5.3 Some Examples

Jae-Kwang Kim

Example 1

- 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,

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

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 $\sigma^2(n-r)/(n^2 m)$, and the three terms are mutually independent. Thus, the variance of $\hat{\theta}_{MI}$ is

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

- 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.

Example 2 Bayesian Bootstrap (Rubin, 1981)

- We first assume that an element of the population takes one of the values $d_1, \cdots, d_K$ with probability $p_1, \cdots, p_K$, respectively. That is, we assume

$$P(Y = d_k) = p_k, \quad \sum_{k=1}^{K} p_k = 1. \tag{2}$$

  Let $y_1, \cdots, y_n$ be an IID sample from (2) and let n_k be the number of y_i equal to d_k.

- The parameter is a vector of probabilities $\mathbf{p} = (p_1, \cdots, p_K)$, such that $\sum_{i=1}^{K} p_i = 1$. In this case, the population mean $\theta = E(Y)$ can be expressed as $\theta = \sum_{i=1}^{K} p_i d_i$ and we need only estimate $\mathbf{p}$.

- Prior distribution for $\mathbf{p} = (p_1, \cdots, p_K)$

$$\pi(\mathbf{p} \mid \boldsymbol{\alpha}) \propto \prod_{k=1}^{K} p_k^{\alpha_0 - 1}$$

  which is Dirichlet distribution with parameter $\boldsymbol{\alpha} = \alpha_0(1, \cdots, 1)'$.
- The posterior distribution of $\mathbf{p}$ is proportional to

$$p(\mathbf{p} \mid n_1, \cdots, n_K) \propto \prod_{k=1}^{K} p_k^{n_k + \alpha_0 - 1} \tag{3}$$

  which is Dirichlet distribution with parameter $(n_1 + \alpha_0, \cdots, n_K + \alpha_0)$.
- The posterior expectation of $\theta = \sum_{i=1}^{K} p_i d_i$ is

$$E\{\theta \mid \text{data}\} = \sum_{i=1}^{K} E(p_i \mid \text{data}) d_i$$

  where $E(p_i \mid n_1, \cdots, n_K) = \frac{n_i + \alpha_0}{\sum_{i=1}^{K}(n_i + \alpha_0)}$.

- Rubin's Bayesian bootstrap is based on the posterior distribution in (3) with $K = n$ and $\alpha_0 = 0$.
- To implement the Bayesian bootstrap to multiple imputation, assume that the first r elements are observed and the remaining $n - r$ elements are missing. The Bayesian bootstrap method can be used to generate the imputed values with the following steps:
  1. [Step 1] From $\mathbf{y}_r = (y_1, \cdots, y_r)$, generate $\mathbf{p}_r^* = (p_1^*, \cdots, p_r^*)$ from Dirichlet distribution with parameter $(1, \cdots, 1)$.
  2. [Step 2] Select the imputed value of y_i by

$$y_i^* = \begin{cases} y_1 & \text{with probability } p_1^* \\ \vdots & \vdots \\ y_r & \text{with probability } p_r^* \end{cases}$$

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

- MI point estimator of $\theta = E(Y)$:

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

where

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

- Define

$$\mu^{*(j)} = \sum_{i=1}^{r} p_i^{*(j)} y_i$$

where $\mathbf{p}^{*(j)}$ is the j-th draw from Dirichlet distribution with parameter $(1, \cdots, 1)$.

- Then,

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

- Under the assumption that y_i are IID with (μ, σ^2), we can show that

$$V\left(\hat{\theta}_{MI}\right) \doteq \frac{1}{r}\sigma^2 + \frac{1}{m}\left(\frac{n-r}{n}\right)^2 \left(\frac{1}{r}\sigma^2 + \frac{1}{n-r}\sigma^2\right). \tag{4}$$

- Thus, the MI estimator under Bayesian bootstrap method is asymptotically equivalent to that under normality.

Remark

Rubin and Schenker (1986) proposed an approximation of this Bayesian bootstrap method, called the approximate Bayesian boostrap (ABB) method, which provides an alternative approach of generating imputed values from the empirical distribution. The ABB method can be described as follows:

1. From $\mathbf{y}_r = (y_1, \cdots, y_r)$, generate a donor set $\mathbf{y}_r^* = (y_1^*, \cdots, y_r^*)$ by bootstrapping. That is, we select

$$y_i^* = \begin{cases} y_1 & \text{with probability } 1/r \\ \dots & \dots \\ y_r & \text{with probability } 1/r \end{cases}$$

independently for each $i = 1, \cdots, r$.

