication k, and $\hat{\theta}_n^{(k)}$ is the k-th replicate of $\hat{\theta}_n$. If $\hat{\theta}_n = \sum_{i=1}^n y_i/n$, then we can write $\hat{\theta}_n^{(k)} = \sum_{i=1}^n w_i^{(k)} y_i$ for some replication weights $w_1^{(k)}, w_2^{(k)}, \cdots, w_n^{(k)}$.

Replication variance estimation (under complete response)

- For example, in the jackknife method, we have $L = n$, $c_k = (n-1)/n$, and

$$w_i^{(k)} = \begin{cases} (n-1)^{-1} & \text{if } i \neq k \\ 0 & \text{if } i = k. \end{cases}$$

If we use the above jackknife method to $\hat{\theta}_n = \sum_{i=1}^{n} y_i/n$, the resulting jackknife estimator in (27) is algebraically equivalent to $n^{-1}(n-1)^{-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$.

Replication variance estimation (under complete response)

- Under some regularity conditions, for $\hat{\theta}_n = g(\bar{y}_n)$, the replication variance estimator of $\hat{\theta}_n$, defined by

$$\hat{V}_{rep}\left(\hat{\theta}_n\right) = \sum_{k=1}^{L} c_k \left(\hat{\theta}_n^{(k)} - \hat{\theta}_n\right)^2, \tag{28}$$

where $\hat{\theta}_n^{(k)} = g(\bar{y}_n^{(k)})$, satisfies

$$\hat{V}_{rep}\left(\hat{\theta}_n\right) \cong \left\{g'(\bar{y}_n)\right\}^2 \hat{V}_{rep}(\bar{y}_n).$$

Thus, the replication variance estimator (28) is asymptotically equivalent to the linearized variance estimator.

Remark

- If the parameter of interest, denoted by ψ, is estimated by $\hat{\psi}$ which is obtained by solving an estimating equation $\sum_{i=1}^{n} U(\psi; y_i) = 0$, then a consistent variance estimator can be obtained by the sandwich formula: The complete-sample variance estimator of $\hat{\psi}$ is

$$\hat{V}\left(\hat{\psi}\right) = \hat{\tau}_u^{-1} \hat{\Omega}_u \hat{\tau}_u^{-1\prime} \tag{29}$$

where

$$\begin{aligned} \hat{\tau}_u &= n^{-1} \sum_{i=1}^{n} \dot{U}\left(\hat{\psi}; \mathbf{y}_i\right) \\ \hat{\Omega}_u &= n^{-1} (n-1)^{-1} \sum_{i=1}^{n} \left(\hat{u}_i - \bar{u}_n\right)^{\otimes 2}, \end{aligned}$$

$\dot{U}(\psi; \mathbf{y}) = \partial U(\psi; \mathbf{y}) / \partial \psi$, $\bar{u}_n = n^{-1} \sum_{i=1}^{n} \hat{u}_i$, and $\hat{u}_i = U(\hat{\psi}; \mathbf{y}_i)$.

- If we want to use the replication method of the form (27), we can construct the replication variance estimator of $\hat{\psi}$ by

$$\hat{V}_{rep}(\hat{\psi}) = \sum_{k=1}^{L} c_k \left(\hat{\psi}^{(k)} - \hat{\psi} \right)^2 \tag{30}$$

  where $\hat{\psi}^{(k)}$ is computed by

$$\hat{U}^{(k)}(\psi) \equiv \sum_{i=1}^{n} w_i^{(k)} U(\psi; y_i) = 0. \tag{31}$$

- One-step approximation of $\hat{\psi}^{(k)}$ is to use

$$\hat{\psi}_1^{(k)} = \hat{\psi} - \left\{ \dot{U}^{(k)}(\hat{\psi}) \right\}^{-1} \hat{U}^{(k)}(\hat{\psi}) \tag{32}$$

or, even more simply, to use

$$\hat{\psi}_1^{(k)} = \hat{\psi} - \left\{ \dot{U}(\hat{\psi}) \right\}^{-1} \hat{U}^{(k)}(\hat{\psi}). \tag{33}$$

The replication variance estimator using (33) is algebraically equivalent to

$$\left\{ \dot{U}(\hat{\psi}) \right\}^{-1} \left[\sum_{k=1}^{n} c_k \left\{ \hat{U}^{(k)}(\hat{\psi}) - \hat{U}(\hat{\psi}) \right\}^{\otimes 2} \right] \left\{ \dot{U}(\hat{\psi}) \right\}^{-1},$$

which is very close to the sandwich variance formula in (29).

