Section 5.5: Multiple Imputation for General Purpose Estimation

Jae-Kwang Kim

Iowa State University

Basic Setup

- $y_i, i = 1, \cdots, n$: random sample from a joint density $f(y;\theta)$.
- The first r elements are observed and the rest $n-r$ elements are missing.
- Assume MAR (Missing At Random)
- The parameter of interest is $\psi = E\{g(Y)\} = \int g(y)f(y:\theta)dy$. Thus, ψ is a function of θ. For example, $g(Y) = I(Y<1)$ leads to $\psi = P(Y<1)$.
- MI estimator is justified when $\hat{\psi}_n$ (complete sample estimator of ψ) is the MLE of ψ. That is, $\hat{\psi}_n = \psi(\hat{\theta}_{MLE,n})$.

- What if $\hat{\psi}_n$ is a MME (Method of Moments Estimator) of ψ ? The MME of ψ is

$$\hat{\psi}_n = \frac{1}{n}\sum_{i=1}^{n} g(y_i).$$

- MI estimator of ψ:

$$\hat{\psi}_{MI} = \frac{1}{m}\sum_{k=1}^{m} \hat{\psi}_I^{(k)}$$

where

$$\hat{\psi}_I^{(k)} = \frac{1}{n}\left\{\sum_{i=1}^{r} g(y_i) + \sum_{i=r+1}^{n} g(y_i^{*(k)})\right\}$$

where $y_i^{*(k)}$ are generated from $f(y \mid \theta^{*(k)})$ and $\theta^{*(k)}$ are generated from $p(\theta \mid y_{obs})$.

- MI variance estimator:

$$\hat{V}_{MI} = W_m + \left(1 + \frac{1}{m}\right) B_m$$

where

$$\begin{aligned} W_m &= \frac{1}{m}\sum_{k=1}^{m} \hat{V}_{n,I}^{(k)} \\ B_m &= \frac{1}{m-1}\sum_{k=1}^{m}\left(\hat{\psi}_I^{(k)} - \bar{\psi}_{MI}\right)^2 . \end{aligned}$$

Here,

$$\hat{V}_{n,I}^{(k)} = \frac{1}{n(n-1)}\left[\sum_{i=1}^{r}\{g(y_i) - \hat{\psi}_I^{(k)}\}^2 + \sum_{i=r+1}^{n}\{g(y_i^{*(k)}) - \hat{\psi}_I^{(k)}\}^2\right]$$

is the variance estimator of $\hat{\psi}_n$ applied to the k-th imputed data.

Is the MI variance estimator approximately unbiased for $V(\hat{\psi}_{MI})$?

- The answer is "Yes" if the variance satisfies

$$V(\hat{\psi}_{MI}) = V(\hat{\psi}_n) + V(\hat{\psi}_{MI} - \hat{\psi}_n). \tag{1}$$

It can be shown that

$$E\{W_m\} = V(\hat{\psi}_n) \tag{2}$$

and

$$E\left\{\left(1 + m^{-1}\right) B_m\right\} = V(\hat{\psi}_{MI} - \hat{\psi}_n) \tag{3}$$

Kim et al (2006) proved (2) and (3) for general cases.

- Meng (1994) called condition (1) "self-efficiency" .

Example 5.10

- Consider the bivariate data (x_i, y_i) of size $n = 200$ where x_i is always observed and y_i is subject to missingness. The sampling distribution of (x_i, y_i) is $x_i \sim N(3, 1)$ and $y_i = -2 + x_i + e_i$ with $e_i \sim N(0, 1)$.
- Multiple imputation can be used to estimate $\theta_1 = E(Y)$ and $\theta_2 = Pr(Y < 1)$. The response mechanism is uniform with response rate 0.6.
- In estimating θ_2, we used a method-of-moment estimator $\hat{\theta}_2 = n^{-1} \sum_{i=1}^{n} I(y_i < 1)$ under complete response. An unbiased estimator for the variance of $\hat{\theta}_2$ is then $\hat{V}_2 = (n-1)^{-1} \hat{\theta}_2 \left(1 - \hat{\theta}_2\right)$. Multiple imputation with size $m = 50$ was used. After multiple imputation, Rubin's variance formula was used.

Simulation Results (based on $B = 1,000$ Monte Carlo samples)

Table: Simulation results of the MI point estimators

Parameter	Mean	$V(\hat{\theta}_n)$	$V(\hat{\theta}_{MI})$	$V(\hat{\theta}_{MI} - \hat{\theta}_n)$	$Cov(\hat{\theta}_n, \hat{\theta}_{MI} - \hat{\theta}_n)$
θ_1	1.00	0.0100	0.0134	0.0035	0.0000
θ_2	0.50	0.00129	0.00137	0.00046	-0.00019

Table: Simulation results of the MI variance estimators

Parameter	$E(W_m)$	$E(B_m)$	Rel. Bias (%)	t-statistics
$V(\hat{\theta}_1)$	0.0100	0.0033	-0.24	-0.08
$V(\hat{\theta}_2)$	0.00125	0.000436	23.08	7.48

