5.2 - 5.3 Multiple Imputation

Jae-Kwang Kim

Basic Setup

- Parameter of interest: ψ defined through $E\{U(\psi; Y)\} = 0$
- $\mathbf{y}_{\rm com} = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$: Complete sample data (without nonresponse)
- From $\mathbf{y}_{\rm com}$, we can obtain $(\hat{\psi}_n, \hat{V}_n)$ where
  - $\hat{\psi}_n = \hat{\psi}(\mathbf{y}_{\rm com})$: Point estimator of ψ, solution to $\sum_{i=1}^n U(\psi; \mathbf{y}_i) = 0$.
  - $\hat{V}_n = \hat{V}(\mathbf{y}_{\rm com})$: Variance estimator of $\hat{\psi}_n$ (obtained by Sandwich formula).
- Under some regularity conditions,

$$\frac{\hat{\psi}_n - \psi}{\sqrt{\hat{V}_n}} \xrightarrow{\mathcal{L}} N(0,1)$$

as $n \to \infty$.

Multiple Imputation method - Rubin's idea

1. Generate m realization of $\mathbf{y}_{mis}^{*(j)}$, $j = 1, 2, \cdots, m$, from the posterior predictive distribution

$$P\left(\mathbf{y}_{mis} \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}\right) = \int P\left(\mathbf{y}_{mis} \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}; \eta\right) P\left(\eta \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}\right) d\eta. \tag{1}$$

2. Obtain $(\hat{\psi}_n^{*(j)}, \hat{V}_n^{*(j)})$ using $\mathbf{y}_{\rm com}^{*(j)} = (\mathbf{y}_{\rm obs}, \mathbf{y}_{\rm mis}^{*(j)})$
3. Combine m estimates:
    - Point estimator of ψ: $\hat{\psi}_{MI} = m^{-1} \sum_{j=1}^{m} \hat{\psi}_n^{*(j)}$
    - Variance estimator of $\hat{\psi}_{MI}$:

$$\hat{V}_{MI} = W_m + \left(1 + m^{-1}\right) B_m \tag{2}$$

where $W_m = m^{-1} \sum_{j=1}^{m} \hat{V}_n^{*(j)}$ and $B_m = (m-1)^{-1} \sum_{j=1}^{m} \left(\hat{\psi}_n^{*(j)} - \hat{\psi}_{MI}\right)^2$.

Two-step approach for Bayesian Imputation

Goal: Generate imputed values from the posterior predictive distribution in (1):

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

1. Generate $\eta^{*(1)}, \cdots, \eta^{*(m)}$ from $P(\eta \mid \mathbf{y}_{\mathrm{obs}}, \boldsymbol{\delta})$.
2. Given the j-th parameter value $\eta^{*(j)} = (\theta^{*(j)}, \phi^{*(j)})$, generate $\mathbf{y}_{\mathrm{mis}}^{*(j)}$ from the conditional distribution $P(\mathbf{y}_{\mathrm{mis}} \mid \mathbf{y}_{\mathrm{obs}}, \boldsymbol{\delta}; \eta^{*(j)})$, where

$$P(\mathbf{y}_{mis} \mid \mathbf{y}_{\mathrm{obs}}, \boldsymbol{\delta}; \eta^{*(j)}) = \frac{f(\mathbf{y}; \theta^{*(j)}) P(\boldsymbol{\delta} \mid \mathbf{y}; \phi^{*(j)})}{\int f(\mathbf{y}; \theta^{*(j)}) P(\boldsymbol{\delta} \mid \mathbf{y}; \phi^{*(j)}) d\mathbf{y}_{mis}}.$$

Bayesian Justification (Rubin, 1987) for $\psi = \eta$

1. If $\hat{\psi}_n = E\{\psi \mid \mathbf{y}_{\rm com}\}$, then, for sufficiently large m, we have

$$
\begin{aligned}
\hat{\psi}_{MI} &= m^{-1} \sum_{k=1}^{m} \hat{\psi}_n^{*(k)} \\
&= m^{-1} \sum_{k=1}^{m} E(\psi \mid \mathbf{y}_{\rm obs}, \mathbf{y}_{\rm mis}^{*(k)}) \\
&\doteq E\{E(\psi \mid \mathbf{y}_{\rm obs}, \mathbf{Y}_{\rm mis}) \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}\} \\
&= E(\psi \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}).
\end{aligned}
$$

