Likelihood-Based Inference and Selection Models

Missing Data with Dropouts

$\boldsymbol{Y}_{obs}$: Measurements actually observed

$\boldsymbol{Y}_{mis}$: Measurements that should be available but are missing

$\mathbf{Y} = (\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis})$: Hypothetical complete data

R = Dropout time

Assume the covariates $\mathbf{X}$ are fully observed.

Data structure (Monotone missingness):

Pattern	Y_1	Y_2	Y_3	R
1	X	X	X	4
2	X	X		3
3	X			2

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

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

Inference Procedures in Selection Models

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

1. Likelihood-based inference: Specify the full likelihoods $L(Y_{obs}, Y_{mis}|X)$, e.g., using GLMMs, and $L(R|Y_{obs}, Y_{mis}, X)$, e.g.,logistic regression.can be used for model fitting.

2. Estimating equation-based Inference: Only specify the first (two) moments of $L(Y_{obs}, Y_{mis}|X)$, and specify $L(R|Y_{obs}, Y_{mis}, X)$, e.g.,logistic regression. Estimate model parameters using modified GEEs, e.g., inverse-probability weighted (IPW) GEEs; augmented inverse probability (AIPW) weighted GEEs.

Remarks

1. Under MAR or MCAR, likelihood-based inference is valid using only the observed data, i.e., usual software are valid using all the observed data.

2. Standard GEEs require MCAR to hold for regression coeffiicent estimators to be consistent.

3. IPW and AIPW GEEs are valid under MAR, and AIPW improves robustness and efficiency of the IPW estimators. (Easily done in SAS)

MAR and MCAR in Likelihood Inference: Ignorability

Recall under MAR,

$$\begin{aligned} f(\boldsymbol{Y}_{obs}, \mathbf{R}; \beta, \gamma) &= f(\mathbf{R}|\boldsymbol{Y}_{obs}; \gamma) \int f(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}; \beta) d\boldsymbol{Y}_{mis} \\ &= f(\mathbf{R}|\boldsymbol{Y}_{obs}; \gamma) f(\boldsymbol{Y}_{obs}; \beta) \end{aligned}$$

and

$$\ell(\beta, \gamma) = \ell(\mathbf{R}|\boldsymbol{Y}_{obs}; \gamma) + \ell(\boldsymbol{Y}_{obs}; \beta) \tag{1}$$

- If the parameters β and γ are distinct in $\ell(\mathbf{R}|\boldsymbol{Y}_{obs}; \gamma)$ and $f(\boldsymbol{Y}_{obs}; \beta)$, likelihood-based inference using only the observed data is valid under MAR and MCAR, i.e., missingness is **ignorable**.
- If $\ell(\mathbf{R}|\boldsymbol{Y}_{obs}; \gamma)$ and $\ell(\boldsymbol{Y}_{obs}; \beta)$ share some common parameters, ignoring the first term on the right hand side of (1) will lead to loss of efficiency but the MLE from maximizing $\ell(\boldsymbol{Y}_{obs}; \beta)$ is still consistent.

Aside: Checking for MAR

- Compare the outcomes(covariates) at time $t_k < t_j$ among those who stay in the study at time t_j with those who drop out at time t_j.

- Run a logistic regression

$$logit(R = j) = \alpha_0 + \alpha_1 Y_{j-1} + \alpha_3 X_{j-1}$$

test for $\alpha_1 = 0$ and $\alpha_2 = 0$.

Inference for Selection Models

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

- The selection model models the marginal distribution of the outcome $\mathbf{Y}|\mathbf{X}$ and models the conditional distribution of the dropout on the outcome $R|\mathbf{Y}, \mathbf{X}$. The regression coefficients from the $\mathbf{Y}|\mathbf{X}$ model hence have attractive population interpretation that is of practical interest.
- Likelihood-based inference: Specify the full likelihood for $L(\mathbf{Y}_{obs}, \mathbf{Y}_{mis}|\mathbf{X})$, e.g., using LMMs or GLMMs, the likelihood $L(R|\mathbf{Y}_{obs}, \mathbf{Y}_{mis}, \mathbf{X})$, e.g., using logistic regression.

Selection Models for Normal Outcome

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{X}_i\beta + \boldsymbol{\epsilon_i} \\ logit\{Pr(R_i = j | R_i \geq j)\} &= \alpha_{j0} + \alpha_1 Y_{i,j-1} + \alpha_2 Y_{i,j} \end{aligned}$$

where $j = 2, \cdots J$, $\mathbf{X}_i$=$n_i \times p$ covariate matrix, β=unknown $p \times 1$ parameter vector, and $\boldsymbol{\epsilon_i} \sim N(\mathbf{0}, \mathbf{V})$.

Remarks:

1. This logistic model corresponds to the logistic model used for discrete survival data, and models the chance a subject drops out at time j given he/she was still in the study at time $j-1$.
2. The dropout model allows for the possibility of missing not at random (NMAR) through α_2, and $\alpha_2 = 0$ corresponds to MAR.
3. The data provide no information about α_2. Hence α_2 is not identifiable from the data, and sensitivity analysis over a range of α_2 is recommended.