Back to Example 4.1

- For the regression imputation in Example 4.1,

$$\hat{\theta}_{Id} = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i y_i + (1-\delta_i)\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)\right\}.$$

- The replication variance estimator of $\hat{\theta}_{Id}$ is computed by

$$\hat{V}_{rep}\left(\hat{\theta}_{Id}\right) = \sum_{k=1}^{L} c_k \left(\hat{\theta}_{Id}^{(k)} - \hat{\theta}_{Id}\right)^2 \tag{34}$$

where

$$\hat{\theta}_{Id}^{(k)} = \sum_{i=1}^{n} w_i^{(k)}\left\{\delta_i y_i + (1-\delta_i)\left(\hat{\beta}_0^{(k)} + \hat{\beta}_1^{(k)} x_i\right)\right\}$$

and $(\hat{\beta}_0^{(k)}, \hat{\beta}_1^{(k)})$ is the solution to

$$\sum_{i=1}^{n} w_i^{(k)}\delta_i\left(y_i - \beta_0 - \beta_1 x_i\right)(1, x_i) = (0,0).$$

Example 4.4

- We now return to the setup of Example 3.11.
- In this case, the deterministically imputed estimator of $\theta = E(Y)$ is constructed by

$$\hat{\theta}_{Id} = n^{-1} \sum_{i=1}^{n} \{\delta_i y_i + (1 - \delta_i)\hat{p}_{0i}\} \tag{35}$$

where $\hat{p}_{0i}$ is the predictor of y_i given x_i and $\delta_i = 0$. That is,

$$\hat{p}_{0i} = \frac{p(x_i; \hat{\beta})\{1 - \pi(x_i, 1; \hat{\phi})\}}{\{1 - p(x_i; \hat{\beta})\}\{1 - \pi(x_i, 0; \hat{\phi})\} + p(x_i; \hat{\beta})\{1 - \pi(x_i, 1; \hat{\phi})\}},$$

where $\hat{\beta}$ and $\hat{\phi}$ are jointly estimated by the EM algorithm.

Example 4.4 (Cont'd)

E-step

$$\bar{S}_1\left(\beta \mid \beta^{(t)}, \phi^{(t)}\right) = \sum_{\delta_i=1}\left\{y_i - p_i(\beta)\right\}\mathbf{x}_i + \sum_{\delta_i=0}\sum_{j=0}^{1} w_{ij(t)}\left\{j - p_i(\beta)\right\}\mathbf{x}_i,$$

where

$$\begin{aligned} w_{ij(t)} &= Pr\left(Y_i = j \mid \mathbf{x}_i, \delta_i = 0; \beta^{(t)}, \phi^{(t)}\right) \\ &= \frac{Pr\left(Y_i = j \mid \mathbf{x}_i; \beta^{(t)}\right) Pr\left(\delta_i = 0 \mid \mathbf{x}_i, j; \phi^{(t)}\right)}{\sum_{y=0}^{1} Pr\left(Y_i = y \mid \mathbf{x}_i; \beta^{(t)}\right) Pr\left(\delta_i = 0 \mid \mathbf{x}_i, y; \phi^{(t)}\right)} \end{aligned}$$

and

$$\begin{aligned} \bar{S}_2\left(\phi \mid \beta^{(t)}, \phi^{(t)}\right) &= \sum_{\delta_i=1}\left\{\delta_i - \pi\left(\mathbf{x}_i, y_i; \phi\right)\right\}\left(\mathbf{x}_i', y_i\right)' \\ &\quad + \sum_{\delta_i=0}\sum_{j=0}^{1} w_{ij(t)}\left\{\delta_i - \pi_i\left(\mathbf{x}_i, j; \phi\right)\right\}\left(\mathbf{x}_i', j\right)'. \end{aligned}$$

Example 4.4 (Cont'd)

M-step

The parameter estimates are updated by solving

$$\left[\bar{S}_1\left(\beta \mid \beta^{(t)}, \phi^{(t)}\right), \bar{S}_2\left(\phi \mid \beta^{(t)}, \phi^{(t)}\right)\right] = (0, 0)$$

for β and ϕ.