Bayesian Justification (Cont'd)

2. Also, if $\hat{V}_n = V\{\psi \mid \mathbf{y}_{\rm com}\}$, then, for sufficiently large m, we have

$$
\begin{aligned}
\hat{V}_{MI} &= W_m + B_m \\
&= \frac{1}{m}\sum_{k=1}^{m} \hat{V}_n^{*(k)} + \frac{1}{m-1}\sum_{k=1}^{m}\left(\hat{\psi}_n^{*(k)} - \bar{\psi}_{MI}\right)^2 \\
&\doteq E\{V(\psi \mid \mathbf{y}_{\rm obs}, \mathbf{Y}_{\rm mis}) \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}\} \\
&\quad + V\{E(\psi \mid \mathbf{y}_{\rm obs}, \mathbf{Y}_{\rm mis}) \mid \mathbf{y}_{obs}, \boldsymbol{\delta}\} \\
&= V\{\psi \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}\}.
\end{aligned}
$$

Theory for multiple imputation

- We are interested in showing

$$\frac{\hat{\psi}_{MI} - \psi}{\sqrt{\hat{V}_{MI}}} \xrightarrow{\mathcal{L}} N(0,1)$$

  as $n \to \infty$.
- Two different cases
  - Case 1: $\psi = \eta$ and $U(\psi) = S(\eta)$
  - Case 2: $\psi \neq \eta$
- An example for Case 2: $Y \sim N(\mu, \sigma^2)$ and the parameter of interest is $\psi = P(Y < c)$.
  - Maximum likelihood estimation: $\hat{\psi} = P(Y < c; \hat{\mu}, \hat{\sigma}^2)$.
  - Method-of-moments estimation: solve $\sum_{i=1}^{n} I(y_i < c) = 0$.

Asymptotic Properties of MI estimator: Case 1

- Interested in the asymptotic properties for $\hat{\eta}_{MI} = m^{-1} \sum_{j=1}^{m} \hat{\eta}_n^{*(j)}$ where $\hat{\eta}_n^{*(j)}$ is computed by solving the imputed score equation for η.
- (Theorem 5.1) We will show that, under some regularity conditions, the MI estimator $\hat{\eta}_{\mathrm{MI}}$ is approximately unbiased for η_0 and has the asymptotic variance

$$V(\hat{\eta}_{MI}) \cong \mathcal{I}_{\mathrm{obs}}^{-1} + m^{-1} \mathcal{J}_{\mathrm{mis}} \mathcal{I}_{\mathrm{obs}}^{-1} \mathcal{J}_{\mathrm{mis}}' + m^{-1} \mathcal{I}_{\mathrm{com}}^{-1} \mathcal{I}_{\mathrm{mis}} \mathcal{I}_{\mathrm{com}}^{-1}. \tag{3}$$

  Thus, for $m \to \infty$, $\hat{\eta}_{MI}$ has the same performance as the MLE of η.
- (Theorem 5.2) Validity of MI variance estimator will also be established.

Setup

- Let $S_{\text{com}}(\eta; \mathbf{y}_{\text{com}})$ be the score function of η under complete response.
- Frequentist Imputation: Let $\hat{\eta}^*$ be the solution to

$$S_{\text{com}}(\eta; \mathbf{y}_{\text{obs}}, \mathbf{y}^*_{\text{mis}}) = 0$$

  where $\mathbf{y}^{*(j)}_{\text{mis}} \sim f(\mathbf{y}_{mis} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \hat{\eta}_p)$ and $\hat{\eta}_p$ is a consistent estimator of η.
- In Bayesian Imputation, instead of using a fixed $\hat{\eta}_p$, we use

$$\eta^*_p \mid (\mathbf{y}_{\text{obs}}, \boldsymbol{\delta}) \sim N(\hat{\eta}_{MLE}, \hat{I}^{-1}_{\text{obs}}), \tag{4}$$

  which is a large-sample approximation of $p(\eta \mid \mathbf{y}_{obs}, \boldsymbol{\delta})$. (Bernstein von-Mises theorem)

