Statistical Methods for Analysis with Missing Data

Lecture 15: identifiability, nonignorability, pattern-mixture models

Mauricio Sadinle

Department of Biostatistics

UNIVERSITY of WASHINGTON

So Far

The approaches that we have covered for handling missing data:

- Ad-hoc approaches (imputation, complete cases)
- Frequentist likelihood-based inference
- Bayesian inference
- Multiple imputation
- Inverse-probability weighting

Something they have in common:

- We have assumed MAR (or MCAR), sometimes avoiding to handle the response mechanism $p(r \mid z)$

Today's Lecture

- What if we want to move away from MAR?
- We will talk about some fundamental issues for handling missing data
  - Identifiability
  - Nonignorability
- This discussion naturally leads to *pattern-mixture models*
- Reading: Chapter 6 of the lecture notes of Davidian and Tsiatis

Back to the Basics: Lecture 1

- Y: study variable
- R: response indicator

$$\underbrace{p(y)}_{\text{what we want}} = p(y \mid R=0)\underbrace{p(R=0) + \underbrace{p(y \mid R=1)p(R=1)}}_{\text{what we can get}}$$

We cannot recover $p(y \mid R = 0)$ nor $p(y)$ from observed data alone

The fundamental problem of inference with missing data: it is impossible without extra, usually untestable, assumptions on how missingness arises

Sample Data

- The *full-data sample* are independent and identically distributed (i.i.d.) draws from some distribution F

$$\{(Z_i, R_i)\}_{i=1}^{n} \overset{i.i.d.}{\sim} F$$

- R_i determines the part of Z_i that we get to observe: $Z_{i(R_i)}$

- We can think of the generative process, for each i:

$$Z_i \implies R_i \implies (Z_{i(R_i)}, R_i)$$

- In this lecture, we delete the subindex i to talk about
  - A generic draw from F
  - What we could recover provided an infinite sample size
  - Separate *identifiability* issues from *estimation* issues

Sample Data

- The *full-data sample* are independent and identically distributed (i.i.d.) draws from some distribution F

$$\{(Z_i, R_i)\}_{i=1}^{n} \overset{i.i.d.}{\sim} F$$

- R_i determines the part of Z_i that we get to observe: $Z_{i(R_i)}$
- We can think of the generative process, for each i:

$$Z_i \implies R_i \implies (Z_{i(R_i)}, R_i)$$

- In this lecture, we delete the subindex i to talk about
  - A generic draw from F
  - What we could recover provided an infinite sample size
  - Separate *identifiability* issues from *estimation* issues

Sample Data

- The *full-data sample* are independent and identically distributed (i.i.d.) draws from some distribution F

$$\{(Z_i, R_i)\}_{i=1}^{n} \overset{i.i.d.}{\sim} F$$

- R_i determines the part of Z_i that we get to observe: $Z_{i(R_i)}$
- We can think of the generative process, for each i:

$$Z_i \implies R_i \implies (Z_{i(R_i)}, R_i)$$

- In this lecture, we delete the subindex i to talk about
  - A generic draw from F
  - What we could recover provided an infinite sample size
  - Separate *identifiability* issues from *estimation* issues

Types of Data

- Full data: (Z, R)
- Observed data: $(Z_{(R)}, R)$
- Missing data: $Z_{(\bar{R})}$

Relationship:

$$(Z, R) = (Z_{(\bar{R})}, Z_{(R)}, R)$$

Distributions of Interest

- Full-data distribution: joint distribution of (Z, R) with density

$$p(z, r) \equiv p(z_{(\bar{r})}, z_{(r)}, r), \text{ for all } r$$

- Observed-data distribution: joint distribution of $(Z_{(R)}, R)$ with density

$$p(z_{(r)}, r) = \int p(z_{(\bar{r})}, z_{(r)}, r)\, dz_{(\bar{r})}, \text{ for all } r$$

- Missing-data distribution, or *extrapolation* distribution: conditional distribution of $Z_{(\bar{R})}$ given $(Z_{(R)}, R)$

$$p(z_{(\bar{r})} \mid z_{(r)}, r) = \frac{p(z_{(\bar{r})}, z_{(r)}, r)}{p(z_{(r)}, r)}, \text{ for all } r$$

Distributions of Interest

- Full-data distribution: joint distribution of (Z, R) with density

$$p(z, r) \equiv p(z_{(\bar{r})}, z_{(r)}, r), \text{ for all } r$$

- Observed-data distribution: joint distribution of $(Z_{(R)}, R)$ with density

$$p(z_{(r)}, r) = \int p(z_{(\bar{r})}, z_{(r)}, r) \, dz_{(\bar{r})}, \text{ for all } r$$

- Missing-data distribution, or *extrapolation* distribution: conditional distribution of $Z_{(\bar{R})}$ given $(Z_{(R)}, R)$

$$p(z_{(\bar{r})} \mid z_{(r)}, r) = \frac{p(z_{(\bar{r})}, z_{(r)}, r)}{p(z_{(r)}, r)}, \text{ for all } r$$

Distributions of Interest

- Full-data distribution: joint distribution of (Z, R) with density

$$p(z, r) \equiv p(z_{(\bar{r})}, z_{(r)}, r), \text{ for all } r$$

- Observed-data distribution: joint distribution of $(Z_{(R)}, R)$ with density

$$p(z_{(r)}, r) = \int p(z_{(\bar{r})}, z_{(r)}, r) \, dz_{(\bar{r})}, \text{ for all } r$$

- Missing-data distribution, or *extrapolation* distribution: conditional distribution of $Z_{(\bar{R})}$ given $(Z_{(R)}, R)$

$$p(z_{(\bar{r})} \mid z_{(r)}, r) = \frac{p(z_{(\bar{r})}, z_{(r)}, r)}{p(z_{(r)}, r)}, \text{ for all } r$$

The Full-Data Distribution

- Joint distribution of (Z, R) with density

$$p(z, r)$$

- Quantities of interest θ (parameters) depend on the full-data distribution

$$p(z,r) \quad \longrightarrow \quad p(z) = \sum_r p(z,r) \quad \longrightarrow \quad \theta = E[f(Z)] = \int f(z)p(z)dz$$

  - For example, say $f(Z) = Z_j$, then

$$\theta_j = E(Z_j) = \int z_j p(z_j) dz_j = \int z_j p(z) dz$$

The Full-Data Distribution

- Joint distribution of (Z, R) with density

$$p(z, r)$$

- Quantities of interest θ (parameters) depend on the full-data distribution

$$p(z, r) \longrightarrow p(z) = \sum_r p(z, r) \longrightarrow \theta = E[f(Z)] = \int f(z)p(z)dz$$