Example 4.4 (Cont'd)

- For replication variance estimation, we can use (34) with

$$\hat{\theta}_{Id}^{(k)} = \sum_{i=1}^{n} w_i^{(k)} \left\{ \delta_i y_i + (1-\delta_i)\hat{p}_{0i}^{(k)} \right\}. \tag{36}$$

where

$$\hat{p}_{0i}^{(k)} = \frac{p(x_i;\hat{\beta}^{(k)})\{1-\pi(x_i,1;\hat{\phi}^{(k)})\}}{\{1-p(x_i;\hat{\beta}^{(k)})\}\pi(x_i,0;\hat{\phi}^{(k)}) + p(x_i;\hat{\beta}^{(k)})\{1-\pi(x_i,1;\hat{\phi}^{(k)})\}}.$$

and $(\hat{\beta}^{(k)}, \hat{\phi}^{(k)})$ is obtained by solving the mean score equations with weights replaced by the replication weights $w_i^{(k)}$.

Example 4.4 (Cont'd)

- That is, $(\hat{\beta}^{(k)}, \hat{\phi}^{(k)})$ is the solution to

$$\begin{aligned}
\bar{S}_1^{(k)}(\beta,\phi) &\equiv \sum_{\delta_i=1} w_i^{(k)} \{y_i - p(\mathbf{x}_i;\beta)\}\,\mathbf{x}_i \\
&\quad + \sum_{\delta_i=0} w_i^{(k)} \sum_{y=0}^{1} w_{iy}^*(\beta,\phi)\{y - p(\mathbf{x}_i;\beta)\}\mathbf{x}_i = 0 \\
\bar{S}_2^{(k)}(\beta,\phi) &\equiv \sum_{\delta_i=1} w_i^{(k)} \{\delta_i - \pi(\mathbf{x}_i, y_i;\phi)\}\,(\mathbf{x}_i', y_i)' \\
&\quad + \sum_{\delta_i=0} w_i^{(k)} \sum_{y=0}^{1} w_{iy}^*(\beta,\phi)\{\delta_i - \pi(\mathbf{x}_i, y;\beta)\}(\mathbf{x}_i', y)' = 0
\end{aligned}$$

and

$$w_{iy}^*(\beta,\phi) = \frac{p(\mathbf{x}_i;\beta)\pi(\mathbf{x}_i,1;\phi)}{\{1-p(\mathbf{x}_i;\beta)\}\pi(\mathbf{x}_i,0;\phi) + p(\mathbf{x}_i;\beta)\pi(\mathbf{x}_i,1;\phi)}.$$

§6 Fractional Imputation

Monte Carlo EM

Remark

- Monte Carlo EM can be used as a frequentist approach to imputation.
- Convergence is not guaranteed (for fixed m).
- E-step can be computationally heavy. (May use MCMC method).

Parametric Fractional Imputation (Kim, 2011)

Parametric fractional imputation

1. More than one (say m) imputed values of $\mathbf{y}_{mis,i}$: $\mathbf{y}_{mis,i}^{*(1)}, \cdots, \mathbf{y}_{mis,i}^{*(m)}$ from some (initial) density $h\left(\mathbf{y}_{mis,i}\right)$.
2. Create weighted data set

$$\left\{\left(w_{ij}^{*}, \mathbf{y}_{ij}^{*}\right); j = 1, 2, \cdots, m; i = 1, 2 \cdots, n\right\}$$

where $\sum_{j=1}^{m} w_{ij}^{*} = 1$, $\mathbf{y}_{ij}^{*} = (\mathbf{y}_{obs,i}, \mathbf{y}_{mis,i}^{*(j)})$

$$w_{ij}^{*} \propto f(\mathbf{y}_{ij}^{*}, \boldsymbol{\delta}_i; \hat{\eta})/h(\mathbf{y}_{mis,i}^{*(j)}),$$

$\hat{\eta}$ is the maximum likelihood estimator of η, and $f\left(\mathbf{y}, \boldsymbol{\delta}; \eta\right)$ is the joint density of $(\mathbf{y}, \boldsymbol{\delta})$.

