Statistical Methods for Analysis with Missing Data

Lecture 10: multiple imputation

Mauricio Sadinle

Department of Biostatistics

UNIVERSITY of WASHINGTON

Previous Lectures

- Introduction to Bayesian inference
- Gibbs sampling from posterior distributions
- General setup for Bayesian inference with missing data
- Ignorability for Bayesian inference (Definition 5.12 in Daniels & Hogan, 2008):
  - MAR
  - Separability: the full-data parameter ϑ can be decomposed as $\vartheta = (\theta, \psi)$, where θ indexes the study-variables model and ψ indexes the response mechanism
  - $\theta \perp\!\!\!\perp \psi$ a priori
- Data augmentation to handle missing data in Bayesian inference

Today's Lecture

Different flavors of *multiple imputation*

- Proper multiple imputation
- Multiple imputation by chained equations

Outline

Multiple Imputation

Single imputation is appealing because of its simplicity, but we shouldn't treat the imputed data as if it was all observed data

- Remember: single imputation leads to overconfidence in results, underestimation of standard errors
- Idea: maybe we can account for the extra uncertainty coming from the fact that we are imputing the missing data

Rubin (1987, *Multiple Imputation for Nonresponse in Surveys*, Wiley) proposed:

- For each individual, *randomly impute* the missing values M times to create M completed datasets
- Run the analysis of interest on each of these M completed datasets
- Combine the results from the M analyses

Multiple Imputation

Single imputation is appealing because of its simplicity, but we shouldn't treat the imputed data as if it was all observed data

- Remember: single imputation leads to overconfidence in results, underestimation of standard errors
- Idea: maybe we can account for the extra uncertainty coming from the fact that we are imputing the missing data

Rubin (1987, *Multiple Imputation for Nonresponse in Surveys*, Wiley) proposed:

- For each individual, *randomly impute* the missing values M times to create M completed datasets
- Run the analysis of interest on each of these M completed datasets
- Combine the results from the M analyses

Multiple Imputation

Single imputation is appealing because of its simplicity, but we shouldn't treat the imputed data as if it was all observed data

- Remember: single imputation leads to overconfidence in results, underestimation of standard errors
- Idea: maybe we can account for the extra uncertainty coming from the fact that we are imputing the missing data

Rubin (1987, *Multiple Imputation for Nonresponse in Surveys*, Wiley) proposed:

- For each individual, *randomly impute* the missing values M times to create M completed datasets
- Run the analysis of interest on each of these M completed datasets
- Combine the results from the M analyses

Multiple Imputation

Single imputation is appealing because of its simplicity, but we shouldn't treat the imputed data as if it was all observed data

- Remember: single imputation leads to overconfidence in results, underestimation of standard errors
- Idea: maybe we can account for the extra uncertainty coming from the fact that we are imputing the missing data

Rubin (1987, *Multiple Imputation for Nonresponse in Surveys*, Wiley) proposed:

- For each individual, *randomly impute* the missing values M times to create M completed datasets
- Run the analysis of interest on each of these M completed datasets
- Combine the results from the M analyses

Proper Multiple Imputation under Ignorability

- For each individual, randomly impute the missing values M times to create M completed datasets
  - Sample $\{\mathbf{Z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
  - Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Run the analysis of interest on each of these M completed datasets

$$\{\ \hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),\ \ \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})]\ \}_{m=1}^{M}$$

- Combine the results of the M analyses

$$\hat{\theta}_{\text{MI}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),$$

$$\hat{V}(\hat{\theta}_{\text{MI}}) = \frac{1}{M}\sum_{m=1}^{M}\hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] + \left(1+\frac{1}{M}\right)\frac{1}{M-1}\sum_{m=1}^{M}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\text{MI}}][\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\text{MI}}]^{T}$$

Proper Multiple Imputation under Ignorability

- For each individual, randomly impute the missing values M times to create M completed datasets
  - Sample $\{\mathbf{Z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
  - Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Run the analysis of interest on each of these M completed datasets

$$\{\ \hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),\ \ \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})]\ \}_{m=1}^{M}$$

- Combine the results of the M analyses

$$\hat{\theta}_{\mathrm{MI}} = \frac{1}{M}\sum_{m=1}^{M} \hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),$$

$$\hat{V}(\hat{\theta}_{\mathrm{MI}}) = \frac{1}{M}\sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] + \left(1+\frac{1}{M}\right)\frac{1}{M-1}\sum_{m=1}^{M}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\mathrm{MI}}][\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\mathrm{MI}}]^{T}$$

Proper Multiple Imputation under Ignorability

- For each individual, randomly impute the missing values M times to create M completed datasets
  - Sample $\{\mathbf{Z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
  - Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Run the analysis of interest on each of these M completed datasets

$$\{\ \hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),\ \ \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})]\ \}_{m=1}^{M}$$

- Combine the results of the M analyses

$$\hat{\theta}_{\mathrm{MI}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),$$

$$\hat{V}(\hat{\theta}_{\mathrm{MI}}) = \frac{1}{M} \sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \frac{1}{M-1} \sum_{m=1}^{M} [\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\mathrm{MI}}][\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\mathrm{MI}}]^T$$

Proper Multiple Imputation under Ignorability

- For each individual, randomly impute the missing values M times to create M completed datasets
  - Sample $\{\mathbf{Z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
  - Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Run the analysis of interest on each of these M completed datasets

$$\{\ \hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),\ \ \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})]\ \}_{m=1}^{M}$$