- For example, say $f(Z) = Z_j$, then

$$\theta_j = E(Z_j) = \int z_j p(z_j) dz_j = \int z_j p(z) dz$$

The Full-Data Distribution

- Joint distribution of (Z, R) with density

$$p(z, r)$$

- Quantities of interest θ (parameters) depend on the full-data distribution

$$p(z, r) \longrightarrow p(z) = \sum_r p(z, r) \longrightarrow \theta = E[f(Z)] = \int f(z)p(z)dz$$

  - For example, say $f(Z) = Z_j$, then

$$\theta_j = E(Z_j) = \int z_j p(z_j)dz_j = \int z_j p(z)dz$$

The Full-Data Distribution

- Joint distribution of (Z, R) with density

$$p(z, r)$$

- Quantities of interest θ (parameters) depend on the full-data distribution

$$p(z, r) \longrightarrow p(z) = \sum_r p(z, r) \longrightarrow \theta = E[f(Z)] = \int f(z)p(z)dz$$

  - For example, say $f(Z) = Z_j$, then

$$\theta_j = E(Z_j) = \int z_j p(z_j) dz_j = \int z_j p(z) dz$$

The Full-Data Distribution

- Joint distribution of (Z, R) with density

$$p(z, r)$$

- Quantities of interest θ (parameters) depend on the full-data distribution

$$p(z, r) \longrightarrow p(z) = \sum_r p(z, r) \longrightarrow \theta = E[f(Z)] = \int f(z)p(z)dz$$

  - For example, say $f(Z) = Z_j$, then

$$\theta_j = E(Z_j) = \int z_j p(z_j) dz_j = \int z_j p(z) dz$$

The Observed-Data Distribution

- Given $R = r$, we observe $Z_{(r)}$
- We can estimate $p(z_{(r)} \mid R = r)$ and $p(R = r)$ from observed data
- The observed-data distribution is all we can hope to recover from data alone

$$p(z_{(r)}, r) = p(z_{(r)} \mid r)p(r)$$

- For example, say you can sample indefinitely from the joint distribution of

$$Z = (Z_1, Z_2) \text{ and } R = (R_1, R_2)$$

  - If $R_j = 0$ you don't see the value of Z_j
- What we can estimate from such data:
  - $p(R = r), r \in \{0, 1\}^2$
  - $p(z_1 \mid R = 10)$
  - $p(z_2 \mid R = 01)$
  - $p(z_1, z_2 \mid R = 11)$

The Observed-Data Distribution

- Given $R = r$, we observe $Z_{(r)}$
- We can estimate $p(z_{(r)} \mid R = r)$ and $p(R = r)$ from observed data
- The observed-data distribution is all we can hope to recover from data alone

$$p(z_{(r)}, r) = p(z_{(r)} \mid r)p(r)$$

  - For example, say you can sample indefinitely from the joint distribution of

$$Z = (Z_1, Z_2) \text{ and } R = (R_1, R_2)$$

    - If $R_j = 0$ you don't see the value of Z_j
  - What we can estimate from such data:
    - $p(R = r), r \in \{0, 1\}^2$
    - $p(z_1 \mid R = 10)$
    - $p(z_2 \mid R = 01)$
    - $p(z_1, z_2 \mid R = 11)$

The Observed-Data Distribution

- Given $R = r$, we observe $Z_{(r)}$
- We can estimate $p(z_{(r)} \mid R = r)$ and $p(R = r)$ from observed data
- The observed-data distribution is all we can hope to recover from data alone

$$p(z_{(r)}, r) = p(z_{(r)} \mid r)p(r)$$

  - For example, say you can sample indefinitely from the joint distribution of

$$Z = (Z_1, Z_2) \text{ and } R = (R_1, R_2)$$

    - If $R_j = 0$ you don't see the value of Z_j
  - What we can estimate from such data:
    - $p(R = r), r \in \{0, 1\}^2$
    - $p(z_1 \mid R = 10)$
    - $p(z_2 \mid R = 01)$
    - $p(z_1, z_2 \mid R = 11)$

The Observed-Data Distribution

- Given $R = r$, we observe $Z_{(r)}$
- We can estimate $p(z_{(r)} \mid R = r)$ and $p(R = r)$ from observed data
- The observed-data distribution is all we can hope to recover from data alone

$$p(z_{(r)}, r) = p(z_{(r)} \mid r)p(r)$$

  - For example, say you can sample indefinitely from the joint distribution of

$$Z = (Z_1, Z_2) \text{ and } R = (R_1, R_2)$$

    - If $R_j = 0$ you don't see the value of Z_j
  - What we can estimate from such data:
    - $p(R = r), r \in \{0, 1\}^2$
    - $p(z_1 \mid R = 10)$
    - $p(z_2 \mid R = 01)$
    - $p(z_1, z_2 \mid R = 11)$

The Observed-Data Distribution

- Given $R = r$, we observe $Z_{(r)}$
- We can estimate $p(z_{(r)} \mid R = r)$ and $p(R = r)$ from observed data
- The observed-data distribution is all we can hope to recover from data alone

$$p(z_{(r)}, r) = p(z_{(r)} \mid r)p(r)$$

  - For example, say you can sample indefinitely from the joint distribution of

$$Z = (Z_1, Z_2) \text{ and } R = (R_1, R_2)$$

    - If $R_j = 0$ you don't see the value of Z_j
  - What we can estimate from such data:
    - $p(R = r), r \in \{0, 1\}^2$
    - $p(z_1 \mid R = 10)$
    - $p(z_2 \mid R = 01)$
    - $p(z_1, z_2 \mid R = 11)$

The Observed-Data Distribution

- Given $R = r$, we observe $Z_{(r)}$
- We can estimate $p(z_{(r)} \mid R = r)$ and $p(R = r)$ from observed data
- The observed-data distribution is all we can hope to recover from data alone

$$p(z_{(r)}, r) = p(z_{(r)} \mid r)p(r)$$

  - For example, say you can sample indefinitely from the joint distribution of

$$Z = (Z_1, Z_2) \text{ and } R = (R_1, R_2)$$

    - If $R_j = 0$ you don't see the value of Z_j
  - What we can estimate from such data:
    - $p(R = r), r \in \{0, 1\}^2$
    - $p(z_1 \mid R = 10)$
    - $p(z_2 \mid R = 01)$
    - $p(z_1, z_2 \mid R = 11)$

The Extrapolation Distribution

- Given $R = r$, we observe $Z_{(r)}$, but we don't observe $Z_{(\bar{r})}$
- There is no way of estimating $p(z_{(\bar{r})} \mid z_{(r)}, r)$ without assumptions

$$\underbrace{p(z_{(\bar{r})}, z_{(r)}, r)}_{\text{what we want}} = \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{how to extrapolate}} \underbrace{p(z_{(r)}, r)}_{\text{what we can get}}$$

- We say that $p(z_{(\bar{r})} \mid z_{(r)}, r)$, and therefore $p(z, r)$, are not *identifiable*
- *Identifying assumptions* explicitly or implicitly amount to constructing $p(z_{(\bar{r})} \mid z_{(r)}, r)$ from $p(z_{(r)}, r)$