Recall: Lemma 4.3

Let $S_I^*(\eta \mid \hat{\eta}_p) = S_{\rm com}(\eta; \mathbf{y}^*)$ be the imputed score function evaluated with $\mathbf{y}^* = (\mathbf{y}_{\rm obs}, \mathbf{y}^*_{\rm mis})$ where $\mathbf{y}^*_{\rm mis}$ is generated from $f(\mathbf{y}_{\rm mis} \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}; \hat{\eta}_p)$. Assume that $\hat{\eta}_p$ converges in probability to η_0. Then, under some regularity conditions, the solution $\hat{\eta}^*$ to $S_I^*(\eta \mid \hat{\eta}_p) = 0$ satisfies

$$\hat{\eta}^* - \eta_0 \cong (\hat{\eta}_{MLE} - \eta_0) + \mathcal{J}_{\rm mis}(\hat{\eta}_p - \hat{\eta}_{MLE}) + \mathcal{I}_{\rm com}^{-1} S^*_{\rm mis}(\hat{\eta}_p) \tag{5}$$

where $\mathcal{J}_{\rm mis} = \mathcal{I}_{\rm com}^{-1}\mathcal{I}_{\rm mis}$ and $S^*_{\rm mis}(\eta) = S_{\rm com}(\eta; \mathbf{y}^*) - S_{\rm obs}(\eta)$.

Bayesian Imputation

- In Bayesian imputation, we use $\hat{\eta}_p = \eta_p^*$ where $\eta_p^* \sim p(\eta \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta})$. Thus, by Lemma 4.3, we have

$$\hat{\eta}^* - \eta_0 \cong \hat{\eta}_{MLE} - \eta_0 + \mathcal{J}_{\rm mis}\left(\eta_p^* - \hat{\eta}_{MLE}\right) + \mathcal{I}_{\rm com}^{-1} Z^* \tag{6}$$

where $Z^* \mid (\mathbf{y}_{obs}, \boldsymbol{\delta}, \eta_p^*) \sim (0, \mathcal{I}_{mis})$. The three terms in (6) are independent.
- Also, for Bayesian imputation, by (4),

$$\left(\eta_p^* - \hat{\eta}_{MLE}\right) \mid (\mathbf{y}_{\rm obs}, \boldsymbol{\delta}) \sim (0, \hat{I}_{\rm obs}^{-1})$$

- The total variance of $\hat{\eta}^*$ is then

$$V\left(\hat{\eta}^*\right) \cong \mathcal{I}_{\rm obs}^{-1} + \mathcal{J}_{\rm mis}\mathcal{I}_{\rm obs}^{-1}\mathcal{J}_{\rm mis}' + \mathcal{I}_{\rm com}^{-1}\mathcal{I}_{\rm mis}\mathcal{I}_{\rm com}^{-1}. \tag{7}$$

Bayesian Imputation

- In multiple imputation, we have $\hat{\eta}_{MI} = m^{-1}\sum_{j=1}^{m}\hat{\eta}^{*(j)}$ and the m values of $\hat{\eta}^{*(j)}$ for $j = 1, \cdots, m$ are independently distributed with

$$\hat{\eta}^{*(j)} - \eta_0 \cong \hat{\eta}_{MLE} - \eta_0 + \mathcal{J}_{\rm mis}\left(\eta_p^{*(j)} - \hat{\eta}_{MLE}\right) + \mathcal{I}_{\rm com}^{-1} Z^{*(j)} \qquad (8)$$

with $Z^{*(j)} \mid (\mathbf{y}_{obs}, \boldsymbol{\delta}, \eta_p^{*(j)}) \sim (0, \mathcal{I}_{mis})$ and

$$\left(\eta_p^{*(j)} - \hat{\eta}_{MLE}\right) \mid (\mathbf{y}_{\rm obs}, \boldsymbol{\delta}) \sim (0, \hat{I}_{\rm obs}^{-1})$$

- The three terms in (8) are orthogonal (i.e. independent). Thus, expression (8) can be called the orthogonalization of the imputed estimator $\hat{\eta}^*$.
- Thus, we can establish Theorem 5.1 (next page).