2. From the donor set $\mathbf{y}_r^* = (y_1^*, \cdots, y_r^*)$, select an imputed value of y_i by

$$y_i^{**} = \begin{cases} y_1^* & \text{with probability } 1/r \\ \dots & \dots \\ y_r^* & \text{with probability } 1/r \end{cases}$$

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

Example 3 (Normal Regression)

- Regression model

$$
\begin{aligned}
y_i &= \mathbf{x}_i'\boldsymbol{\beta} + e_i, \\
e_i &\overset{i.i.d.}{\sim} N\left(0, \sigma^2\right).
\end{aligned} \tag{5}
$$

The p-dimensional $\mathbf{x}_i$'s are observed on the complete sample and are assumed to be fixed.

- We assume that the first r units are the respondents. Let $\mathbf{y}_r = (y_1, y_2, \cdots, y_r)'$ and $X_r = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_r)'$. Also, let $\mathbf{y}_{n-r} = (y_{r+1}, y_{r+2}, \cdots, y_n)'$ and $X_{n-r} = (\mathbf{x}_{r+1}, \mathbf{x}_{r+2}, \cdots, \mathbf{x}_n)'$.

Recall that, under complete response (i.e. $r = n$)

$$\begin{aligned} p(\beta, \sigma^2 \mid y) &\propto (\sigma^2)^{-n/2-1} \exp\left(-\frac{1}{2\sigma^2}\|y - X\beta\|^2\right) \\ &= (\sigma^2)^{-(n-p)/2-1} \exp\left(-\frac{1}{2\sigma^2}\|y - X\hat{\beta}\|^2\right) \\ &\quad \times(\sigma^2)^{-p/2} \exp\left(-\frac{1}{2\sigma^2}(\beta - \hat{\beta})^\top (X^\top X)(\beta - \hat{\beta})\right), \end{aligned}$$

where $\hat{\beta} = (X^\top X)^{-1} X^\top y$. Since $p(\beta \mid y, \sigma^2) = N(\beta \mid \hat{\beta}, \sigma^2 (X^\top X)^{-1})$,

$$p(\sigma^2 \mid y) = \frac{p(\beta, \sigma^2 \mid y)}{p(\beta \mid y, \sigma^2)} = (\sigma^2)^{-(n-p)/2-1} \exp\left(-\frac{1}{2\sigma^2}\|y - X\hat{\beta}\|^2\right).$$

This implies that

$$\sigma^2 \mid y \sim \text{Inverse-Gamma}\left(\frac{n-p}{2}, \frac{\|y - X\hat{\beta}\|^2}{2}\right).$$

The suggested method of Bayesian imputation is as follows:

1. Draw

$$\sigma^{*2} \mid \mathbf{y}_r \overset{i.i.d.}{\sim} \sim \text{Inverse-Gamma}\left(\frac{r-p}{2}, \frac{(r-p)\hat{\sigma}_r^2}{2}\right), \tag{6}$$

   and

$$\boldsymbol{\beta}^* \mid (\mathbf{y}_r, \sigma^*) \overset{i.i.d.}{\sim} N\left(\hat{\boldsymbol{\beta}}_r, (X_r'X_r)^{-1}\sigma^{*2}\right), \tag{7}$$

   where $\hat{\sigma}_r^2 = (r-p)^{-1}\mathbf{y}_r'\left[I - X_r(X_r'X_r)^{-1}X_r'\right]\mathbf{y}_r$ and $\hat{\boldsymbol{\beta}}_r = (X_r'X_r)^{-1}X_r'\mathbf{y}_r$.

2. For each missing unit $j = r+1, \cdots, n$, draw

$$e_j^* \mid (\boldsymbol{\beta}^*, \sigma^*) \overset{i.i.d.}{\sim} N\left(0, \sigma^{*2}\right).$$

   Then, $y_j^* = \mathbf{x}_j\boldsymbol{\beta}^* + e_j^*$ is the imputed value associated with unit j.

At the j-th repetition of multiple imputation ($j = 1, ..., m$), we can calculate the imputed version of the full sample estimators

$$\hat{\boldsymbol{\beta}}_{I(j)} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \left\{ \sum_{i=1}^{r} \mathbf{x}_i y_i + \sum_{i=r+1}^{n} \mathbf{x}_i y_i^{*(j)} \right\}$$

and

$$\hat{V}_{I(j)} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \hat{\sigma}^2_{I(j)},$$

where

$$\hat{\sigma}^2_{I(j)} = (n-p)^{-1} \left\{ \sum_{i=1}^{r} \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_{I(j)} \right)^2 + \sum_{i=r+1}^{n} \left(y_i^{*(j)} - \mathbf{x}_i' \hat{\boldsymbol{\beta}}_{I(j)} \right)^2 \right\}.$$