- Combine the results of the M analyses

$$\hat{\theta}_{\text{MI}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),$$

$$\hat{V}(\hat{\theta}_{\text{MI}}) = \frac{1}{M}\sum_{m=1}^{M}\hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] +$$

$$\left(1+\frac{1}{M}\right)\frac{1}{M-1}\sum_{m=1}^{M}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\text{MI}}][\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\text{MI}}]^{T}$$

Proper Multiple Imputation under Ignorability

- For each individual, randomly impute the missing values M times to create M completed datasets
  - Sample $\{\mathbf{Z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
  - Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Run the analysis of interest on each of these M completed datasets

$$\{ \ \hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}), \ \ \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] \ \}_{m=1}^{M}$$

- Combine the results of the M analyses

$$\hat{\theta}_{\text{MI}} = \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}),$$

$$\hat{V}(\hat{\theta}_{\text{MI}}) = \frac{1}{M} \sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \frac{1}{M-1} \sum_{m=1}^{M} [\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\text{MI}}][\hat{\theta}_{\text{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}) - \hat{\theta}_{\text{MI}}]^T$$

Multiple Imputation

What's the justification for this procedure?

- MI can be justified from a Bayesian point of view
- Actual practice of MI is an approximation of the Bayesian procedure

Bayesian Derivation of Multiple Imputation

Recall that, given a prior $p(\theta, \phi)$, Bayesian inferences are based on the posterior distribution

$$p(\theta, \phi \mid \mathbf{r}, \mathbf{z_r}) = \frac{p(\theta, \phi) \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta) \; d\mathbf{z}_{\bar{\mathbf{r}}}}{p(\mathbf{r}, \mathbf{z_r})}$$

Bayesian Derivation of Multiple Imputation

$$
\begin{aligned}
p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) &= \frac{p(\theta, \phi) \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta) \; d\mathbf{z}_{\bar{\mathbf{r}}}}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}), \text{ with } \{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_m \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})
\end{aligned}
$$

Bayesian Derivation of Multiple Imputation

$$
\begin{aligned}
p(\theta, \phi \mid \mathbf{r}, \mathbf{z_r}) &= \frac{p(\theta, \phi) \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta) \; d\mathbf{z}_{\bar{\mathbf{r}}}}{p(\mathbf{r}, \mathbf{z_r})} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z_r})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z_r})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z_r})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\theta, \phi \mid \mathbf{r}, \mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}) p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z_r}) \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}), \;\; \text{with } \{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_m \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z_r})
\end{aligned}
$$

Bayesian Derivation of Multiple Imputation

$$
\begin{aligned}
p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) &= \frac{p(\theta, \phi) \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta) \; d\mathbf{z}_{\bar{\mathbf{r}}}}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}), \;\; \text{with } \{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_m \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})
\end{aligned}
$$

Bayesian Derivation of Multiple Imputation

$$
\begin{aligned}
p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) &= \frac{p(\theta, \phi) \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta) \ d\mathbf{z}_{\bar{\mathbf{r}}}}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \ d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z})} \ d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \ d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \ d\mathbf{z}_{\bar{\mathbf{r}}} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}), \ \text{with } \{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_m \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})
\end{aligned}
$$

Bayesian Derivation of Multiple Imputation

$$
\begin{aligned}
p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) &= \frac{p(\theta, \phi) \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta) \, d\mathbf{z}_{\bar{\mathbf{r}}}}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \, d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z})} \, d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \, d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \, d\mathbf{z}_{\bar{\mathbf{r}}} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}), \text{ with } \{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_m \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})
\end{aligned}
$$

Bayesian Derivation of Multiple Imputation

$$
\begin{aligned}
p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) &= \frac{p(\theta, \phi) \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta) \; d\mathbf{z}_{\bar{\mathbf{r}}}}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} \frac{p(\theta, \phi) p(\mathbf{r} \mid \mathbf{z}, \phi) p(\mathbf{z} \mid \theta)}{p(\mathbf{r}, \mathbf{z})} \frac{p(\mathbf{r}, \mathbf{z})}{p(\mathbf{r}, \mathbf{z}_{\mathbf{r}})} \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&= \int_{\mathcal{Z}_{\bar{\mathbf{r}}}} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \; d\mathbf{z}_{\bar{\mathbf{r}}} \\
&\approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)}), \;\text{ with } \{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_m \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})
\end{aligned}
$$

Bayesian Derivation of Multiple Imputation

A purely Bayesian version of MI (without assuming ignorability):

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

"In a nutshell, Rubin's MI is simply a size m Monte Carlo simulation" (Xie & Meng, 2017, Statistica Sinica)[1]

[1] http://dx.doi.org/10.5705/ss.2014.067

Bayesian Derivation of Multiple Imputation

A purely Bayesian version of MI (without assuming ignorability):

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

"In a nutshell, Rubin's MI is simply a size m Monte Carlo simulation" (Xie & Meng, 2017, Statistica Sinica)[1]

[1] http://dx.doi.org/10.5705/ss.2014.067

Bayesian Derivation of Multiple Imputation

A purely Bayesian version of MI (without assuming ignorability):

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

"*In a nutshell, Rubin's MI is simply a size m Monte Carlo simulation*" (Xie & Meng, 2017, Statistica Sinica)[1]