Theorem 5.1

Assume that the posterior distribution $p\left(\eta \mid \mathbf{y}_{\rm obs}, \boldsymbol{\delta}\right)$ is asymptotically normal with mean $\hat{\boldsymbol{\eta}}_{MLE}$ and variance $\{I_{\rm obs}(\hat{\eta}_{MLE})\}^{-1}$, where $I_{\rm obs}(\boldsymbol{\eta})$ is the observed Fisher information derived from the marginal density of $(\mathbf{y}_{\rm obs}, \boldsymbol{\delta})$. Under some regularity conditions, the MI estimator $\hat{\eta}_{MI}$ is approximately unbiased for η_0 and has the asymptotic variance

$$V\left(\hat{\eta}_{MI}\right) \cong \mathcal{I}_{\rm obs}^{-1} + m^{-1}\mathcal{J}_{\rm mis}\mathcal{I}_{\rm obs}^{-1}\mathcal{J}_{\rm mis}' + m^{-1}\mathcal{I}_{\rm com}^{-1}\mathcal{I}_{\rm mis}\mathcal{I}_{\rm com}^{-1}. \tag{9}$$

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

Theorem 5.2

Assume that m preliminary values of η, denoted by $\eta_p^{*(1)}, \cdots, \eta_p^{*(m)}$, are independently generated from a normal distribution with mean $\hat{\eta}_{MLE}$ and variance matrix $\{I_{\text{obs}}(\hat{\eta}_{MLE})\}^{-1}$. Assume that the complete sample variance estimator $\hat{V}$ satisfies

$$E\left\{\hat{V}_I^{*(j)}\right\} \cong \mathcal{I}_{\text{com}}^{-1}, \tag{10}$$

where $\hat{V}_I^{*(j)}$ is the naive variance estimator that is computed by applying $\hat{V}$ to the j-th imputed data $\mathbf{y}^{*(j)}$. Then, the Rubin's variance estimator (2) is asymptotically unbiased for the variance of the MI estimator $\hat{\eta}_{MI}$.

Lemma 5.1

Let $X_1, \cdots, X_m$ be identically distributed (not necessarily independent) with mean θ and covariance

$$Cov(X_i, X_j) = \begin{cases} \sigma_{11} & \text{if } i = j \\ \sigma_{12} & \text{otherwise.} \end{cases}$$

Let $B_m = (m-1)^{-1} \sum_{i=1}^{m} (X_i - \bar{X}_m)^2$ with $\bar{X}_m = m^{-1} \sum_{i=1}^{m} X_i$. Then we have

$$E(B_m) = \sigma_{11} - \sigma_{12}.$$

Proof of Lemma 5.1

Proof of Theorem 5.2

By Lemma 5.1, we have

$$E\left(B_m\right) = V(\hat{\eta}^{*(1)}) - Cov(\hat{\eta}^{*(1)}, \hat{\eta}^{*(2)}).$$

Using (8), we can obtain

$$\begin{aligned} E\left(B_m\right) &= V\left\{\mathcal{J}_{\text{mis}}\left(\eta_p^{*(1)} - \hat{\eta}_{MLE}\right)\right\} + V\left\{\mathcal{I}_{\text{com}}^{-1} Z^{*(1)}\right\} \\ &= \mathcal{J}_{\text{mis}}\mathcal{I}_{\text{obs}}^{-1}\mathcal{J}_{\text{mis}}' + \mathcal{I}_{\text{com}}^{-1}\mathcal{I}_{\text{mis}}\mathcal{I}_{\text{com}}^{-1} \end{aligned}$$