Discussion

- For $\theta = \theta_1$,

$$\begin{aligned} V(\hat{\theta}_{MI}) &= V(\hat{\theta}_n) + V(\hat{\theta}_{MI} - \hat{\theta}_n) \\ &= \frac{\sigma_y^2}{n} + \left(\frac{1}{r} - \frac{1}{n}\right)\sigma_e^2 \\ &= \frac{2}{200} + \left(\frac{1}{120} - \frac{1}{200}\right) \cdot 1 = 0.010 + 0.0033. \end{aligned}$$

- In the simulation result, $E(W_m) = 0.0100$ and $E(B_m) = 0.0033$. Thus, W_m estimates $V(\hat{\theta}_n)$ and B_m estimates $V(\hat{\theta}_{MI} - \hat{\theta}_n)$ nearly unbiasedly.
- The test statistic for testing $E(\hat{V}_{MI}) = V(\hat{\theta}_{MI})$ is very small ($T = -0.08$), which confirms unbiasedness.

Discussion

- For $\theta = \theta_2$, we have

$$\begin{aligned} V(\hat{\theta}_{MI}) &= V(\hat{\theta}_n) + V(\hat{\theta}_{MI} - \hat{\theta}_n) + 2Cov(\hat{\theta}_n, \hat{\theta}_{MI} - \hat{\theta}_n) \\ &\doteq 0.00129 + 0.00046 + 2 \cdot (-0.00019) = 0.00137, \end{aligned}$$

  while

$$\begin{aligned} E(\hat{V}_{MI}) &= E(W_m) + (1 + m^{-1})E(B_m) \\ &= 0.00125 + 1.02 \cdot 0.000436 = 0.00169 > 0.00137. \end{aligned}$$

- Thus, the MI variance estimator overestimates the variance because it ignores the covariance term between $\hat{\theta}_n$ and $\hat{\theta}_{MI} - \hat{\theta}_n$.

Theory for MI under MME

- Bayesian imputation: For each $j = 1, \cdots, m$,
  1. Generate $\eta^{*(j)}$ be the j-th sample from $p(\eta \mid \mathbf{y}_{obs}, \boldsymbol{\delta}) \doteq N(\hat{\eta}, \hat{I}_{obs}^{-1})$.
  2. Generate $y_{i,\mathrm{mis}}^{*(j)}$ from $f(y_{i,\mathrm{mis}} \mid y_{i,\mathrm{obs}}, \delta_i; \eta^{*(j)})$.
- The j-th imputed estimator of $\psi = E\{g(Y)\}$:

$$
\begin{aligned}
\hat{\psi}_I^{*(j)} &= \frac{1}{n}\sum_{i=1}^{n} g(y_{i,\mathrm{obs}}, y_{i,\mathrm{mis}}^{*(j)}) \\
&= \frac{1}{n}\left[\sum_{i=1}^{n} m_g(y_{i,\mathrm{obs}}; \eta^{*(j)}) + \sum_{i=1}^{n} e_i^{*(j)}\right]
\end{aligned}
$$

  where $m_g(y_{i,\mathrm{obs}}; \eta) = E\{g(Y_i) \mid y_{i,\mathrm{obs}}; \eta\}$ and

$$
e_i^{*(j)} = g(y_{i,\mathrm{obs}}, y_{i,\mathrm{mis}}^{*(j)}) - m_g(y_{i,\mathrm{obs}}; \eta^{*(j)}).
$$

- If y_i is completely observed, then $m_g(y_{i,\mathrm{obs}}; \eta) = g(y_i)$ and $e_i^{*(j)} = 0$.

Theory for MI under MME

- Using Taylor linearization, we obtain

$$m_g(y_{i,\mathrm{obs}};\eta^{*(j)}) \cong m_g(y_{i,\mathrm{obs}};\hat{\eta}) + \dot{m}_g(y_{i,\mathrm{obs}};\hat{\eta})(\eta^{*(j)} - \hat{\eta})$$

where

$$\dot{m}_g(y_{i,\mathrm{obs}};\eta) = \frac{\partial}{\partial \eta'} m_g(y_{i,\mathrm{obs}};\eta).$$

- Thus, we can obtain

$$\hat{\psi}_I^{*(j)} \cong \hat{\psi}_p + \left\{ \frac{1}{n}\sum_{i=1}^{n} \dot{m}_g(y_{i,\mathrm{obs}};\hat{\eta}) \right\} (\eta^{*(j)} - \hat{\eta}) + \frac{1}{n}\sum_{i=1}^{n} e_i^{*(j)}$$

where

$$\hat{\psi}_p = \frac{1}{n}\sum_{i=1}^{n} m_g(y_{i,\mathrm{obs}};\hat{\eta}).$$

Theory for MI under MME

- MI estimator of ψ:

$$\hat{\psi}_{MI} \cong \hat{\psi}_p + \left\{ \frac{1}{n} \sum_{i=1}^{n} \dot{m}_g(y_{i,\text{obs}}; \hat{\eta}) \right\} \frac{1}{m} \sum_{j=1}^{m} (\eta^{*(j)} - \hat{\eta}) + \frac{1}{nm} \sum_{j=1}^{m} \sum_{i=1}^{n} e_i^{*(j)}$$