The Extrapolation Distribution

- Given $R = r$, we observe $Z_{(r)}$, but we don't observe $Z_{(\bar{r})}$
- There is no way of estimating $p(z_{(\bar{r})} \mid z_{(r)}, r)$ without assumptions

$$\underbrace{p(z_{(\bar{r})}, z_{(r)}, r)}_{\text{what we want}} = \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{how to extrapolate}} \underbrace{p(z_{(r)}, r)}_{\text{what we can get}}$$

- We say that $p(z_{(\bar{r})} \mid z_{(r)}, r)$, and therefore $p(z, r)$, are not *identifiable*
- *Identifying assumptions* explicitly or implicitly amount to constructing $p(z_{(\bar{r})} \mid z_{(r)}, r)$ from $p(z_{(r)}, r)$

The Extrapolation Distribution

- Given $R = r$, we observe $Z_{(r)}$, but we don't observe $Z_{(\bar{r})}$
- There is no way of estimating $p(z_{(\bar{r})} \mid z_{(r)}, r)$ without assumptions

$$\underbrace{p(z_{(\bar{r})}, z_{(r)}, r)}_{\text{what we want}} = \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{how to extrapolate}} \underbrace{p(z_{(r)}, r)}_{\text{what we can get}}$$

- We say that $p(z_{(\bar{r})} \mid z_{(r)}, r)$, and therefore $p(z, r)$, are not *identifiable*
- *Identifying assumptions* explicitly or implicitly amount to constructing $p(z_{(\bar{r})} \mid z_{(r)}, r)$ from $p(z_{(r)}, r)$

The Extrapolation Distribution

- Given $R = r$, we observe $Z_{(r)}$, but we don't observe $Z_{(\bar{r})}$
- There is no way of estimating $p(z_{(\bar{r})} \mid z_{(r)}, r)$ without assumptions

$$\underbrace{p(z_{(\bar{r})}, z_{(r)}, r)}_{\text{what we want}} = \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{how to extrapolate}} \underbrace{p(z_{(r)}, r)}_{\text{what we can get}}$$

- We say that $p(z_{(\bar{r})} \mid z_{(r)}, r)$, and therefore $p(z, r)$, are not *identifiable*
- *Identifying assumptions* explicitly or implicitly amount to constructing $p(z_{(\bar{r})} \mid z_{(r)}, r)$ from $p(z_{(r)}, r)$

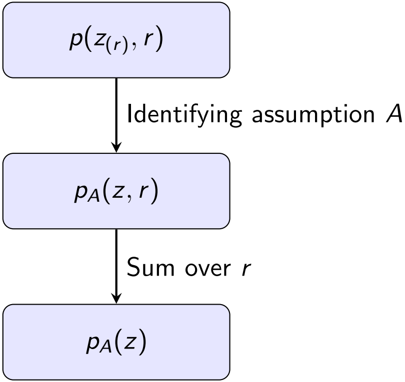
The Extrapolation Distribution

- Given $R = r$, we observe $Z_{(r)}$, but we don't observe $Z_{(\bar{r})}$
- There is no way of estimating $p(z_{(\bar{r})} \mid z_{(r)}, r)$ without assumptions

$$\underbrace{p(z_{(\bar{r})}, z_{(r)}, r)}_{\text{what we want}} = \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{how to extrapolate}} \underbrace{p(z_{(r)}, r)}_{\text{what we can get}}$$

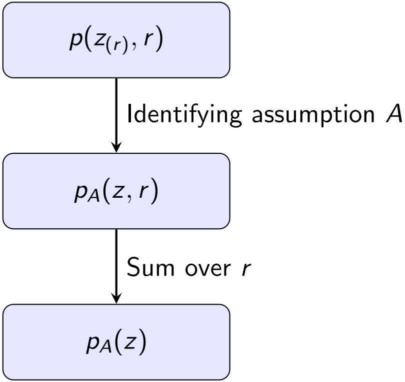
- We say that $p(z_{(\bar{r})} \mid z_{(r)}, r)$, and therefore $p(z, r)$, are not *identifiable*
- *Identifying assumptions* explicitly or implicitly amount to constructing $p(z_{(\bar{r})} \mid z_{(r)}, r)$ from $p(z_{(r)}, r)$

General Identification Strategy



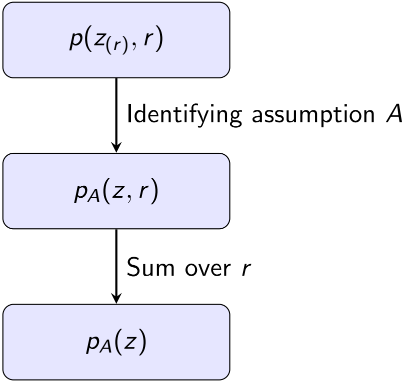
- Note that MAR (ignorability) gives you a shortcut to go from $p(z_{(r)}, r)$ to $p_{MAR}(z)$
- Otherwise, how do people specify identifying assumptions?

General Identification Strategy



- Note that MAR (ignorability) gives you a shortcut to go from $p(z_{(r)}, r)$ to $p_{MAR}(z)$
- Otherwise, how do people specify identifying assumptions?

General Identification Strategy



- Note that MAR (ignorability) gives you a shortcut to go from $p(z_{(r)}, r)$ to $p_{MAR}(z)$
- Otherwise, how do people specify identifying assumptions?

Factorizations of the Full-Data Distribution

Selection model factorization:

$$p(z, r) = p(r \mid z)p(z)$$

- The response mechanism $p(r \mid z)$ represents the way in which values of study variables get *selected* into the sample
- Natural factorization when we initially had a model $\{p(z \mid \theta)\}_\theta$ in mind, say had we not had missing data
- Allows us to continue using model $\{p(z \mid \theta)\}_\theta$
- Identifying assumptions are expressed as restriction on response mechanism $p(r \mid z)$
- We have focused on this approach so far under MAR:

$$p(r \mid z) = p(r \mid z_{(r)})$$

Factorizations of the Full-Data Distribution

Selection model factorization:

$$p(z, r) = p(r \mid z)p(z)$$

- The response mechanism $p(r \mid z)$ represents the way in which values of study variables get *selected* into the sample
- Natural factorization when we initially had a model $\{p(z \mid \theta)\}_\theta$ in mind, say had we not had missing data
- Allows us to continue using model $\{p(z \mid \theta)\}_\theta$
- Identifying assumptions are expressed as restriction on response mechanism $p(r \mid z)$
- We have focused on this approach so far under MAR:

$$p(r \mid z) = p(r \mid z_{(r)})$$

Factorizations of the Full-Data Distribution

Selection model factorization:

$$p(z, r) = p(r \mid z)p(z)$$

- The response mechanism $p(r \mid z)$ represents the way in which values of study variables get *selected* into the sample
- Natural factorization when we initially had a model $\{p(z \mid \theta)\}_\theta$ in mind, say had we not had missing data
- Allows us to continue using model $\{p(z \mid \theta)\}_\theta$
- Identifying assumptions are expressed as restriction on response mechanism $p(r \mid z)$
- We have focused on this approach so far under MAR:

$$p(r \mid z) = p(r \mid z_{(r)})$$

Factorizations of the Full-Data Distribution

Selection model factorization:

$$p(z, r) = p(r \mid z)p(z)$$

- The response mechanism $p(r \mid z)$ represents the way in which values of study variables get *selected* into the sample
- Natural factorization when we initially had a model $\{p(z \mid \theta)\}_\theta$ in mind, say had we not had missing data
- Allows us to continue using model $\{p(z \mid \theta)\}_\theta$
- Identifying assumptions are expressed as restriction on response mechanism $p(r \mid z)$
- We have focused on this approach so far under MAR:

$$p(r \mid z) = p(r \mid z_{(r)})$$

Factorizations of the Full-Data Distribution

Selection model factorization:

$$p(z, r) = p(r \mid z)p(z)$$

- The response mechanism $p(r \mid z)$ represents the way in which values of study variables get *selected* into the sample
- Natural factorization when we initially had a model $\{p(z \mid \theta)\}_\theta$ in mind, say had we not had missing data
- Allows us to continue using model $\{p(z \mid \theta)\}_\theta$
- Identifying assumptions are expressed as restriction on response mechanism $p(r \mid z)$
- We have focused on this approach so far under MAR:

$$p(r \mid z) = p(r \mid z_{(r)})$$

Factorizations of the Full-Data Distribution

Pattern-mixture model factorization:

$$p(z, r) = p(z \mid r)p(r)$$

- Requires models for distribution of Z given each value $R = r$
- Distribution of study variables is obtained as a *mixture* of *pattern*-specific models

$$p(z) = \sum_r p(z \mid r)p(r)$$

- This gives an alternative approach for handling missing data

Factorizations of the Full-Data Distribution

Pattern-mixture model factorization:

$$p(z, r) = p(z \mid r)p(r)$$

- Requires models for distribution of Z given each value $R = r$
- Distribution of study variables is obtained as a *mixture* of *pattern*-specific models

$$p(z) = \sum_{r} p(z \mid r)p(r)$$

- This gives an alternative approach for handling missing data

Factorizations of the Full-Data Distribution

Pattern-mixture model factorization:

$$p(z, r) = p(z \mid r)p(r)$$

- Requires models for distribution of Z given each value $R = r$
- Distribution of study variables is obtained as a *mixture* of *pattern*-specific models

$$p(z) = \sum_r p(z \mid r)p(r)$$

- This gives an alternative approach for handling missing data

Pattern-Mixture Models

- The pattern-mixture model factorization explicitly reveals:

$$p(z) = \sum_r p(z \mid r)p(r)$$

$$= \sum_r p(z_{(\bar{r})} \mid z_{(r)}, r)p(z_{(r)} \mid r)p(r)$$

$$= \sum_r \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{needs identifying assumption}} \underbrace{p(z_{(r)} \mid r)p(r)}_{\text{can be estimated from data}}$$

- Explicitly shows what needs identifying assumptions and what can be obtained from data alone

Pattern-Mixture Models

- The pattern-mixture model factorization explicitly reveals:

$$
\begin{aligned}
p(z) &= \sum_r p(z \mid r) p(r) \\
&= \sum_r p(z_{(\bar{r})} \mid z_{(r)}, r) p(z_{(r)} \mid r) p(r) \\
&= \sum_r \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{needs identifying assumption}} \underbrace{p(z_{(r)} \mid r) p(r)}_{\text{can be estimated from data}}
\end{aligned}
$$

- Explicitly shows what needs identifying assumptions and what can be obtained from data alone

Pattern-Mixture Models

- The pattern-mixture model factorization explicitly reveals:

$$
\begin{aligned}
p(z) &= \sum_r p(z \mid r)p(r) \\
&= \sum_r p(z_{(\bar{r})} \mid z_{(r)}, r)p(z_{(r)} \mid r)p(r) \\
&= \sum_r \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{needs identifying assumption}} \quad \underbrace{p(z_{(r)} \mid r)p(r)}_{\text{can be estimated from data}}
\end{aligned}
$$

- Explicitly shows what needs identifying assumptions and what can be obtained from data alone

Pattern-Mixture Models

- The pattern-mixture model factorization explicitly reveals:

$$
\begin{aligned}
p(z) &= \sum_r p(z \mid r)p(r) \\
&= \sum_r p(z_{(\bar{r})} \mid z_{(r)}, r)p(z_{(r)} \mid r)p(r) \\
&= \sum_r \overbrace{p(z_{(\bar{r})} \mid z_{(r)}, r)}^{\text{needs identifying assumption}} \underbrace{p(z_{(r)} \mid r)p(r)}_{\text{can be estimated from data}}
\end{aligned}
$$

- Explicitly shows what needs identifying assumptions and what can be obtained from data alone

Identifying Assumptions for Pattern-Mixture Models

- Identifying assumptions in the framework of pattern mixture models amount to specifying how to construct

$$\{p(z_{(\bar{r})} \mid z_{(r)}, r)\}_r$$

from

$$\{p(z_{(r)}, r)\}_r$$

- Once $p_A(z_{(\bar{r})} \mid z_{(r)}, r)$ is specified, according to an assumption A, this defines a full-data density

$$p_A(z_{(\bar{r})}, z_{(r)}, r) = p_A(z_{(\bar{r})} \mid z_{(r)}, r)p(z_{(r)}, r)$$

- Note that this in turn implies a response mechanism

$$p_A(r \mid z_{(\bar{r})}, z_{(r)}) = \frac{p_A(z_{(\bar{r})}, z_{(r)}, r)}{\sum_{r'} p_A(z_{(\bar{r}')}, z_{(r')}, r')}$$

- Assumptions that lead to response mechanisms that are not particular cases of MAR are *nonignorable*

Identifying Assumptions for Pattern-Mixture Models

- Identifying assumptions in the framework of pattern mixture models amount to specifying how to construct

$$\{p(z_{(\bar{r})} \mid z_{(r)}, r)\}_r$$

from

$$\{p(z_{(r)}, r)\}_r$$

- Once $p_A(z_{(\bar{r})} \mid z_{(r)}, r)$ is specified, according to an assumption A, this defines a full-data density

$$p_A(z_{(\bar{r})}, z_{(r)}, r) = p_A(z_{(\bar{r})} \mid z_{(r)}, r)p(z_{(r)}, r)$$