Thus, by assumption (10), we have

$$E\left\{\hat{V}_{MI}(\hat{\eta}_{MI})\right\} \cong \mathcal{I}_{\text{com}}^{-1} + \left(1 + m^{-1}\right)\left(\mathcal{J}_{\text{mis}}\mathcal{I}_{\text{obs}}^{-1}\mathcal{J}_{\text{mis}}' + \mathcal{I}_{\text{com}}^{-1}\mathcal{I}_{\text{mis}}\mathcal{I}_{\text{com}}^{-1}\right). \qquad (11)$$

Using matrix algebra, we have

$$\left(A + BCB'\right)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + B'A^{-1}B\right)^{-1}B'A^{-1}$$

and

$$\left(C^{-1} + B'A^{-1}B\right)^{-1} = C - CB'\left(A + BCB'\right)^{-1}BC,$$

which leads to

$$\left(A + BCB'\right)^{-1} = A^{-1} - A^{-1}BCB'A^{-1} + A^{-1}BCB'\left(A + BCB'\right)^{-1}BCB'A^{-1}.$$

Applying the above equality to $A = \mathcal{I}_{\rm com}$, $B = I$, and $C = -\mathcal{I}_{\rm mis}$, we have

$$\mathcal{I}_{\rm obs}^{-1} = \mathcal{I}_{\rm com}^{-1} + \mathcal{J}_{\rm mis}\mathcal{I}_{\rm obs}^{-1}\mathcal{J}_{\rm mis}' + \mathcal{I}_{\rm com}^{-1}\mathcal{I}_{\rm mis}\mathcal{I}_{\rm com}^{-1} \tag{12}$$

and (11) reduces to

$$E\left\{\hat{V}_{MI}(\hat{\eta}_{MI})\right\} \cong \mathcal{I}_{\rm obs}^{-1} + m^{-1}\mathcal{J}_{\rm mis}\mathcal{I}_{\rm obs}^{-1}\mathcal{J}_{\rm mis}' + m^{-1}\mathcal{I}_{\rm com}^{-1}\mathcal{I}_{\rm mis}\mathcal{I}_{\rm com}^{-1}, \tag{13}$$

which shows the asymptotic unbiasedness of Rubin's variance estimator by (9).

Example 5.5

- Let $y_1, \cdots, y_n$ be IID observations from $N(\mu, \sigma^2)$ and only the first r elements are observed and the remaining $n - r$ elements are missing. Assume that the response mechanism is MAR.
- Bayesian imputation:
  1. The j-th posterior values of (μ, σ^2) are generated from

     $$\sigma^{2*(j)} \mid \mathbf{y}_r \sim \text{Inv. Gamma}\left(\frac{r-1}{2}, \frac{(r-1)S_r^2}{2}\right)$$

     and

     $$\mu^{*(j)} \mid (\mathbf{y}_r, \sigma^{2*(j)}) \sim N\left(\bar{y}_r, r^{-1}\sigma^{2*(j)}\right)$$

     where $\mathbf{y}_r = (y_1, \cdots, y_r)$, $\bar{y}_r = r^{-1}\sum_{i=1}^r y_i$, and $\hat{S}_r^2 = (r-1)^{-1}\sum_{i=1}^r (y_i - \bar{y}_r)^2$.
  2. Given the posterior sample $(\mu^{*(j)}, \sigma^{2*(j)})$, the imputed values are generated from

     $$y_i^{*(j)} \mid \left(\mathbf{y}_r, \mu^{*(j)}, \sigma^{2*(j)}\right) \sim N\left(\mu^{*(j)}, \sigma^{2*(j)}\right)$$

     independently for $i = r+1, \cdots, n$.