[1] http://dx.doi.org/10.5705/ss.2014.067

Bayesian Derivation of Multiple Imputation

A purely Bayesian version of MI (without assuming ignorability):

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

"*In a nutshell, Rubin's MI is simply a size m Monte Carlo simulation*" (Xie & Meng, 2017, Statistica Sinica)[1]

[1] http://dx.doi.org/10.5705/ss.2014.067

Bayesian Derivation of Multiple Imputation

A purely Bayesian version of MI (without assuming ignorability):

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta, \phi \mid \mathbf{r}, \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

"*In a nutshell, Rubin's MI is simply a size m Monte Carlo simulation*" (Xie & Meng, 2017, Statistica Sinica)[1]

[1] http://dx.doi.org/10.5705/ss.2014.067

Bayesian Derivation of MI under Ignorability

A purely Bayesian version of MI under ignorability:

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta \mid \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

For the rest of today's session we will assume ignorability

Bayesian Derivation of MI under Ignorability

A purely Bayesian version of MI under ignorability:

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta \mid \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

For the rest of today's session we will assume ignorability

Bayesian Derivation of MI under Ignorability

A purely Bayesian version of MI under ignorability:

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta \mid \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

For the rest of today's session we will assume ignorability

Bayesian Derivation of MI under Ignorability

A purely Bayesian version of MI under ignorability:

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta \mid \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

For the rest of today's session we will assume ignorability

Bayesian Derivation of MI under Ignorability

A purely Bayesian version of MI under ignorability:

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$
- Create M completed datasets $\{(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\}_{m=1}^{M}$
- Obtain posteriors using each completed dataset $p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$
- Combine individual posteriors

$$p(\theta \mid \mathbf{z}_{\mathbf{r}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})$$

For the rest of today's session we will assume ignorability

Bayesian Derivation of MI under Ignorability

A further approximation based on Bernstein–von Mises theorem

- Heuristically, we say that asymptotically

$$p(\theta \mid \mathbf{z}_{\mathbf{r}}) \approx N(\hat{\theta}_{\mathrm{MLE}}, \hat{V}[\hat{\theta}_{\mathrm{MLE}}])$$

- Therefore, for large sample sizes, instead of averaging the individual posteriors we only need the posterior mean and covariance

$$\begin{aligned}
\hat{\theta}_{\mathrm{MLE}} &\approx E(\theta \mid \mathbf{z}_{\mathbf{r}}) \\
&= E_{\mathbf{z}_{\bar{\mathbf{r}}}}[E(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) \mid \mathbf{z}_{\mathbf{r}}] \\
&\approx \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})
\end{aligned}$$

$$\begin{aligned}
\hat{V}[\hat{\theta}_{\mathrm{MLE}}] &\approx V(\theta \mid \mathbf{z}_{\mathbf{r}}) \\
&= E_{\mathbf{z}_{\bar{\mathbf{r}}}}[V(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) \mid \mathbf{z}_{\mathbf{r}}] + V_{\mathbf{z}_{\bar{\mathbf{r}}}}[E(\theta \mid \mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}) \mid \mathbf{z}_{\mathbf{r}}] \\
&\approx \frac{1}{M} \sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \hat{V}_{\mathbf{z}_{\bar{\mathbf{r}}}}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{\mathbf{r}}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})]
\end{aligned}$$

Bayesian Derivation of MI under Ignorability

A further approximation based on Bernstein–von Mises theorem

- Heuristically, we say that asymptotically

$$p(\theta \mid \mathbf{z_r}) \approx N(\hat{\theta}_{\text{MLE}}, \hat{V}[\hat{\theta}_{\text{MLE}}])$$

- Therefore, for large sample sizes, instead of averaging the individual posteriors we only need the posterior mean and covariance

$$\begin{aligned} \hat{\theta}_{\text{MLE}} &\approx E(\theta \mid \mathbf{z_r}) \\ &= E_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\ &\approx \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)}) \end{aligned}$$

$$\begin{aligned} \hat{V}[\hat{\theta}_{\text{MLE}}] &\approx V(\theta \mid \mathbf{z_r}) \\ &= E_{\mathbf{z_{\bar{r}}}}[V(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] + V_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\ &\approx \frac{1}{M} \sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \hat{V}_{\mathbf{z_{\bar{r}}}}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})] \end{aligned}$$

Bayesian Derivation of MI under Ignorability

A further approximation based on Bernstein–von Mises theorem

- Heuristically, we say that asymptotically

$$p(\theta \mid \mathbf{z_r}) \approx N(\hat{\theta}_{\text{MLE}}, \hat{V}[\hat{\theta}_{\text{MLE}}])$$

- Therefore, for large sample sizes, instead of averaging the individual posteriors we only need the posterior mean and covariance

$$\begin{aligned}
\hat{\theta}_{\text{MLE}} &\approx E(\theta \mid \mathbf{z_r}) \\
&= E_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\
&\approx \frac{1}{M}\sum_{m=1}^{M} \hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})
\end{aligned}$$

$$\begin{aligned}
\hat{V}[\hat{\theta}_{\text{MLE}}] &\approx V(\theta \mid \mathbf{z_r}) \\
&= E_{\mathbf{z_{\bar{r}}}}[V(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] + V_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\
&\approx \frac{1}{M}\sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \hat{V}_{\mathbf{z_{\bar{r}}}}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})]
\end{aligned}$$