3. The weight w_{ij}^{*} are the normalized importance weights and can be called fractional weights.

Remark

- Importance sampling idea: For sufficiently large m,

$$\sum_{j=1}^{m} w_{ij}^{*} g\left(y_{ij}^{*}\right) \cong \frac{\int g(y_i) \frac{f(y_i,\delta_i;\hat{\eta})}{h(y_{mis,i})} h(y_{mis,i}) dy_{mis,i}}{\int \frac{f(y_i,\delta_i;\hat{\eta})}{h(y_{mis,i})} h(y_{mis,i}) dy_{mis,i}} = E\left\{g\left(y_i\right) \mid y_{obs,i}, \delta_i; \hat{\eta}\right\}$$

  for any g such that the expectation exists.

- In the importance sampling literature, $h(\cdot)$ is called proposal distribution and $f(\cdot)$ is called target distribution.
- Do not need to compute the conditional distribution $f(y_{mis,i} \mid y_{obs,i}, \delta_i; \eta)$. Only the joint distribution $f(y_{obs,i}, y_{mis,i}, \delta_i; \eta)$ is needed because

$$\frac{f(y_{obs,i}, y_{mis,i}^{*(j)}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(j)})}{\sum_{k=1}^{m} f(y_{obs,i}, y_{mis,i}^{*(k)}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(k)})} = \frac{f(y_{mis,i}^{*(j)} \mid y_{obs,i}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(j)})}{\sum_{k=1}^{m} f(y_{mis,i}^{*(k)} \mid y_{obs,i}, \delta_i; \hat{\eta})/h(y_{i,mis}^{*(k)})}.$$

EM algorithm by fractional imputation

1. Imputation-step: generate $y_{i,\mathrm{mis}}^{*(j)} \sim h\left(y_{i,mis}\right)$.
2. Weighting-step: compute

$$w_{ij(t)}^{*} \propto f\left(y_{ij}^{*}, \delta_i; \hat{\eta}_{(t)}\right) / h(y_{i,mis}^{*(j)})$$

   where $\sum_{j=1}^{m} w_{ij(t)}^{*} = 1$.

3. M-step: update

$$\hat{\eta}^{(t+1)} = \arg\max \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij(t)}^{*} \log f\left(\eta; y_{ij}^{*}, \delta_i\right).$$

4. (Optional) Check if $w_{ij(t)}^{*}$ is too large for some j. If so, set $h(y_{i,mis}) = f(y_{i,\mathrm{mis}} \mid y_{i,\mathrm{obs}}; \hat{\eta}_t)$ and goto Step 1.
5. Repeat Step 2 - Step 4 until convergence.

Remark

- “Imputation Step” + “Weighting Step” = E-step.
- The imputed values are not changed for each EM iteration. Only the fractional weights are changed.
  1. Computationally efficient (because we use importance sampling only once).
  2. Convergence is achieved (because the imputed values are not changed). See Theorem 4.5.
- For sufficiently large t, $\hat{\eta}^{(t)} \longrightarrow \hat{\eta}^*$. Also, for sufficiently large m, $\hat{\eta}^* \longrightarrow \hat{\eta}_{MLE}$.
- For estimation of ψ in $E\{U(\psi; Y)\} = 0$, simply use

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^* U(\psi; \mathbf{y}_{ij}^*) = 0$$

where $w_{ij}^* = w_{ij}^*(\hat{\eta})$ and $\hat{\eta}$ is obtained from the above EM algorithm.