- Note that this in turn implies a response mechanism

$$p_A(r \mid z_{(\bar{r})}, z_{(r)}) = \frac{p_A(z_{(\bar{r})}, z_{(r)}, r)}{\sum_{r'} p_A(z_{(\bar{r}')}, z_{(r')}, r')}$$

- Assumptions that lead to response mechanisms that are not particular cases of MAR are *nonignorable*

Identifying Assumptions for Pattern-Mixture Models

- Identifying assumptions in the framework of pattern mixture models amount to specifying how to construct

$$\{p(z_{(\bar{r})} \mid z_{(r)}, r)\}_r$$

from

$$\{p(z_{(r)}, r)\}_r$$

- Once $p_A(z_{(\bar{r})} \mid z_{(r)}, r)$ is specified, according to an assumption A, this defines a full-data density

$$p_A(z_{(\bar{r})}, z_{(r)}, r) = p_A(z_{(\bar{r})} \mid z_{(r)}, r)p(z_{(r)}, r)$$

- Note that this in turn implies a response mechanism

$$p_A(r \mid z_{(\bar{r})}, z_{(r)}) = \frac{p_A(z_{(\bar{r})}, z_{(r)}, r)}{\sum_{r'} p_A(z_{(\bar{r}')}, z_{(r')}, r')}$$

- Assumptions that lead to response mechanisms that are not particular cases of MAR are *nonignorable*

Identifying Assumptions for Pattern-Mixture Models

- Identifying assumptions in the framework of pattern mixture models amount to specifying how to construct

$$\{p(z_{(\bar{r})} \mid z_{(r)}, r)\}_r$$

from

$$\{p(z_{(r)}, r)\}_r$$

- Once $p_A(z_{(\bar{r})} \mid z_{(r)}, r)$ is specified, according to an assumption A, this defines a full-data density

$$p_A(z_{(\bar{r})}, z_{(r)}, r) = p_A(z_{(\bar{r})} \mid z_{(r)}, r) p(z_{(r)}, r)$$

- Note that this in turn implies a response mechanism

$$p_A(r \mid z_{(\bar{r})}, z_{(r)}) = \frac{p_A(z_{(\bar{r})}, z_{(r)}, r)}{\sum_{r'} p_A(z_{(\bar{r}')}, z_{(r')}, r')}$$

- Assumptions that lead to response mechanisms that are not particular cases of MAR are *nonignorable*

Comments on Pattern-Mixture Models

Advantages:

- Identifiability requirements are more explicit than with selection models: easier to understand what is it that you are assuming
- Provides a natural framework for sensitivity analyses

Limitations:

- We cannot continue using model $\{p(z \mid \theta)\}_\theta$
- Parameters of scientific interest do not explicitly appear in the model
- Requires per-pattern model, say $\{p(z_{(r)} \mid r, \theta_r)\}_{\theta_r}$
- For general pattern of nonresponse we would need $2^K - 1$ models, one for each pattern in $\{0, 1\}^K$ (minus $\mathbf{0}_K$)
- Most developments under this approach assume monotone nonresponse (e.g., dropout)

Comments on Pattern-Mixture Models

Advantages:

- Identifiability requirements are more explicit than with selection models: easier to understand what is it that you are assuming
- Provides a natural framework for sensitivity analyses

Limitations:

- We cannot continue using model $\{p(z \mid \theta)\}_\theta$
- Parameters of scientific interest do not explicitly appear in the model
- Requires per-pattern model, say $\{p(z_{(r)} \mid r, \theta_r)\}_{\theta_r}$
- For general pattern of nonresponse we would need $2^K - 1$ models, one for each pattern in $\{0,1\}^K$ (minus $\mathbf{0}_K$)
- Most developments under this approach assume monotone nonresponse (e.g., dropout)

Comments on Pattern-Mixture Models

Advantages:

- Identifiability requirements are more explicit than with selection models: easier to understand what is it that you are assuming
- Provides a natural framework for sensitivity analyses

Limitations:

- We cannot continue using model $\{p(z \mid \theta)\}_\theta$
- Parameters of scientific interest do not explicitly appear in the model
- Requires per-pattern model, say $\{p(z_{(r)} \mid r, \theta_r)\}_{\theta_r}$
- For general pattern of nonresponse we would need $2^K - 1$ models, one for each pattern in $\{0,1\}^K$ (minus $\mathbf{0}_K$)
- Most developments under this approach assume monotone nonresponse (e.g., dropout)

Comments on Pattern-Mixture Models

Advantages:

- Identifiability requirements are more explicit than with selection models: easier to understand what is it that you are assuming
- Provides a natural framework for sensitivity analyses

Limitations:

- We cannot continue using model $\{p(z \mid \theta)\}_\theta$
- Parameters of scientific interest do not explicitly appear in the model
- Requires per-pattern model, say $\{p(z_{(r)} \mid r, \theta_r)\}_{\theta_r}$
- For general pattern of nonresponse we would need $2^K - 1$ models, one for each pattern in $\{0, 1\}^K$ (minus $\mathbf{0}_K$)
- Most developments under this approach assume monotone nonresponse (e.g., dropout)

Comments on Pattern-Mixture Models

Advantages:

- Identifiability requirements are more explicit than with selection models: easier to understand what is it that you are assuming
- Provides a natural framework for sensitivity analyses

Limitations:

- We cannot continue using model $\{p(z \mid \theta)\}_\theta$
- Parameters of scientific interest do not explicitly appear in the model
- Requires per-pattern model, say $\{p(z_{(r)} \mid r, \theta_r)\}_{\theta_r}$
- For general pattern of nonresponse we would need $2^K - 1$ models, one for each pattern in $\{0,1\}^K$ (minus $\mathbf{0}_K$)
- Most developments under this approach assume monotone nonresponse (e.g., dropout)

Comments on Pattern-Mixture Models

Advantages:

- Identifiability requirements are more explicit than with selection models: easier to understand what is it that you are assuming
- Provides a natural framework for sensitivity analyses

Limitations:

- We cannot continue using model $\{p(z \mid \theta)\}_\theta$
- Parameters of scientific interest do not explicitly appear in the model
- Requires per-pattern model, say $\{p(z_{(r)} \mid r, \theta_r)\}_{\theta_r}$
- For general pattern of nonresponse we would need $2^K - 1$ models, one for each pattern in $\{0,1\}^K$ (minus $\mathbf{0}_K$)
- Most developments under this approach assume monotone nonresponse (e.g., dropout)

Comments on Pattern-Mixture Models

Advantages:

- Identifiability requirements are more explicit than with selection models: easier to understand what is it that you are assuming
- Provides a natural framework for sensitivity analyses

Limitations:

- We cannot continue using model $\{p(z \mid \theta)\}_\theta$
- Parameters of scientific interest do not explicitly appear in the model
- Requires per-pattern model, say $\{p(z_{(r)} \mid r, \theta_r)\}_{\theta_r}$
- For general pattern of nonresponse we would need $2^K - 1$ models, one for each pattern in $\{0,1\}^K$ (minus $\mathbf{0}_K$)
- Most developments under this approach assume monotone nonresponse (e.g., dropout)