- Variance of $\hat{\psi}_{MI}$:

$$V(\hat{\psi}_{MI}) = V(\hat{\psi}_p) + m^{-1} \left\{ \mathcal{D}_g \mathcal{I}_{obs}^{-1} \mathcal{D}_g' + n^{-1} E[V\{g(Y) \mid \mathbf{y}_{\text{obs}}\}] \right\},$$

where

$$\mathcal{D}_g = p \lim_n \frac{1}{n} \sum_{i=1}^{n} \dot{m}_g(y_{i,\text{obs}}; \hat{\eta}).$$

Remark 1

Let $\hat{\psi}_{MME,n} = n^{-1} \sum_{i=1}^{n} g(y_i)$ be the MME of ψ using the full sample. Note that

$$
\begin{aligned}
\hat{\psi}_{MME,n} - \hat{\psi}_p &= \frac{1}{n} \sum_{i=1}^{n} \{g(y_i) - m_g(y_{i,\mathrm{obs}}; \hat{\eta})\} \\
&\cong \frac{1}{n} \sum_{i=1}^{n} \{g(y_i) - m_g(y_{i,\mathrm{obs}}; \eta_0)\} + \mathcal{D}_g \left(\hat{\eta} - \eta_0 \right).
\end{aligned}
$$

Thus,

$$
V\left(\hat{\psi}_{MME,n} - \hat{\psi}_p \right) \cong \mathcal{D}_g \mathcal{I}_{obs}^{-1} \mathcal{D}_g' + n^{-1} E[V\{g(Y) \mid Y_{\mathrm{obs}}\}]
$$

Variance estimation

- Now, let's investigate the bias of MI variance estimator. We first assume that W_m is unbiased for $V(\hat{\psi}_{MME,n})$.
- Using Lemma 5.2, we can obtain

$$E(B_m) = \mathcal{D}_g \mathcal{I}_{obs}^{-1} \mathcal{D}_g' + n^{-1} E[V\{g(Y) \mid Y_{\rm obs}\}]$$

- By Remark 1, we have

$$E(B_m) = V(\hat{\psi}_p - \hat{\psi}_{MME,n}).$$

- Therefore, the bias of Rubin's variance estimator is

$$\begin{aligned} E(\hat{V}_{MI}) - V(\hat{\psi}_{MI}) &= V(\hat{\psi}_{MME,n}) + V(\hat{\psi}_p - \hat{\psi}_{MME,n}) - V(\hat{\psi}_p) \\ &= 2Cov(\hat{\psi}_{MME,n}, \hat{\psi}_{MME,n} - \hat{\psi}_p). \qquad (4) \end{aligned}$$

Example

- Let (x_i, y_i) be the vector of random variables in the sample, where x_i is always observed and y_i is subject to missingness.
- Let $f(y \mid x; \theta)$ be the prediction model for y with unknown parameter θ.
- Let $\psi = E\{g(Y)\}$ and the MME of ψ, $\hat{\psi}_n = n^{-1} \sum_{i=1}^{n} g(y_i)$ for estimation of ψ.
- Bayesian imputation is used for multiple imputation:
  1. Generate $\theta^* \sim N(\hat{\theta}, \hat{I}_{obs}^{-1})$.
  2. For $\delta_i = 0$, generate $y_i^* \sim f(y_i \mid x_i; \theta^*)$.
- Let's investigate the bias of MI variance estimator using (4).

Example

- Note that

$$\hat{\psi}_p = \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_i g(y_i) + (1-\delta_i)m(x_i;\hat{\theta})\right\}.$$

- Thus,

$$\hat{\psi}_p \cong \frac{1}{n}\sum_{i=1}^{n}[\delta_i g(y_i) + (1-\delta_i)E\{g(Y_i) \mid x_i\}] + \mathcal{D}_g(\hat{\theta}-\theta_0)$$

where

$$\begin{aligned}\mathcal{D}_g &= p\lim_n \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\partial}{\partial\theta}E\{g(Y) \mid x_i;\theta_0\} \\ &= (1-p)E\{\dot{m}_g(X) \mid \delta = 0\}.\end{aligned}$$

Example

- Thus,

$$\hat{\psi}_{MME,n} - \hat{\psi}_p \cong \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)e_i - r^{-1}\mathcal{D}_g\mathcal{I}_\theta^{-1}\sum_{i=1}^{n}\delta_i S(\theta; x_i, y_i)$$

- We can establish that

$$\begin{aligned} & Cov(\hat{\psi}_{MME,n}, \hat{\psi}_{MME,n} - \hat{\psi}_p) \\ = \; & n^{-1}(1-p)E[V\{g(Y) \mid X\} \mid \delta = 0] \\ & -n^{-1}(1-p)E\{\dot{m}_g(X) \mid \delta = 0\}'\mathcal{I}_\theta^{-1}E\{\dot{m}_g(X) \mid \delta = 1\}. \end{aligned}$$

- In many cases, the covariance term is positive. See Yang and Kim (2016) for more details.