Theorem 4.5 (Theorem 1 of Kim (2011))

Theorem

Let

$$Q^*(\eta \mid \hat{\eta}_{(t)}) = \sum_{i=1}^{n} \sum_{j=1}^{m} w^*_{ij(t)} \log f \left(\eta; y^*_{ij}, \delta_i \right).$$

If

$$Q^*(\hat{\eta}_{(t+1)} \mid \hat{\eta}_{(t)}) \geq Q^*(\hat{\eta}_{(t)} \mid \hat{\eta}_{(t)}) \tag{37}$$

then

$$l^*_{\rm obs}(\hat{\eta}_{(t+1)}) \geq l^*_{\rm obs}(\hat{\eta}_{(t)}), \tag{38}$$

where $l^*_{\rm obs}(\eta) = \sum_{i=1}^{n} \ln\{f^*_{obs(i)}(\mathbf{y}_{i,obs}, \boldsymbol{\delta}_i; \eta)\}$ *and*

$$f^*_{obs(i)}(\mathbf{y}_{i,obs}, \boldsymbol{\delta}_i; \eta) = \frac{1}{m} \sum_{j=1}^{m} f(\mathbf{y}^*_{ij}, \boldsymbol{\delta}_i; \eta) / h_m(\mathbf{y}^{*(j)}_{i,\rm mis}).$$

Proof

By using Jensen's inequality,

$$\begin{aligned} l^*_{\text{obs}}(\hat{\eta}_{(t+1)}) - l^*_{\text{obs}}(\hat{\eta}_{(t)}) &= \sum_{i=1}^{n} \ln \left\{ \sum_{j=1}^{m} w^*_{ij(t)} \frac{f(\mathbf{y}^*_{ij}, \boldsymbol{\delta}_i; \hat{\eta}_{(t+1)})}{f(\mathbf{y}^*_{ij}, \boldsymbol{\delta}_i; \hat{\eta}_{(t)})} \right\} \\ &\geq \sum_{i=1}^{n} \sum_{j=1}^{m} w^*_{ij(t)} \ln \left\{ \frac{f(\mathbf{y}^*_{ij}, \boldsymbol{\delta}_i; \hat{\eta}_{(t+1)})}{f(\mathbf{y}^*_{ij}, \boldsymbol{\delta}_i; \hat{\eta}_{(t)})} \right\} \\ &= Q^*(\hat{\eta}_{(t+1)} \mid \hat{\eta}_{(t)}) - Q^*(\hat{\eta}_{(t)} \mid \hat{\eta}_{(t)}). \end{aligned}$$

Therefore, (37) implies (38).

Example 4.11: Return to Example 3.15

- Fractional imputation

1. Imputation Step: Generate $y_i^{*(1)}, \cdots, y_i^{*(m)}$ from $f\left(y_i \mid x_i; \hat{\theta}_{(0)}\right)$.
2. Weighting Step: Using the m imputed values generated from Step 1, compute the fractional weights by

$$w_{ij(t)}^* \propto \frac{f\left(y_i^{*(j)} \mid x_i; \hat{\theta}_{(t)}\right)}{f\left(y_i^{*(j)} \mid x_i; \hat{\theta}_{(0)}\right)} \left\{1 - \pi(x_i, y_i^{*(j)}; \hat{\phi}_{(t)})\right\}$$

where

$$\pi(x_i, y_i; \hat{\phi}) = \frac{\exp\left(\hat{\phi}_0 + \hat{\phi}_1 x_i + \hat{\phi}_2 y_i\right)}{1 + \exp\left(\hat{\phi}_0 + \hat{\phi}_1 x_i + \hat{\phi}_2 y_i\right)}.$$

Example 4.11

- Using the imputed data and the fractional weights, the M-step can be implemented by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij(t)}^{*} S\left(\theta; x_i, y_i^{*(j)}\right) = 0$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij(t)}^{*}\left\{\delta_i - \pi(\phi; x_i, y_i^{*(j)})\right\}\left(1, x_i, y_i^{*(j)}\right) = 0, \qquad (39)$$

where $S(\theta; x_i, y_i) = \partial \log f(y_i \mid x_i; \theta)/\partial\theta$.