Bayesian Derivation of MI under Ignorability

A further approximation based on Bernstein–von Mises theorem

- Heuristically, we say that asymptotically

$$p(\theta \mid \mathbf{z_r}) \approx N(\hat{\theta}_{\text{MLE}}, \hat{V}[\hat{\theta}_{\text{MLE}}])$$

- Therefore, for large sample sizes, instead of averaging the individual posteriors we only need the posterior mean and covariance

$$\begin{aligned}
\hat{\theta}_{\text{MLE}} &\approx E(\theta \mid \mathbf{z_r}) \\
&= E_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\
&\approx \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})
\end{aligned}$$

$$\begin{aligned}
\hat{V}[\hat{\theta}_{\text{MLE}}] &\approx V(\theta \mid \mathbf{z_r}) \\
&= E_{\mathbf{z_{\bar{r}}}}[V(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] + V_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\
&\approx \frac{1}{M} \sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \hat{V}_{\mathbf{z_{\bar{r}}}}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})]
\end{aligned}$$

Bayesian Derivation of MI under Ignorability

A further approximation based on Bernstein–von Mises theorem

- Heuristically, we say that asymptotically

$$p(\theta \mid \mathbf{z_r}) \approx N(\hat{\theta}_{\text{MLE}}, \hat{V}[\hat{\theta}_{\text{MLE}}])$$

- Therefore, for large sample sizes, instead of averaging the individual posteriors we only need the posterior mean and covariance

$$\begin{aligned}\hat{\theta}_{\text{MLE}} &\approx E(\theta \mid \mathbf{z_r}) \\ &= E_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\ &\approx \frac{1}{M}\sum_{m=1}^{M} \hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})\end{aligned}$$

$$\begin{aligned}\hat{V}[\hat{\theta}_{\text{MLE}}] &\approx V(\theta \mid \mathbf{z_r}) \\ &= E_{\mathbf{z_{\bar{r}}}}[V(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] + V_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\ &\approx \frac{1}{M}\sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \hat{V}_{\mathbf{z_{\bar{r}}}}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})]\end{aligned}$$

Bayesian Derivation of MI under Ignorability

A further approximation based on Bernstein–von Mises theorem

- Heuristically, we say that asymptotically

$$p(\theta \mid \mathbf{z_r}) \approx N(\hat{\theta}_{\text{MLE}}, \hat{V}[\hat{\theta}_{\text{MLE}}])$$

- Therefore, for large sample sizes, instead of averaging the individual posteriors we only need the posterior mean and covariance

$$\begin{aligned}\hat{\theta}_{\text{MLE}} &\approx E(\theta \mid \mathbf{z_r}) \\ &= E_{\mathbf{z}_{\bar{\mathbf{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}) \mid \mathbf{z_r}] \\ &\approx \frac{1}{M}\sum_{m=1}^{M} \hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})\end{aligned}$$

$$\begin{aligned}\hat{V}[\hat{\theta}_{\text{MLE}}] &\approx V(\theta \mid \mathbf{z_r}) \\ &= E_{\mathbf{z}_{\bar{\mathbf{r}}}}[V(\theta \mid \mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}) \mid \mathbf{z_r}] + V_{\mathbf{z}_{\bar{\mathbf{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}) \mid \mathbf{z_r}] \\ &\approx \frac{1}{M}\sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \hat{V}_{\mathbf{z}_{\bar{\mathbf{r}}}}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z}_{\bar{\mathbf{r}}}^{(m)})]\end{aligned}$$

Bayesian Derivation of MI under Ignorability

A further approximation based on Bernstein–von Mises theorem

- Heuristically, we say that asymptotically

$$p(\theta \mid \mathbf{z_r}) \approx N(\hat{\theta}_{\text{MLE}}, \hat{V}[\hat{\theta}_{\text{MLE}}])$$

- Therefore, for large sample sizes, instead of averaging the individual posteriors we only need the posterior mean and covariance

$$\begin{aligned}
\hat{\theta}_{\text{MLE}} &\approx E(\theta \mid \mathbf{z_r}) \\
&= E_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\
&\approx \frac{1}{M} \sum_{m=1}^{M} \hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})
\end{aligned}$$

$$\begin{aligned}
\hat{V}[\hat{\theta}_{\text{MLE}}] &\approx V(\theta \mid \mathbf{z_r}) \\
&= E_{\mathbf{z_{\bar{r}}}}[V(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] + V_{\mathbf{z_{\bar{r}}}}[E(\theta \mid \mathbf{z_r}, \mathbf{z_{\bar{r}}}) \mid \mathbf{z_r}] \\
&\approx \frac{1}{M} \sum_{m=1}^{M} \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})] + \left(1 + \frac{1}{M}\right) \hat{V}_{\mathbf{z_{\bar{r}}}}[\hat{\theta}_{\text{MLE}}(\mathbf{z_r}, \mathbf{z_{\bar{r}}}^{(m)})]
\end{aligned}$$

Bayesian Derivation of MI under Ignorability

Where does $\left(1 + \frac{1}{M}\right)$ come from?

- Adjustment for finite number of imputations
- Derived under an extra set of assumptions (Section 3.3 of Rubin (1987))
- Negligible for a moderate number of imputations

Comments on Multiple Imputation

Wait a second!

