Pattern Mixture Models

Model Classes

Likelihood Function:

$$L(Y_{obs}, R) = \int L(Y_{obs}, Y_{mis}, R) dY_{mis}$$

Two classes of models:

1. Selection model (Diggle and Kenward, 1994):

Partition the likelihood as

$$L(Y_{obs}, Y_{mis}, R) = L(Y_{obs}, Y_{mis}) L(R | Y_{obs}, Y_{mis})$$

2. Pattern Mixture Model (Little, 1993, 1994):

Partition the likelihood as

$$L(Y_{obs}, Y_{mis}, R) = L(R) L(Y_{obs}, Y_{mis} | R)$$

Pattern Mixture Model

$$L(\mathbf{Y}_{obs}, R|X) = \int L(R|\mathbf{X})L(\mathbf{Y}_{obs}, \mathbf{Y}_{mis}|\mathbf{X}, R)d\mathbf{Y}_{mis}$$

The joint distribution of $(\mathbf{Y}_i, R_i)$ is factored as:

$$[\mathbf{Y}_i, R_i|\mathbf{X}_i] = [R_i|\mathbf{X}_i][\mathbf{Y}_i|\mathbf{X}_i, R_i]$$

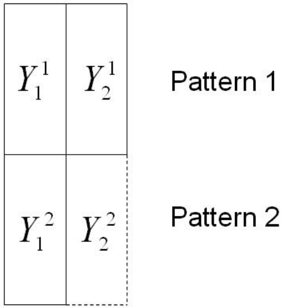
where

- $[\mathbf{Y}_i|\mathbf{X}_i, R_i]$ models the relationship between Y and X within each missing data pattern.
- $[R_i|\mathbf{X}_i]$ models the drop-out mechanism marginally as a function of covariates.

Pattern Mixture Model: Remarks

- Pattern-mixture models stratify the data by missing data patterns and then allow for different relationships between Y and X for different patterns.
- When drop-out is at random, this approach can avoid a full specification of the drop-out mechanism.
- Complicated when there are too many missing patterns.
- Note: Pattern-mixture model has no marginal mean interpretation

Identifiability in Pattern Mixture Models



$$
\begin{aligned}
\mathbf{Y}^1 = (\mathbf{Y}_1^1, \mathbf{Y}_2^1) &\sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}^{(1)}) \\
\mathbf{Y}^2 = (\mathbf{Y}_1^2, \mathbf{Y}_2^2) &\sim N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}^{(2)})
\end{aligned}
$$

Remarks:

- This model assumes the outcome has different means and covariances for different patterns.
- The parameters in Pattern 1, $\boldsymbol{\mu}^{(1)} = (\boldsymbol{\mu}_1^{(1)}, \boldsymbol{\mu}_2^{(2)})$and $\boldsymbol{\Sigma}^{(1)}$ both can be estimated directly from the data.

Identifiability in Pattern Mixture Models (2)

- The parameters in Pattern 2, $\mu_1^{(2)}$ and $\sigma_{11}^{(2)}$can be estimated from the data but $\mu_2^{(2)}$, $\sigma_{12}^{(2)}$, $\sigma_{22}^{(2)}$ are not estimable from the data.
- Assumptions are needed to make parameter identifiable, e.g., assume $\mathbf{\Sigma}^{(1)} = \mathbf{\Sigma}^{(2)}$and perform sensitivity analysis of $\mu_2^{(2)}$.
- Pattern mixture models allow one to see easily which parameters are not identifiable.
- $\mathbf{Y}$ is a mixture of two normal distributions with the marginal mean

  $$E(\mathbf{Y}) = \sum_{j=1}^{2} Pr(R = j + 1)\boldsymbol{\mu}^{(j)}.$$

  Compare to selection models, where the marginal distribution of $\mathbf{Y}$ is normal.
- Selection models and pattern mixture models are equivalent under MCAR.

Parameter Interpretation in Pattern Mixture Models

Regression Pattern Mixture Model

$$\begin{aligned}\boldsymbol{\mu}_{ij}^{(1)} &= \mathbf{X}_{ij}^T \beta^{(1)} \\ \boldsymbol{\mu}_{ij}^{(2)} &= \mathbf{X}_{ij}^T \beta^{(2)}\end{aligned}$$

- $\beta^{(1)}$is the regression coefficient vector if longitudinal data are complete (Pattern 1).
- $\beta^{(2)}$ is the regression coefficient vector among those subjects who drop out at time 2.
- Both $\beta^{(1)}$and $\beta^{(2)}$need to be interpreted conditional on dropout patterns and are less interesting in practice.

Parameter Interpretation in Pattern Mixture Models (2)

- Marginal covariate effect:

$$\hat{\beta} \;=\; \sum_{j=1}^{J} \pi_j \widehat{\beta^{(j)}},$$

where $\pi_j = Pr(R = j + 1)$ is the probability of each dropout pattern.

- Can use the Delta Method to obtain standard errors.
    - uncertainty in the estimates of $\beta' s$.
    - uncertainty in the sample proportion estimates $\pi' s$.