Example 4.12: Back to Example 3.18 (GLMM)

- Level 1 model
  $$y_{ij} \sim f_1(y_{ij} \mid x_{ij}, a_i; \theta_1)$$
  for some fixed θ_1 and a_i random.
- Level 2 model
  $$a_i \sim f_2(a_i; \theta_2)$$
- Latent variable: a_i
- We are interested in generating a_i from

$$p(a_i \mid \mathbf{x}_i, \mathbf{y}_i; \theta_1, \theta) \propto \left\{ \prod_{j=1}^{n_i} f_1(y_{ij} \mid x_{ij}, a_i; \theta_1) \right\} f_2(a_i; \theta_2)$$

Example 4.12 (Cont'd)

- E-step
  1. Imputation Step: Generate $a_i^{*(1)}, \cdots, a_i^{*(m)}$ from $f_2(a_i; \hat{\theta}_2^{(t)})$.
  2. Weighting Step: Using the m imputed values generated from Step 1, compute the fractional weights by

$$w_{ij(t)}^{*} \propto g_1(\mathbf{y}_i \mid \mathbf{x}_i, a_i^{*(j)}; \hat{\theta}_1^{(t)})$$

  where $g_1(\mathbf{y}_i \mid \mathbf{x}_i, a_i; \hat{\theta}_1) = \prod_{j=1}^{n_i} f_1(y_{ij} \mid x_{ij}, a_i; \theta_1)$.

- M-step: Update the parameters by solving

$$\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij(t)}^{*} S_1\left(\theta_1; \mathbf{x}_i, \mathbf{y}_i, a_i^{*(j)}\right) = 0$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij(t)}^{*} S_2\left(\theta_2; a_i^{*(j)}\right) = 0.$$

Example 4.13: Measurement error model

- Interested in estimating θ in $f(y \mid x; \theta)$.
- Instead of observing x, we observe z which can be highly correlated with x.
- Thus, z is an instrumental variable for x:

$$f(y \mid x, z) = f(y \mid x)$$

  and

$$f(y \mid z = a) \neq f(y \mid z = b)$$

  for $a \neq b$.
- In addition to original sample, we have a separate calibration sample that observes (x_i, z_i).

Example 4.13 (Cont'd)

Table: Data Structure

	Z	X	Y
Calibration Sample	o	o	
Original Sample	o		o

- The goal is to generate x in the original sample from

$$\begin{aligned} f\left(x_i \mid z_i, y_i\right) & \propto f\left(x_i \mid z_i\right) f\left(y_i \mid x_i, z_i\right) \\ & = f\left(x_i \mid z_i\right) f\left(y_i \mid x_i\right) \end{aligned}$$

- Obtain a consistent estimator $\hat{f}(x \mid z)$ from calibration sample.
- E-step
  1. Generate $x_i^{*(1)}, \cdots, x_i^{*(m)}$ from $\hat{f}(x_i \mid z_i)$.
  2. Compute the fractional weights associated with $x_i^{*(j)}$ by

$$w_{ij}^* \propto f(y_i \mid x_i^{*(j)}; \hat{\theta})$$

- M-step: Solve the weighted score equation for θ.

Remarks for Computation

- Recall that, writing

$$\bar{S}^*(\eta \mid \eta) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_{ij}^*(\eta) S(\eta; \mathbf{y}_{ij}^*, \boldsymbol{\delta}_i)$$

where $w_{ij}^*(\eta)$ is the fractional weight associated with y_{ij}^*, denoted by

$$w_{ij}^*(\eta) = \frac{f(\mathbf{y}_{ij}^*, \boldsymbol{\delta}_i; \eta)/h_m(\mathbf{y}_{i,\mathrm{mis}}^{*(j)})}{\sum_{k=1}^{m} f(\mathbf{y}_{ik}^*, \boldsymbol{\delta}_i; \eta)/h_m(\mathbf{y}_{i,\mathrm{mis}}^{*(k)})}, \qquad (40)$$

and $S(\eta; \mathbf{y}, \boldsymbol{\delta}) = \partial \log f(\mathbf{y}, \boldsymbol{\delta}; \eta)/\partial \eta$, the EM algorithm for fractional imputation can be expressed as

$$\hat{\eta}^{(t+1)} \leftarrow \text{ solve } \bar{S}^*(\eta \mid \eta^{(t)}) = 0.$$