- Let $\theta = E(Y)$ be the parameter of interest and the MI estimator of θ can be expressed as

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

where

$$\hat{\theta}_I^{(j)} = \frac{1}{n} \left\{ \sum_{i=1}^{r} y_i + \sum_{i=r+1}^{n} y_i^{*(j)} \right\}.$$

- Then, using the same orthogonalization in (8), we have

$$\hat{\theta}_{MI} = \bar{y}_r + \frac{n-r}{nm}\sum_{j=1}^{m}\left(\mu^{*(j)} - \bar{y}_r\right) + \frac{1}{nm}\sum_{i=r+1}^{n}\sum_{j=1}^{m}\left(y_i^{*(j)} - \mu^{*(j)}\right).$$

- Asymptotically, the first term has mean μ and variance $r^{-1}\sigma^2$, the second term has mean zero and variance $(1-r/n)^2\sigma^2/(mr)$, the third term has mean zero and variance $\sigma^2(n-r)/(n^2m)$, and the three terms are mutually independent.
- Thus, the variance of $\hat{\theta}_{MI}$ is

$$V\left(\hat{\theta}_{MI}\right) = \frac{1}{r}\sigma^2 + \frac{1}{m}\left(\frac{n-r}{n}\right)^2\left(\frac{1}{r}\sigma^2 + \frac{1}{n-r}\sigma^2\right). \qquad (14)$$

- For variance estimation, note that

$$
\begin{aligned}
V(y_i^{*(j)}) &= V(\bar{y}_r) + V(\mu^{*(j)} - \bar{y}_r) + V(y_i^{*(j)} - \mu^{*(j)}) \\
&= \frac{1}{r}\sigma^2 + \frac{1}{r}\sigma^2 \left(\frac{r+1}{r-1}\right) + \sigma^2 \left(\frac{r+1}{r-1}\right) \\
&\cong \sigma^2.
\end{aligned}
$$

- Writing

$$
\begin{aligned}
\hat{V}_I^{(j)}(\hat{\theta}) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left\{ \tilde{y}_i^{*(j)} - \frac{1}{n} \sum_{k=1}^{n} \tilde{y}_k^{*(j)} \right\}^2 \\
&= \frac{1}{n(n-1)} \left\{ \sum_{i=1}^{n} \left(\tilde{y}_i^{*(j)} - \mu \right)^2 - n \left(\frac{1}{n} \sum_{k=1}^{n} \tilde{y}_k^{*(j)} - \mu \right)^2 \right\}
\end{aligned}
$$

where $\tilde{y}_i^* = \delta_i y_i + (1-\delta_i) y_i^{*(j)}$, we have

$$
\begin{aligned}
E\left\{ \hat{V}_I^{(j)}(\hat{\theta}) \right\} &= \frac{1}{n(n-1)} \left\{ \sum_{i=1}^{n} E\left(\tilde{y}_i^{*(j)} - \mu \right)^2 - nV\left(\frac{1}{n} \sum_{k=1}^{n} \tilde{y}_k^{*(j)} \right) \right\} \\
&\cong n^{-1}\sigma^2
\end{aligned}
$$

which shows that $E(W_m) \cong V(\hat{\theta}_n)$.

- Also,

$$
\begin{aligned}
E(B_m) &= V\left(\hat{\theta}_I^{*(1)}\right) - Cov\left(\hat{\theta}_I^{*(1)}, \hat{\theta}_I^{*(2)}\right) \\
&= V\left\{\frac{n-r}{n}\left(\mu^{*(1)} - \bar{y}_r\right) + \frac{1}{n}\sum_{i=r+1}^{n}\left(y_i^{*(1)} - \mu^{*(1)}\right)\right\} \\
&\cong \left(\frac{n-r}{n}\right)^2\left(\frac{1}{r} + \frac{1}{n-r}\right)\sigma^2 \\
&= \left(\frac{1}{r} - \frac{1}{n}\right)\sigma^2.
\end{aligned}
$$

Thus, Rubin's variance estimator satisfies

$$
E\left\{\hat{V}_{MI}(\hat{\theta}_{MI})\right\} \cong \frac{1}{r}\sigma^2 + \frac{1}{m}\left(\frac{n-r}{n}\right)^2\left(\frac{1}{r}\sigma^2 + \frac{1}{n-r}\sigma^2\right) \cong V\left(\hat{\theta}_{MI}\right),
$$

which shows the asymptotic unbiasedness of the Rubin's variance estimator.