Dropout in Longitudinal Study

If missingness comes only from subjects dropping out

- Missingness patterns are uniquely summarized by the dropout time

$$D = 1 + \sum_{j=1}^{T} R_j$$

- The *observed data* are obtained as realizations of

$$(Z_{(D)}, D)$$

where, if $D = d$, $Z_{(d)} = (Z_1, \ldots, Z_{d-1})$ and $Z_{(\bar{d})} = (Z_d, \ldots, Z_T)$

- Pattern-mixture model requires modeling
  - $p(D = d)$: simply take empirical frequency
  - $p(z_{(d)} \mid D = d)$: depends on variable type

Dropout in Longitudinal Study

If missingness comes only from subjects dropping out

- Missingness patterns are uniquely summarized by the dropout time

$$D = 1 + \sum_{j=1}^{T} R_j$$

- The *observed data* are obtained as realizations of

$$(Z_{(D)}, D)$$

where, if $D = d$, $Z_{(d)} = (Z_1, \ldots, Z_{d-1})$ and $Z_{(\bar{d})} = (Z_d, \ldots, Z_T)$

- Pattern-mixture model requires modeling
  - $p(D = d)$: simply take empirical frequency
  - $p(z_{(d)} \mid D = d)$: depends on variable type

Dropout in Longitudinal Study

If missingness comes only from subjects dropping out

- Missingness patterns are uniquely summarized by the dropout time

$$D = 1 + \sum_{j=1}^{T} R_j$$

- The *observed data* are obtained as realizations of

$$(Z_{(D)}, D)$$

  where, if $D = d$, $Z_{(d)} = (Z_1, \ldots, Z_{d-1})$ and $Z_{(\bar{d})} = (Z_d, \ldots, Z_T)$

- Pattern-mixture model requires modeling
  - $p(D = d)$: simply take empirical frequency
  - $p(z_{(d)} \mid D = d)$: depends on variable type

Dropout in Longitudinal Study

If missingness comes only from subjects dropping out

- Missingness patterns are uniquely summarized by the dropout time

$$D = 1 + \sum_{j=1}^{T} R_j$$

- The *observed data* are obtained as realizations of

$$(Z_{(D)}, D)$$

where, if $D = d$, $Z_{(d)} = (Z_1, \dots, Z_{d-1})$ and $Z_{(\bar{d})} = (Z_d, \dots, Z_T)$

- Pattern-mixture model requires modeling
  - $p(D = d)$: simply take empirical frequency
  - $p(z_{(d)} \mid D = d)$: depends on variable type

Dropout in Longitudinal Study

If missingness comes only from subjects dropping out

- Missingness patterns are uniquely summarized by the dropout time

$$D = 1 + \sum_{j=1}^{T} R_j$$

- The *observed data* are obtained as realizations of

$$(Z_{(D)}, D)$$

where, if $D = d$, $Z_{(d)} = (Z_1, \ldots, Z_{d-1})$ and $Z_{(\bar{d})} = (Z_d, \ldots, Z_T)$

- Pattern-mixture model requires modeling
  - $p(D = d)$: simply take empirical frequency
  - $p(z_{(d)} \mid D = d)$: depends on variable type

A Simple Pattern-Mixture Model Under Dropout

In some situations, the following simple implementation of pattern-mixture models (PMMs) might be reasonable

- Idea: for each dropout group, model observed data and extrapolate to missing data
- Example:
  - For each d, fit
    $$E(Y_j \mid D = d) = \beta_{0d} + \beta_{1d} t_j,$$
    using data from $j < d$, and predict for $j \geq d$
  - This implies
    $$E(Y_j) = E[E(Y_j \mid D = d)] = \sum_d p(D = d)\beta_{0d} + t_j \sum_d p(D = d)\beta_{1d}$$
  - All parameters $p(d)$, β_{0d}, β_{1d}, $d = 1, \ldots, T$, can be directly estimated from the observed data (provided dropout starts at time $D = 3$)
- Note that this approach imposes parametric assumptions on the evolution of means over time, and assumes that this trend can be extrapolated

A Simple Pattern-Mixture Model Under Dropout

In some situations, the following simple implementation of pattern-mixture models (PMMs) might be reasonable

- Idea: for each dropout group, model observed data and extrapolate to missing data
- Example:
  - For each d, fit
    $$E(Y_j \mid D = d) = \beta_{0d} + \beta_{1d} t_j,$$
    using data from $j < d$, and predict for $j \geq d$
  - This implies
    $$E(Y_j) = E[E(Y_j \mid D = d)] = \sum_d p(D = d)\beta_{0d} + t_j \sum_d p(D = d)\beta_{1d}$$
  - All parameters $p(d)$, β_{0d}, β_{1d}, $d = 1, \ldots, T$, can be directly estimated from the observed data (provided dropout starts at time $D = 3$)
- Note that this approach imposes parametric assumptions on the evolution of means over time, and assumes that this trend can be extrapolated

A Simple Pattern-Mixture Model Under Dropout

In some situations, the following simple implementation of pattern-mixture models (PMMs) might be reasonable

- Idea: for each dropout group, model observed data and extrapolate to missing data
- Example:
  - For each d, fit
    $$E(Y_j \mid D = d) = \beta_{0d} + \beta_{1d} t_j,$$
    using data from $j < d$, and predict for $j \geq d$
  - This implies
    $$E(Y_j) = E[E(Y_j \mid D = d)] = \sum_d p(D = d)\beta_{0d} + t_j \sum_d p(D = d)\beta_{1d}$$
  - All parameters $p(d)$, β_{0d}, β_{1d}, $d = 1, \ldots, T$, can be directly estimated from the observed data (provided dropout starts at time $D = 3$)
- Note that this approach imposes parametric assumptions on the evolution of means over time, and assumes that this trend can be extrapolated

A Simple Pattern-Mixture Model Under Dropout

In some situations, the following simple implementation of pattern-mixture models (PMMs) might be reasonable

- Idea: for each dropout group, model observed data and extrapolate to missing data
- Example:
  - For each d, fit
    $$E(Y_j \mid D = d) = \beta_{0d} + \beta_{1d} t_j,$$
    using data from $j < d$, and predict for $j \geq d$
  - This implies
    $$E(Y_j) = E[E(Y_j \mid D = d)] = \sum_d p(D = d)\beta_{0d} + t_j \sum_d p(D = d)\beta_{1d}$$
  - All parameters $p(d)$, β_{0d}, β_{1d}, $d = 1, \dots, T$, can be directly estimated from the observed data (provided dropout starts at time $D = 3$)
- Note that this approach imposes parametric assumptions on the evolution of means over time, and assumes that this trend can be extrapolated