Remarks for Computation

- Instead of EM algorithm, Newton-type algorithm can also be used. The Newton-type algorithm for computing the MLE from the fractionally imputed data is given by

$$\hat{\eta}^{(t+1)} = \hat{\eta}^{(t)} + \left\{ I_{obs}^{*}(\hat{\eta}^{(t)}) \right\}^{-1} \bar{S}^{*}(\hat{\eta}^{(t)} \mid \hat{\eta}^{(t)})$$

where

$$\begin{aligned} I_{obs}^{*}(\eta) &= -\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}^{*}(\eta) \dot{S}(\eta; \mathbf{y}_{ij}^{*}, \boldsymbol{\delta}_i) \\ &\quad - \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}^{*}(\eta) \left\{ S(\eta; \mathbf{y}_{ij}^{*}, \boldsymbol{\delta}_i) - \bar{S}_i^{*}(\eta) \right\}^{\otimes 2}, \end{aligned}$$

$\dot{S}(\eta; \mathbf{y}, \boldsymbol{\delta}) = \partial S(\eta; \mathbf{y}, \boldsymbol{\delta}) / \partial \eta$ and $\bar{S}_i^{*}(\eta) = \sum_{j=1}^{M} w_{ij}^{*}(\eta) \dot{S}(\eta; \mathbf{y}_{ij}^{*}, \boldsymbol{\delta}_i)$.

Estimation of general parameter

- Parameter Ψ is defined through $E\{U(\Psi; Y)\} = 0$.
- The FI estimator of Ψ is computed by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}(\hat{\eta})U(\Psi; \mathbf{y}_{ij}^{*}) = 0. \tag{41}$$

Note that $\hat{\eta}$ is the solution to

$$\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}\left(\hat{\eta}\right) S\left(\hat{\eta}; \mathbf{y}_{ij}^{*}\right) = 0.$$

Estimation of general parameter

- We can use either linearization method or replication method for variance estimation. For linearization method, using Theorem 4.2, we can use sandwich formula

$$\hat{V}\left(\hat{\Psi}\right) = \hat{\tau}_q^{-1}\hat{\Omega}_q\hat{\tau}_q^{-1\prime} \tag{42}$$

  where

$$\begin{aligned} \hat{\tau}_q &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}\dot{U}\left(\hat{\Psi};\mathbf{y}_{ij}^{*}\right) \\ \hat{\Omega}_q &= n^{-1}(n-1)^{-1}\sum_{i=1}^{n}\left(\hat{q}_i^{*}-\bar{q}_n^{*}\right)^{\otimes 2}, \end{aligned}$$

  with $\hat{q}_i^{*} = \bar{U}_i^{*} + \hat{\kappa}\bar{S}_i^{*}$, where $(\bar{U}_i^{*}, \bar{S}_i^{*}) = \sum_{j=1}^{m} w_{ij}^{*}(U_{ij}^{*}, S_{ij}^{*})$, $U_{ij}^{*} = U(\hat{\Psi};\mathbf{y}_{ij}^{*})$, $S_{ij}^{*} = S(\hat{\eta};\mathbf{y}_{ij}^{*})$, and

$$\hat{\kappa} = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^{*}(\hat{\eta})\left(U_{ij}^{*}-\bar{U}_i^{*}\right)S_{ij}^{*}\left\{I_{\text{obs}}^{*}(\hat{\eta})\right\}^{-1}$$

Estimation of general parameter

- For replication method, we first obtain the k-th replicate $\hat{\eta}^{(k)}$ of $\hat{\eta}$ by solving

  $$\sum_{i=1}^{n}\sum_{j=1}^{m} w_i^{(k)} w_{ij}^{*}\left(\eta\right) S\left(\eta; \mathbf{y}_{ij}^{*}\right) = 0.$$