- What is $p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$ and how do we sample from it?

$$p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}) = \int_{\theta} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, \theta) p(\theta \mid \mathbf{z}_{\mathbf{r}}) d\theta$$

- So, to obtain a draw from conditional distribution $\mathbf{Z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}$ we can
  - Draw $\theta^{(m)}$ from $p(\theta \mid \mathbf{z}_{\mathbf{r}})$
  - Draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$ from $p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, \theta^{(m)})$

Comments on Multiple Imputation

Wait a second!

- What is $p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$ and how do we sample from it?

$$p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}) = \int_{\theta} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, \theta) p(\theta \mid \mathbf{z}_{\mathbf{r}}) d\theta$$

- So, to obtain a draw from conditional distribution $\mathbf{Z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}$ we can
  - Draw $\theta^{(m)}$ from $p(\theta \mid \mathbf{z}_{\mathbf{r}})$
  - Draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$ from $p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, \theta^{(m)})$

Comments on Multiple Imputation

Therefore, to approximate $p(\theta \mid \mathbf{z_r})$ via MI, part of what we need to do is

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z_r})$ to create M completed datasets
- To obtain a draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$, we need to
  - Draw $\theta^{(m)}$ from $p(\theta \mid \mathbf{z_r})$
  - Draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$ from $p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z_r}, \theta^{(m)})$

Confused anyone?

Comments on Multiple Imputation

Therefore, to approximate $p(\theta \mid \mathbf{z_r})$ via MI, part of what we need to do is

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z_r})$ to create M completed datasets
- To obtain a draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$, we need to
  - Draw $\theta^{(m)}$ from $p(\theta \mid \mathbf{z_r})$
  - Draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$ from $p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z_r}, \theta^{(m)})$

Confused anyone?

Comments on Multiple Imputation

Therefore, to approximate $p(\theta \mid \mathbf{z_r})$ via MI, part of what we need to do is

- Sample $\{\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}\}_{m=1}^{M} \overset{iid}{\sim} p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z_r})$ to create M completed datasets
- To obtain a draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$, we need to
  - Draw $\theta^{(m)}$ from $p(\theta \mid \mathbf{z_r})$
  - Draw $\mathbf{z}_{\bar{\mathbf{r}}}^{(m)}$ from $p(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z_r}, \theta^{(m)})$

Confused anyone?

Comments on Multiple Imputation

- You need to sample from $p(\theta \mid \mathbf{z_r})$ to approximate $p(\theta \mid \mathbf{z_r})$ via MI
  - If you can directly work with $p(\theta \mid \mathbf{z_r})$, then MI seems pointless (e.g., if you are doing the imputation and the analysis)
- Rubin's motivation for MI:
  - A statistical agency needs to publish a dataset with missingness
  - It instead publishes M completed datasets
  - Analysts run analyses on each completed dataset and combine results
  - Analysts don't have to worry about the missing data problem

Comments on Multiple Imputation

- You need to sample from $p(\theta \mid \mathbf{z_r})$ to approximate $p(\theta \mid \mathbf{z_r})$ via MI
  - If you can directly work with $p(\theta \mid \mathbf{z_r})$, then MI seems pointless (e.g., if you are doing the imputation and the analysis)
- Rubin's motivation for MI:
  - A statistical agency needs to publish a dataset with missingness
  - It instead publishes M completed datasets
  - Analysts run analyses on each completed dataset and combine results
  - Analysts don't have to worry about the missing data problem

Comments on Multiple Imputation

Just a “tiny” detail:

- Analysts don’t usually use the same model used by the imputer!
- Models might be *uncongenial*

Outline

Uncongeniality

Xiao-Li Meng (1994, Statistical Science) described this issue:

- Multiple imputations are based on an *imputation model*
- Analysts use an *analysis procedure*
- Imputation and analysis might be incompatible or *uncongenial*

Uncongeniality

Xiao-Li Meng (1994, Statistical Science) described this issue:

- Multiple imputations are based on an *imputation model*
- Analysts use an *analysis procedure*
- Imputation and analysis might be incompatible or *uncongenial*

Uncongeniality

Xiao-Li Meng (1994, Statistical Science) described this issue:

- Multiple imputations are based on an *imputation model*
- Analysts use an *analysis procedure*
- Imputation and analysis might be incompatible or *uncongenial*

Uncongeniality

Meng (1994):

- From the analyst's point of view:
  - $\mathcal{P}_{obs}$: inferential procedure with *incomplete* data, e.g. derived from $L_{obs}(\theta \mid \mathbf{z}_{(\mathbf{r})})$ and $p(\theta)$, summarized by $\{\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})}), \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})})]\}$
  - $\mathcal{P}_{com}$: inferential procedure with *complete* data, e.g. derived from $L(\theta \mid \mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})$ and $p(\theta)$, summarized by $\{\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})}), \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})]\}$
- From the imputer's point of view:
  - Imputations are drawn from a conditional distribution $g(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, A)$, where A corresponds to extra variables that might be available to the imputer but not to the analyst

Uncongeniality

Meng (1994):

- From the analyst's point of view:
  - $\mathcal{P}_{obs}$: inferential procedure with *incomplete* data, e.g. derived from $L_{obs}(\theta \mid \mathbf{z}_{(\mathbf{r})})$ and $p(\theta)$, summarized by $\{\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})}), \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})})]\}$
  - $\mathcal{P}_{com}$: inferential procedure with *complete* data, e.g. derived from $L(\theta \mid \mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})$ and $p(\theta)$, summarized by $\{\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})}), \hat{V}[\hat{\theta}_{\mathrm{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})]\}$
- From the imputer's point of view:
  - Imputations are drawn from a conditional distribution $g(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, A)$, where A corresponds to extra variables that might be available to the imputer but not to the analyst