A Simple Pattern-Mixture Model Under Dropout

In some situations, the following simple implementation of pattern-mixture models (PMMs) might be reasonable

- Idea: for each dropout group, model observed data and extrapolate to missing data
- Example:
  - For each d, fit
    $$E(Y_j \mid D = d) = \beta_{0d} + \beta_{1d} t_j,$$
    using data from $j < d$, and predict for $j \geq d$
  - This implies
    $$E(Y_j) = E[E(Y_j \mid D = d)] = \sum_d p(D = d)\beta_{0d} + t_j \sum_d p(D = d)\beta_{1d}$$
  - All parameters $p(d)$, β_{0d}, β_{1d}, $d = 1, \ldots, T$, can be directly estimated from the observed data (provided dropout starts at time $D = 3$)
- Note that this approach imposes parametric assumptions on the evolution of means over time, and assumes that this trend can be extrapolated

Identifying Assumptions for PMMs Under Dropout

- In general, how to obtain $p(z_{(\bar{d})} \mid z_{(d)}, d)$ from $p(z_{(d)}, d)$?
- Note that

$$p(z_{(\bar{d})} \mid z_{(d)}, d) = p(z_d, \ldots, z_T \mid z_1, \ldots, z_{d-1}, d)$$
$$= \prod_{\ell=d}^{T} p(z_\ell \mid z_1, \ldots, z_{d-1}, \ldots, z_{\ell-1}, d)$$
$$= \prod_{\ell=d}^{T} p(z_\ell \mid z_{(\ell)}, d)$$

- Thus, we need to think how to obtain $p(z_\ell \mid z_{(\ell)}, d)$ for each $\ell \geq d$, $d = 1, \ldots, T$

Identifying Assumptions for PMMs Under Dropout

- In general, how to obtain $p(z_{(\bar{d})} \mid z_{(d)}, d)$ from $p(z_{(d)}, d)$?

- Note that

$$
\begin{aligned}
p(z_{(\bar{d})} \mid z_{(d)}, d) &= p(z_d, \ldots, z_T \mid z_1, \ldots, z_{d-1}, d) \\
&= \prod_{\ell=d}^{T} p(z_\ell \mid z_1, \ldots, z_{d-1}, \ldots, z_{\ell-1}, d) \\
&= \prod_{\ell=d}^{T} p(z_\ell \mid z_{(\ell)}, d)
\end{aligned}
$$

- Thus, we need to think how to obtain $p(z_\ell \mid z_{(\ell)}, d)$ for each $\ell \geq d$, $d = 1, \ldots, T$

Identifying Assumptions for PMMs Under Dropout

- In general, how to obtain $p(z_{(\bar{d})} \mid z_{(d)}, d)$ from $p(z_{(d)}, d)$?
- Note that

$$p(z_{(\bar{d})} \mid z_{(d)}, d) = p(z_d, \ldots, z_T \mid z_1, \ldots, z_{d-1}, d)$$
$$= \prod_{\ell=d}^{T} p(z_\ell \mid z_1, \ldots, z_{d-1}, \ldots, z_{\ell-1}, d)$$
$$= \prod_{\ell=d}^{T} p(z_\ell \mid z_{(\ell)}, d)$$

- Thus, we need to think how to obtain $p(z_\ell \mid z_{(\ell)}, d)$ for each $\ell \geq d$, $d = 1, \ldots, T$

Identifying Assumptions for PMMs Under Dropout

- In general, how to obtain $p(z_{(\bar{d})} \mid z_{(d)}, d)$ from $p(z_{(d)}, d)$?
- Note that

$$
\begin{aligned}
p(z_{(\bar{d})} \mid z_{(d)}, d) &= p(z_d, \ldots, z_T \mid z_1, \ldots, z_{d-1}, d) \\
&= \prod_{\ell=d}^{T} p(z_\ell \mid z_1, \ldots, z_{d-1}, \ldots, z_{\ell-1}, d) \\
&= \prod_{\ell=d}^{T} p(z_\ell \mid z_{(\ell)}, d)
\end{aligned}
$$

- Thus, we need to think how to obtain $p(z_\ell \mid z_{(\ell)}, d)$ for each $\ell \geq d$, $d = 1, \ldots, T$

Identifying Assumptions for PMMs Under Dropout

- In general, how to obtain $p(z_{(\bar{d})} \mid z_{(d)}, d)$ from $p(z_{(d)}, d)$?

- Note that

$$\begin{aligned} p(z_{(\bar{d})} \mid z_{(d)}, d) &= p(z_d, \ldots, z_T \mid z_1, \ldots, z_{d-1}, d) \\ &= \prod_{\ell=d}^{T} p(z_\ell \mid z_1, \ldots, z_{d-1}, \ldots, z_{\ell-1}, d) \\ &= \prod_{\ell=d}^{T} p(z_\ell \mid z_{(\ell)}, d) \end{aligned}$$

- Thus, we need to think how to obtain $p(z_\ell \mid z_{(\ell)}, d)$ for each $\ell \geq d$, $d = 1, \ldots, T$

The Complete-Case Identifying Assumption

Little (JASA 1993) proposed to tie the extrapolation distributions to the distribution of complete cases:

$$p_{CC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D = T + 1),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- The distributions for $D = T + 1$ are identifiable from the complete cases
- This strategy could also be used with nonmonotone missingness
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.

The Complete-Case Identifying Assumption

Little (JASA 1993) proposed to tie the extrapolation distributions to the distribution of complete cases:

$$p_{CC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D = T + 1),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- The distributions for $D = T + 1$ are identifiable from the complete cases
- This strategy could also be used with nonmonotone missingness
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.

The Complete-Case Identifying Assumption

Little (JASA 1993) proposed to tie the extrapolation distributions to the distribution of complete cases:

$$p_{CC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D = T + 1),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- The distributions for $D = T + 1$ are identifiable from the complete cases
- This strategy could also be used with nonmonotone missingness
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.

The Neighboring-Case Identifying Assumption

The extrapolation distributions could also be obtained from the closest dropout pattern where ℓ is available:

$$p_{NC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D = \ell + 1),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- Among observations with $D = \ell + 1$ we get to observe z_ℓ and $z_{(\ell)}$
- We could think that observations with $D = \ell + 1$ are the best for basing extrapolation of the values of Z_ℓ
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.

The Neighboring-Case Identifying Assumption

The extrapolation distributions could also be obtained from the closest dropout pattern where ℓ is available:

$$p_{NC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D = \ell + 1),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- Among observations with $D = \ell + 1$ we get to observe z_ℓ and $z_{(\ell)}$
- We could think that observations with $D = \ell + 1$ are the best for basing extrapolation of the values of Z_ℓ
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.