The proposed point estimator for the regression coefficient based on m repeated imputations is

$$\hat{\boldsymbol{\beta}}_{MI} = \frac{1}{m} \sum_{j=1}^{m} \hat{\boldsymbol{\beta}}_{I(j)} \tag{8}$$

and the proposed estimator for the variance of $\hat{\boldsymbol{\beta}}_{MI}$ is given by

$$\hat{V}_{MI} = W_m + \left(1 + m^{-1}\right) B_m \tag{9}$$

with

$$W_m = \frac{1}{m} \sum_{j=1}^{m} \hat{V}_{I(j)}. \tag{10}$$

Since we can write

$$\hat{\boldsymbol{\beta}}_{MI} = \lim_{M\to\infty} \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}'\right)^{-1} \left[\sum_{i=1}^{r} \mathbf{x}_i y_i + \sum_{i=r+1}^{n} \mathbf{x}_i \left(m^{-1} \sum_{k=1}^{m} y_i^{*(j)}\right)\right],$$

where

$$y_i^{*(j)} = \mathbf{x}_i' \hat{\boldsymbol{\beta}}_r + \mathbf{x}_i' \left(\boldsymbol{\beta}^{*(j)} - \hat{\boldsymbol{\beta}}_r\right) + e_i^{*(j)},$$

we can decompose it into three independent components as

$$\hat{\boldsymbol{\beta}}_{MI} = \hat{\boldsymbol{\beta}}_r + \frac{1}{m} \sum_{j=1}^{m} \sum_{i=r+1}^{n} \mathbf{h}_i \mathbf{x}_i' \left(\boldsymbol{\beta}^{*(j)} - \hat{\boldsymbol{\beta}}_r\right) + \frac{1}{m} \sum_{j=1}^{m} \sum_{i=r+1}^{n} \mathbf{h}_i e_i^{*(j)}, \qquad (11)$$

where $\hat{\boldsymbol{\beta}}_r = \left(X_r' X_r\right)^{-1} X_r' \mathbf{y}_r$, $\mathbf{h}_i = \left(X_n' X_n\right)^{-1} \mathbf{x}_i$ and $\boldsymbol{\beta}^{*(j)}$ is the j-th realization of the parameter values generated from posterior distribution (7).

The total variance is

$$\begin{aligned} V\left(\hat{\beta}_{MI}\right) &\cong \left(X_r'X_r\right)^{-1}\sigma^2 \\ &+ m^{-1}\left(X_n'X_n\right)^{-1}\left(X_{n-r}'X_{n-r}\right)\left(X_r'X_r\right)^{-1}\left(X_{n-r}'X_{n-r}\right)\left(X_n'X_n\right)^{-1}\sigma^2 \\ &+ m^{-1}\left(X_n'X_n\right)^{-1}\left(X_{n-r}'X_{n-r}\right)\left(X_n'X_n\right)^{-1}\sigma^2. \end{aligned}$$

Using some matrix algebra,

$$\begin{aligned} \left(X_r'X_r\right)^{-1} &= \left(X_n'X_n - X_{n-r}'X_{n-r}\right)^{-1} \\ &= \left(X_n'X_n\right)^{-1} + \left(X_n'X_n\right)^{-1}\left(X_{n-r}'X_{n-r}\right)\left(X_n'X_n\right)^{-1} \\ &+ \left(X_n'X_n\right)^{-1}\left(X_{n-r}'X_{n-r}\right)\left(X_r'X_r\right)^{-1}\left(X_{n-r}'X_{n-r}\right)\left(X_n'X_n\right)^{-1}, \end{aligned}$$

we can write

$$V\left(\hat{\beta}_{MI}\right) \cong \left(X_r'X_r\right)^{-1}\sigma^2 + m^{-1}\left\{\left(X_r'X_r\right)^{-1} - \left(X_n'X_n\right)^{-1}\right\}\sigma^2.$$

Also, it can be shown that

$$E(W_m) \cong (X_n'X_n)^{-1}\sigma^2$$

and, using Lemma 5.2,

$$\begin{aligned} E(B_m) &\cong (X_n'X_n)^{-1}\left(X_{n-r}'X_{n-r}\right)(X_n'X_n)^{-1}\sigma^2 \\ &+ (X_n'X_n)^{-1}\left(X_{n-r}'X_{n-r}\right)(X_r'X_r)^{-1}\left(X_{n-r}'X_{n-r}\right)(X_n'X_n)^{-1}\sigma^2 \\ &= \left\{(X_r'X_r)^{-1} - (X_n'X_n)^{-1}\right\}\sigma^2. \end{aligned}$$

Thus, the asymptotic unbiasedness of the Rubin's variance estimator can be established.