Uncongeniality

Meng (1994):

- From the analyst's point of view:
  - $\mathcal{P}_{obs}$: inferential procedure with *incomplete* data, e.g. derived from $L_{obs}(\theta \mid \mathbf{z}_{(\mathbf{r})})$ and $p(\theta)$, summarized by $\{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})})]\}$
  - $\mathcal{P}_{com}$: inferential procedure with *complete* data, e.g. derived from $L(\theta \mid \mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})$ and $p(\theta)$, summarized by $\{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})]\}$
- From the imputer's point of view:
  - Imputations are drawn from a conditional distribution $g(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, A)$, where A corresponds to extra variables that might be available to the imputer but not to the analyst

Uncongeniality

Meng (1994):

- *Congeniality* requires the existence of a *Bayesian model* (prior and likelihood) $\mathcal{F}$ such that
  - The posterior mean and variance of θ under $\mathcal{F}$ given *incomplete* data are asymptotically the same as $\mathcal{P}_{obs} \equiv \{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})})]\}$
  - The posterior mean and variance of θ under $\mathcal{F}$ given *complete* data are asymptotically the same as $\mathcal{P}_{com} \equiv \{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})]\}$
  - The posterior predictive density $f(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$ under $\mathcal{F}$ is identical to $g(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, A)$ (from where imputations are drawn)

Uncongeniality

Meng (1994):

- *Congeniality* requires the existence of a *Bayesian model* (prior and likelihood) $\mathcal{F}$ such that
  - The posterior mean and variance of θ under $\mathcal{F}$ given *incomplete* data are asymptotically the same as $\mathcal{P}_{obs} \equiv \{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})})]\}$
  - The posterior mean and variance of θ under $\mathcal{F}$ given *complete* data are asymptotically the same as $\mathcal{P}_{com} \equiv \{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})]\}$
  - The posterior predictive density $f(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$ under $\mathcal{F}$ is identical to $g(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, A)$ (from where imputations are drawn)

Uncongeniality

Meng (1994):

- *Congeniality* requires the existence of a *Bayesian model* (prior and likelihood) $\mathcal{F}$ such that
  - The posterior mean and variance of θ under $\mathcal{F}$ given *incomplete* data are asymptotically the same as $\mathcal{P}_{obs} \equiv \{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})})]\}$
  - The posterior mean and variance of θ under $\mathcal{F}$ given *complete* data are asymptotically the same as $\mathcal{P}_{com} \equiv \{\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})}), \hat{V}[\hat{\theta}_{\text{MLE}}(\mathbf{z}_{(\mathbf{r})}, \mathbf{z}_{(\bar{\mathbf{r}})})]\}$
  - The posterior predictive density $f(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}})$ under $\mathcal{F}$ is identical to $g(\mathbf{z}_{\bar{\mathbf{r}}} \mid \mathbf{z}_{\mathbf{r}}, A)$ (from where imputations are drawn)

Uncongeniality

Xie & Meng (2017, Statistica Sinica):

"*When the imputation model class and the (embedded) analyst's model class differ, the behavior of Rubin's rules becomes very complicated, capable of producing inconsistent variance estimators, a matter that has received recurrent criticisms*"

- Confidence intervals might not be valid (less coverage than desired)

Uncongeniality

Xie & Meng (2017, Statistica Sinica):

"*When the imputation model class and the (embedded) analyst's model class differ, the behavior of Rubin's rules becomes very complicated, capable of producing inconsistent variance estimators, a matter that has received recurrent criticisms*"

- Confidence intervals might not be valid (less coverage than desired)

Uncongeniality

Xie & Meng (2017, Statistica Sinica) conclude:

- Imputation model should be more general than analysis model to obtain valid confidence intervals from Rubin's rules
- "*In general, uncongenality should be regarded as the rule rather than the exception, and a simple confidence valid procedure to combat any degree of uncongenality is to double Rubin's MI variance estimate*"

Uncongeniality

Xie & Meng (2017, Statistica Sinica) conclude:

- Imputation model should be more general than analysis model to obtain valid confidence intervals from Rubin's rules
- *"In general, uncongenality should be regarded as the rule rather than the exception, and a simple confidence valid procedure to combat any degree of uncongenality is to double Rubin's MI variance estimate"*

Uncongeniality

Xie & Meng (2017, Statistica Sinica) conclude:

- Imputation model should be more general than analysis model to obtain valid confidence intervals from Rubin's rules
- "*In general, uncongenality should be regarded as the rule rather than the exception, and a simple confidence valid procedure to combat any degree of uncongenality is to double Rubin's MI variance estimate*"

Outline

Modeling Multivariate Distributions

- The imputation part of multiple imputation requires a model for the joint distribution of the study variables
- Which models are common for multivariate distributions?
  - Multivariate normal for continuous variables
  - Multinomial for categorical variables
- What if we have a mix of variable types?
  - Counts
  - Continuous, some nonnegative, some skewed
  - Categorical, some nominal, some ordinal
  - Mixed type, perhaps zero inflated
- Flexible models for variables of mixed type *do* exist, but they are a current area of research (e.g., Murray & Reiter, JASA 2016)

Modeling Multivariate Distributions

- The imputation part of multiple imputation requires a model for the joint distribution of the study variables