The Neighboring-Case Identifying Assumption

The extrapolation distributions could also be obtained from the closest dropout pattern where ℓ is available:

$$p_{NC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D = \ell + 1),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- Among observations with $D = \ell + 1$ we get to observe z_ℓ and $z_{(\ell)}$
- We could think that observations with $D = \ell + 1$ are the best for basing extrapolation of the values of Z_ℓ
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.

The Available-Case Identifying Assumption

Here, the extrapolation distributions are obtained from all available cases where ℓ is available:

$$p_{AC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D > \ell),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- Among observations with $D > \ell$ we get to observe z_ℓ and $z_{(\ell)}$
- We could think that this approach maximizes the use of available information for basing extrapolation of the values of Z_ℓ
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.
- HW4: under monotone nonresponse, the AC assumption is equivalent to MAR

The Available-Case Identifying Assumption

Here, the extrapolation distributions are obtained from all available cases where ℓ is available:

$$p_{AC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D > \ell),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- Among observations with $D > \ell$ we get to observe z_ℓ and $z_{(\ell)}$
- We could think that this approach maximizes the use of available information for basing extrapolation of the values of Z_ℓ
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.
- HW4: under monotone nonresponse, the AC assumption is equivalent to MAR

The Available-Case Identifying Assumption

Here, the extrapolation distributions are obtained from all available cases where ℓ is available:

$$p_{AC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D > \ell),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- Among observations with $D > \ell$ we get to observe z_ℓ and $z_{(\ell)}$
- We could think that this approach maximizes the use of available information for basing extrapolation of the values of Z_ℓ
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.
- HW4: under monotone nonresponse, the AC assumption is equivalent to MAR

The Available-Case Identifying Assumption

Here, the extrapolation distributions are obtained from all available cases where ℓ is available:

$$p_{AC}(z_\ell \mid z_{(\ell)}, D = d) \equiv p(z_\ell \mid z_{(\ell)}, D > \ell),$$

for all $\ell \geq d$, $d = 1, \ldots, T$.

- Among observations with $D > \ell$ we get to observe z_ℓ and $z_{(\ell)}$
- We could think that this approach maximizes the use of available information for basing extrapolation of the values of Z_ℓ
- HW4: say $T = 3$, write down this restriction for $\ell \geq d$, $d = 1, 2, 3$.
- HW4: under monotone nonresponse, the AC assumption is equivalent to MAR

Observational Equivalence

- Two full-data distributions are said to be *observationally equivalent* if their implied observed-data distributions are the same

- This is, say I have two full-data distributions

$$p_A(z_{(\bar{r})}, z_{(r)}, r)$$

and

$$p_B(z_{(\bar{r})}, z_{(r)}, r)$$

if

$$\int p_A(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})} = \int p_B(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})}$$

for all $(z_{(r)}, r)$, then they are *observationally equivalent*

- HW4: the full-data distributions obtained under the CC, NC, and AC restrictions are observationally equivalent

- This is an important feature in *sensitivity analysis*! (next class)

Observational Equivalence

- Two full-data distributions are said to be *observationally equivalent* if their implied observed-data distributions are the same

- This is, say I have two full-data distributions

$$p_A(z_{(\bar{r})}, z_{(r)}, r)$$

and

$$p_B(z_{(\bar{r})}, z_{(r)}, r)$$

if

$$\int p_A(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})} = \int p_B(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})}$$

for all $(z_{(r)}, r)$, then they are *observationally equivalent*

- HW4: the full-data distributions obtained under the CC, NC, and AC restrictions are observationally equivalent

- This is an important feature in *sensitivity analysis*! (next class)

Observational Equivalence

- Two full-data distributions are said to be *observationally equivalent* if their implied observed-data distributions are the same
- This is, say I have two full-data distributions

$$p_A(z_{(\bar{r})}, z_{(r)}, r)$$

and

$$p_B(z_{(\bar{r})}, z_{(r)}, r)$$

if

$$\int p_A(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})} = \int p_B(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})}$$

for all $(z_{(r)}, r)$, then they are *observationally equivalent*

- HW4: the full-data distributions obtained under the CC, NC, and AC restrictions are observationally equivalent
- This is an important feature in *sensitivity analysis*! (next class)

Observational Equivalence

- Two full-data distributions are said to be *observationally equivalent* if their implied observed-data distributions are the same
- This is, say I have two full-data distributions

$$p_A(z_{(\bar{r})}, z_{(r)}, r)$$

and

$$p_B(z_{(\bar{r})}, z_{(r)}, r)$$

if

$$\int p_A(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})} = \int p_B(z_{(\bar{r})}, z_{(r)}, r)\ dz_{(\bar{r})}$$

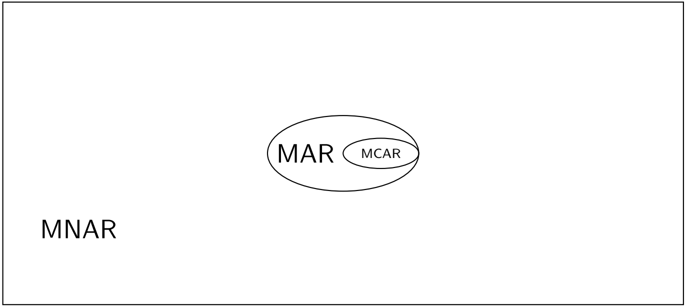
for all $(z_{(r)}, r)$, then they are *observationally equivalent*

- HW4: the full-data distributions obtained under the CC, NC, and AC restrictions are observationally equivalent
- This is an important feature in *sensitivity analysis*! (next class)

Summary

Main take-aways from today's lecture:

- The fundamental problem of inference with missing data: *it is impossible without extra, usually untestable, assumptions on how missingness arises*
- Pattern-mixture models provide an alternative way of thinking about missing data
- Remember the universe of missing-data assumptions:



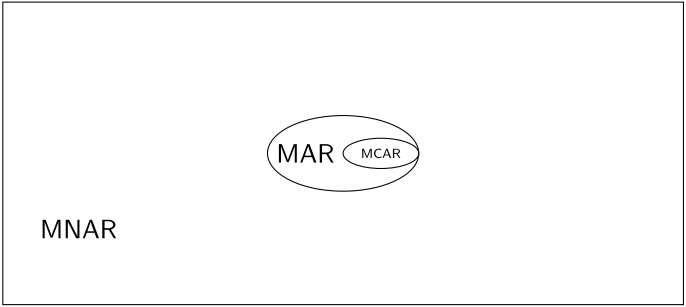
Next lecture:

- More on nonignorable missing data (MNAR), and sensitivity analysis

Summary

Main take-aways from today's lecture:

- The fundamental problem of inference with missing data: *it is impossible without extra, usually untestable, assumptions on how missingness arises*
- Pattern-mixture models provide an alternative way of thinking about missing data
- Remember the universe of missing-data assumptions:



Next lecture:

- More on nonignorable missing data (MNAR), and sensitivity analysis