  Once $\hat{\eta}^{(k)}$ is obtained then the k-th replicate $\hat{\Psi}^{(k)}$ of $\hat{\Psi}$ is obtained by solving

  $$\sum_{i=1}^{n}\sum_{j=1}^{m} w_i^{(k)} w_{ij}^{*}(\hat{\eta}^{(k)}) U(\Psi; \mathbf{y}_{ij}^{*}) = 0$$

  for ψ.
- The replication variance estimator of $\hat{\Psi}$ from (41) is obtained by

  $$\hat{V}_{rep}(\hat{\Psi}) = \sum_{k=1}^{L} c_k \left(\hat{\Psi}^{(k)} - \hat{\Psi}\right)^2.$$

Example 4.15: Nonparametric Fractional Imputation

- Bivariate data: (x_i, y_i)
- x_i are completely observed but y_i is subject to missingness.
- Joint distribution of (x, y) completely unspecified.
- Assume MAR in the sense that $P(\delta = 1 \mid x, y)$ does not depend on y.
- Without loss of generality, assume that $\delta_i = 1$ for $i = 1, \cdots, r$ and $\delta_i = 0$ for $i = r + 1, \cdots, n$.
- We are only interested in estimating $\theta = E(Y)$.

Example 4.15 (Cont'd)

- Let $K_h(x_i, x_j) = K((x_i - x_j)/h)$ be the Kernel function with bandwidth h such that $K(x) \geq 0$ and

$$\int K(x)dx = 1, \quad \int xK(x)dx = 0, \quad \sigma_K^2 \equiv \int x^2 K(x)dx > 0.$$

  Examples include the following:

  - Boxcar kernel: $K(x) = \frac{1}{2} I(x)$
  - Gaussian kernel: $K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$
  - Epanechnikov kernel: $K(x) = \frac{3}{4}(1 - x^2)I(x)$
  - Tricube Kernel: $K(x) = \frac{70}{81}(1 - |x|^3)^3 I(x)$

  where

$$I(x) = \begin{cases} 1 & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1. \end{cases}$$

Example 4.15 (Cont'd)

- Nonparametric regression estimator of $m(x) = E(Y \mid x)$:

$$\hat{m}(x) = \sum_{i=1}^{r} l_i(x) y_i \tag{43}$$

where

$$l_i(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_j K\left(\frac{x-x_j}{h}\right)}.$$

Estimator in (43) is often called Nadaraya-Watson kernel estimator.

- Under some regularity conditions and under the optimal choice of h (with $h^* = O(n^{-1/5})$), it can be shown that

$$E\left[\{\hat{m}(x) - m(x)\}^2\right] = O(n^{-4/5}).$$

Thus, its convergence rate is slower than that of parametric one.

Example 4.15 (Cont'd)

- However, the imputed estimator of θ using (43) can achieve the $\sqrt{n}$-consistency. That is,

$$\hat{\theta}_{NP} = \frac{1}{n} \left\{ \sum_{i=1}^{r} y_i + \sum_{i=r+1}^{n} \hat{m}(x_i) \right\} \tag{44}$$

  achieves

$$\sqrt{n} \left(\hat{\theta}_{NP} - \theta \right) \longrightarrow N(0, \sigma^2) \tag{45}$$

  where $\sigma^2 = E\{v(x)/\pi(x)\} + V\{m(x)\}$, $m(x) = E(y \mid x)$, $v(x) = V(y \mid x)$ and $\pi(x) = E(\delta \mid x)$.

- Result (45) was first proved by Cheng (1994).

Example 4.15 (Cont'd)

- We can express $\hat{\theta}_{NP}$ in (45) as a nonparametric fractional imputation (NFI) estimator of the form

$$\hat{\theta}_{NFI} = \frac{1}{n}\left\{\sum_{i=1}^{r} y_i + \sum_{j=r+1}^{n}\sum_{i=1}^{r} w_{ij}^{*} y_i^{*(j)}\right\}$$

where $w_{ij}^{*} = l_i(x_j)$, which is defined after (43), and $y_i^{*(j)} = y_i$.

- Variance estimation can be implemented by a resampling method, such as bootstrap.