- Which models are common for multivariate distributions?
  - Multivariate normal for continuous variables
  - Multinomial for categorical variables

- What if we have a mix of variable types?
  - Counts
  - Continuous, some nonnegative, some skewed
  - Categorical, some nominal, some ordinal
  - Mixed type, perhaps zero inflated

- Flexible models for variables of mixed type *do* exist, but they are a current area of research (e.g., Murray & Reiter, JASA 2016)

Modeling Multivariate Distributions

- The imputation part of multiple imputation requires a model for the joint distribution of the study variables
- Which models are common for multivariate distributions?
  - Multivariate normal for continuous variables
  - Multinomial for categorical variables
- What if we have a mix of variable types?
  - Counts
  - Continuous, some nonnegative, some skewed
  - Categorical, some nominal, some ordinal
  - Mixed type, perhaps zero inflated
- Flexible models for variables of mixed type *do* exist, but they are a current area of research (e.g., Murray & Reiter, JASA 2016)

Modeling Conditional Distributions

On the other hand, we know a lot about, and have a lot of software for, modeling response variables of different types in a regression context

- Continuous response: linear regression
- Binary response: logistic regression
- In general: generalized linear models

Imputing One Variable

Say $Z = (Y_1, \ldots, Y_K)$, and only Y_1 is subject to missingness

- We only need to model $Y_1 \mid Y_2, \ldots, Y_K$, say using

$$p(y_1 \mid y_2, \ldots, y_K, \theta)$$

- To impute missing Y_1's via multiple imputation, we need to
  - Draw $\theta^{(m)}$ from $p(\theta \mid \mathbf{z}_\mathbf{r}) \propto p(\theta) \prod_{i:r_i=1} p(y_{i1} \mid y_{i2}, \ldots, y_{iK}, \theta)$
  - Draw $y_{i1}^{(m)}$ from $p(y_1 \mid y_{i2}, \ldots, y_{iK}, \theta^{(m)})$ if y_{i1} is missing

Imputing One Variable

Say $Z = (Y_1, \ldots, Y_K)$, and only Y_1 is subject to missingness

- We only need to model $Y_1 \mid Y_2, \ldots, Y_K$, say using

$$p(y_1 \mid y_2, \ldots, y_K, \theta)$$

- To impute missing Y_1's via multiple imputation, we need to
  - Draw $\theta^{(m)}$ from $p(\theta \mid \mathbf{z_r}) \propto p(\theta) \prod_{i:r_i=1} p(y_{i1} \mid y_{i2}, \ldots, y_{iK}, \theta)$
  - Draw $y_{i1}^{(m)}$ from $p(y_1 \mid y_{i2}, \ldots, y_{iK}, \theta^{(m)})$ if y_{i1} is missing

Multivariate Imputation by Chained Equations

Multivariate Imputation by Chained Equations (MICE) (van Buuren 2007, van Buuren and Groothuis-Oudshoorn 2011[2]) is an ad-hoc multiple imputation procedure that builds on the previous idea

- If each $Y_1, \ldots, Y_K$ is subject to missingness, we can posit K different regression models

$$
\begin{gathered}
p_1(y_1 \mid y_{-1}, \theta_1) \\
p_2(y_2 \mid y_{-2}, \theta_2) \\
\vdots \\
p_K(y_K \mid y_{-K}, \theta_K)
\end{gathered}
$$

- θ_k: parameters of the kth conditional distribution
- $y_{-k} = (y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_K)$
- Key idea: use these models to sequentially impute, one variable at a time. Repeat this over a number of iterations

[2] https://www.jstatsoft.org/article/view/v045i03/v45i03.pdf

Multivariate Imputation by Chained Equations

Multivariate Imputation by Chained Equations (MICE) (van Buuren 2007, van Buuren and Groothuis-Oudshoorn 2011[2]) is an ad-hoc multiple imputation procedure that builds on the previous idea

- If each $Y_1, \ldots, Y_K$ is subject to missingness, we can posit K different regression models

$$
\begin{gathered}
p_1(y_1 \mid y_{-1}, \theta_1) \\
p_2(y_2 \mid y_{-2}, \theta_2) \\
\vdots \\
p_K(y_K \mid y_{-K}, \theta_K)
\end{gathered}
$$

- θ_k: parameters of the kth conditional distribution
- $y_{-k} = (y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_K)$
- Key idea: use these models to sequentially impute, one variable at a time. Repeat this over a number of iterations

[2] https://www.jstatsoft.org/article/view/v045i03/v45i03.pdf

Multivariate Imputation by Chained Equations

Multivariate Imputation by Chained Equations (MICE) (van Buuren 2007, van Buuren and Groothuis-Oudshoorn 2011[2]) is an ad-hoc multiple imputation procedure that builds on the previous idea

- If each $Y_1, \ldots, Y_K$ is subject to missingness, we can posit K different regression models

$$\begin{gathered} p_1(y_1 \mid y_{-1}, \theta_1) \\ p_2(y_2 \mid y_{-2}, \theta_2) \\ \vdots \\ p_K(y_K \mid y_{-K}, \theta_K) \end{gathered}$$

- θ_k: parameters of the kth conditional distribution
- $y_{-k} = (y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_K)$
- Key idea: use these models to sequentially impute, one variable at a time. Repeat this over a number of iterations