Parameter Interpretation in Pattern Mixture Models (3)

- The parameters for incomplete data are not estimable if no assumption is made.

- The paramters can be identified if assumptions are made across different dropout patterns, e.g., the slopes are the same across different dropout patterms for incomplete subjects.

- Explicit assumptions on the form of the dependence of missing mechanism (R) on Y are avoided in pattern mixture models. Note this is required in selection models.

Schizophrenia Study

- 312 patients received drug therapy for schizophrenia; 101 patients received a placebo

- Variables:
    - subject ID number
    - Outcome = IMPS79: overall severity of illness (continuous)
    - WEEK: 0,1,3,6
    - DRUG: 0=placebo 1=drug (chlorpromazine, fluphenazine, or thioridazine)
    - SEX: 0=female 1=male

Trial sample size:

		Time		
Group	0	1	3	6
Placebo ($n = 108$)	107	105	87	70
Drug ($n = 329$)	327	321	287	265

Note: The drug group combines three treatments.

- 102 of 437 subjects did not complete the trial by the end of the study.
- Assume the time trajectory (intercept and slope) is the same across different patterns for those subjects who dropped out from the study.

Pattern Mixture Model for Schizophrenia study:

$$
\begin{aligned}
IMPS79_{ij} \quad = \quad & \beta_0 + \beta_1 Drug_i + \beta_2 \sqrt{week}_i + \beta_3 Drug_i \times \sqrt{week}_i \\
& \beta_0^D \times Drop_i + \beta_1^D (Drug_i \times Drop_i) + \beta_2^D (\sqrt{week}_i \times Drop_i) \\
& + \beta_3^D (Drug_i \times \sqrt{week}_i \times Drop_i) \\
& + b_{i0} + b_{i1} \sqrt{week}_i + \epsilon_{ij}
\end{aligned}
$$

- $Drop_i$: indicator of whether the subject dropped out of study.
- β_0, β_1, β_2 and β_3 are the coefficients for the subgroup who completed the trial.
- β_0^D, β_1^D, β_2^D and β_3^D are the difference of coefficients for those dropped out of the trial compared to the completers.

Pattern-mixture averaged results:

- Final estimates are obtained by averaging over missing-data patterns

$$\hat{\hat{\boldsymbol{\beta}}} = \hat{\pi}_c \hat{\boldsymbol{\beta}}_c + \hat{\pi}_d \hat{\boldsymbol{\beta}}_d$$

  where π_c =prob of completing the trial and π_d= prob of dropping out.

- In this example, we use drug-specific sample proportions as estimates of missing-data pattern proportions
- Note when $Drug = 0$:
  - $drop = 0$:

$$E[IMPS79_{ij}] = \quad \beta_0 + \beta_2 \sqrt{week}_i$$

  - $drop = 1$:

$$E[IMPS79_{ij}] \quad = \quad \beta_0 + \beta_2 \sqrt{week}_i + \beta_0^D + \beta_2^D \sqrt{week}_i$$

Pattern-mixture averaged results (2)

- Therefore,

$$\begin{aligned}\hat{\hat{\beta}}_0 &= \hat{\pi}_{c0} \times \hat{\beta}_0 + \hat{\pi}_{d0} \times (\hat{\beta}_0 + \hat{\beta_0^D}) \\ &= \hat{\beta}_0 + \hat{\pi}_{d0} \times \hat{\beta_0^D}\end{aligned}$$

$$\begin{aligned}\hat{\hat{\beta}}_2 &= \hat{\pi}_{c0} \times \hat{\beta}_2 + \hat{\pi}_{d0} \times (\hat{\beta}_2 + \hat{\beta_2^D}) \\ &= \hat{\beta}_2 + \hat{\pi}_{d0} \times \hat{\beta_2^D}\end{aligned}$$

  where $\hat{\pi}_{c0}$ and $\hat{\pi}_{d0}$ are the completer proportion and the dropout proportion in the placebo group respectively.

- Similarly, we can get

$$\hat{\hat{\beta}}_1 = \hat{\beta}_1 + \hat{\pi}_{d1} \times \hat{\beta_1^D}$$

  and

$$\hat{\hat{\beta}}_3 = \hat{\beta}_3 + \hat{\pi}_{d1} \times \hat{\beta_3^D}$$

  where $\hat{\pi}_{d1}$ is the dropout proportion in the treatment group.

Analysis Results

NIMH Schizophrenia Study: Severity across Time
MML Estimates (se) *random intercept and slope models*

	Completers $N = 335$	*All cases* $N = 437$	*Pattern mixture* $N = 437$
intercept	5.221 (.109)	5.348 (.088)	5.334 (.089)
Drug (0=P; 1=D)	0.202 (.123)	0.046 (.101)	0.124 (.105)
Time (sqrt wk)	-0.393 (.073)	-0.336 (.068)	-0.305 (.071)
Drug by Time	-0.539 (.083)	-0.641 (.078)	-0.662 (.078)