[2] https://www.jstatsoft.org/article/view/v045i03/v45i03.pdf

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$

- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$\theta_1^{(t)} \sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i:r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1)$$

$$y_{i1}^{(t)} \sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1}$$

$$\vdots$$

$$\theta_K^{(t)} \sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i:r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K)$$

$$y_{iK}^{(t)} \sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}$$

- Iterate for a number of times

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$

- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$\theta_1^{(t)} \sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i:r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1)$$

$$y_{i1}^{(t)} \sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1}$$

$$\vdots$$

$$\theta_K^{(t)} \sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i:r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K)$$

$$y_{iK}^{(t)} \sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}$$

- Iterate for a number of times

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$

- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$\theta_1^{(t)} \sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i:r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1)$$

$$y_{i1}^{(t)} \sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1}$$

$$\vdots$$

$$\theta_K^{(t)} \sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i:r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K)$$

$$y_{iK}^{(t)} \sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}$$

- Iterate for a number of times

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$
- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$
\begin{aligned}
\theta_1^{(t)} &\sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i:r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1) \\
y_{i1}^{(t)} &\sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1} \\
&\vdots \\
\theta_K^{(t)} &\sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i:r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K) \\
y_{iK}^{(t)} &\sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}
\end{aligned}
$$

- Iterate for a number of times

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$
- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$\theta_1^{(t)} \sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i:r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1)$$

$$y_{i1}^{(t)} \sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1}$$

$$\vdots$$

$$\theta_K^{(t)} \sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i:r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K)$$

$$y_{iK}^{(t)} \sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}$$

- Iterate for a number of times

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$
- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$\theta_1^{(t)} \sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i:r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1)$$

$$y_{i1}^{(t)} \sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1}$$

$$\vdots$$

$$\theta_K^{(t)} \sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i:r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K)$$

$$y_{iK}^{(t)} \sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}$$

- Iterate for a number of times

Multivariate Imputation by Chained Equations

The MICE algorithm:

- Initialize the algorithm by randomly imputing the missing values of each variable/column by observed values of that variable/column. Denote this initial completed data as $\mathbf{y}_1^{(0)}, \ldots, \mathbf{y}_K^{(0)}$
- Run a pseudo Gibbs/Data Augmentation sampler, with tth iteration:

$$\theta_1^{(t)} \sim p_1(\theta_1 \mid \mathbf{y}_{1(\mathbf{r}_1)}, \mathbf{y}_2^{(t-1)}, \ldots, \mathbf{y}_K^{(t-1)}) \propto p_1(\theta_1) \prod_{i:r_{i1}=1} p_1(y_{i1} \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1)$$

$$y_{i1}^{(t)} \sim p_1(y_1 \mid y_{i2}^{(t-1)}, \ldots, y_{iK}^{(t-1)}, \theta_1^{(t)}), \text{ for all missing } y_{i1}$$

$$\vdots$$

$$\theta_K^{(t)} \sim p_K(\theta_K \mid \mathbf{y}_{K(\mathbf{r}_K)}, \mathbf{y}_1^{(t)}, \ldots, \mathbf{y}_{K-1}^{(t)}) \propto p_K(\theta_K) \prod_{i:r_{iK}=1} p_K(y_{iK} \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K)$$

$$y_{iK}^{(t)} \sim p_K(y_K \mid y_{i1}^{(t)}, \ldots, y_{i,K-1}^{(t)}, \theta_K^{(t)}), \text{ for all missing } y_{iK}$$

- Iterate for a number of times

Multivariate Imputation by Chained Equations

Comments:

- MICE is implemented in R, in the package `mice`
- Authors of `mice` suggest running the algorithm for 10 to 20 iterations
- `mice` package gives you m imputed datasets from m runs of the previous algorithm
- The idea is to use Rubin's combining rules with these m datasets

Caveats:

- Lack of theoretical study of this method, although incredibly popular!
- In general, the K conditional models will not be compatible, that is, there might not exist a joint distribution with such conditionals!

Multivariate Imputation by Chained Equations

Comments:

- MICE is implemented in R, in the package `mice`
- Authors of `mice` suggest running the algorithm for 10 to 20 iterations
- `mice` package gives you m imputed datasets from m runs of the previous algorithm
- The idea is to use Rubin's combining rules with these m datasets

Caveats:

- Lack of theoretical study of this method, although incredibly popular!
- In general, the K conditional models will not be compatible, that is, there might not exist a joint distribution with such conditionals!

Outline

Proper Multiple Imputation

Uncongeniality

Multivariate Imputation by Chained Equations

Summary

Summary

Main take-aways from today's lecture:

- Multiple Imputation:
  - Monte Carlo approximation of proper Bayesian procedure
  - Designed in the context of a statistical agency that needs to release complete datasets
  - Goal is to account for imputation uncertainty
  - Uncongeniality generally leads to invalidity of inferences based on Rubin's combining rules
  - MICE is a practical implementation of multiple imputation that builds on Gibbs sampling ideas, but lacks theoretical guarantees

Next lecture:

- R session 3

Summary

Main take-aways from today's lecture:

- Multiple Imputation:
  - Monte Carlo approximation of proper Bayesian procedure
  - Designed in the context of a statistical agency that needs to release complete datasets
  - Goal is to account for imputation uncertainty
  - Uncongeniality generally leads to invalidity of inferences based on Rubin's combining rules
  - MICE is a practical implementation of multiple imputation that builds on Gibbs sampling ideas, but lacks theoretical guarantees

Next lecture:

- R session 3