Selection Models for Normal Outcome (2)

Note

$$logit\{Pr(R_i = j|R_i \geq j)\} = \alpha_{j0} + \alpha_1 Y_{i,j-1} + \alpha_2 Y_{i,j}$$

$\Rightarrow$

$$P_{ij} = Pr(R_i = j|R_i \geq j) = \frac{exp^{\alpha_{j0}+\alpha_1 Y_{i,j-1}+\alpha_2 Y_{i,j}}}{1 + exp^{\alpha_{j0}+\alpha_1 Y_{i,j-1}+\alpha_2 Y_{i,j}}}$$

$\Rightarrow$If $R_i < J + 1$, then

$$\begin{aligned}
\pi_i &= Pr(i\text{th subject drop outs at } R_i) \\
&= \prod_{j=2}^{R_i-1}(1 - P_{ij})P_{i,R_i} \\
&= \frac{exp^{\alpha_{R_i,0}+\alpha_1 Y_{i,R_i-1}+\alpha_2 Y_{i,R_i}}}{\prod_{j=2}^{R_i}\{1 + exp^{\alpha_{j0}+\alpha_1 Y_{i,j-1}+\alpha_2 Y_{i,j}}\}}
\end{aligned}$$

If $R_i = J + 1$ (complete case), then

$$\begin{aligned}
\pi_i &= Pr(i\text{th subject completes the study}) \\
&= \prod_{j=2}^{J}(1 - P_{ij}) = \frac{1}{\prod_{j=2}^{J}\{1 + exp^{\alpha_{j0}+\alpha_1 Y_{i,j-1}+\alpha_2 Y_{i,j}}\}}
\end{aligned}$$

Complete data likelihood for (β, α):

$$
\begin{aligned}
& L_C(\mathbf{Y}_{obs}, \mathbf{Y}_{mis}, R; \boldsymbol{\beta}, \boldsymbol{\alpha}) \\
= \; & L(\boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}; \beta) L(\mathbf{R} | \boldsymbol{Y}_{obs}, \boldsymbol{Y}_{mis}; \boldsymbol{\alpha}) \\
\propto \; & \prod_{i=1}^{m} \{ |\mathbf{V}|^{-\frac{1}{2}} e^{(Y_i - \mathbf{X}_i \beta)^T \mathbf{V}^{-1} (\mathbf{Y}_i - \mathbf{X}_i \beta)} \left\{ \frac{exp^{\alpha_{R_i,0} + \alpha_1 Y_{i,R_i - 1} + \alpha_2 Y_{i,R_i}}}{\prod_{j=2}^{R_i} \{ 1 + exp^{\alpha_{j0} + \alpha_1 Y_{i,j-1} + \alpha_2 Y_{i,j}} \}} \right\}^{I(R_i < J+1)} \\
& \left\{ \frac{1}{\prod_{j=2}^{J} \{ 1 + exp^{\alpha_{j0} + \alpha_1 Y_{i,j-1} + \alpha_2 Y_{i,j}} \}} \right\}^{I(R_i = J+1)}
\end{aligned}
$$

Observed data likelihood for (β, α) :

$$
\begin{aligned}
& \prod_{i=1}^{m} L(\mathbf{Y}_{i,obs} | \mathbf{X}_i; \boldsymbol{\beta}, \boldsymbol{\alpha}) \\
= \; & \prod_{i=1}^{m} \int L(Y_{i1}, \cdots, Y_{iR_i}, \mathbf{X}_i; \beta) L(R_i | Y_{i1}, \cdots, Y_{iR_i}, \mathbf{X}_i, \alpha_0, \alpha_1, \alpha_2) d\mathbf{Y}_{iR_i}
\end{aligned}
$$

Remarks: The integrated likelihood is maximized wrt $\boldsymbol{\beta}, \alpha_0, \alpha_1$for fixed α_2 and a sensitivity analysis is performed by varying α_2 around 0 (MAR).

Example: Methadone Treatment Practice Survey

- National panel survey (88, 90, 95) of 172 drug abuse treatment units.
- Major goal: Have the treatment practices changed over time?
- **Treatment Practice Measures:**
  - Maximum methadone dose levels
- **Covariates:**
  - Client characteristics, e.g., % African American
  - unit characteristics, e.g., ownership (profit, non-profit).

Example: Methadone Treatment Practice Survey

Means of the Outcome Variables by Year

Outcome	1988 ($m = 172$)	1990 ($m = 140$)	1995 ($m = 116$)
Max dose	78.0	81.9	93.0

Comparison of means for responding (Res) and dropping out units (DO)

	1988 Data		1990 Data	
Outcome	Res. '90 ($n = 140$)	DO '90 ($n = 32$)	Res. '95 ($n = 116$)	DO '95 ($n = 24$)
Upper Limit Dose	78.0	77.9	83.6	74.0

Example: Methadone Treatment Practice Survey

- Let

$$P_{ij} = Pr\{R = j | R_i \geq j, \mathbf{H}_{i,j-1}, Y_{ij}\}$$

  where $\mathbf{H}_{i,j-1}$ is the history data of Y and $\mathbf{X}$

- Assumed Model of P_{ij}:

$$logit(P_{ij}) = \psi_{j0} + \psi_1 Y_{i,j-1} + \psi_2 Y_{ij}$$

- This model allows dropout to be **nonignorable**.
- α_2 is not identifiable from the observed data.

Example: Parameter Estimates and Estimated SEs

	Log upper-limit dose (mg/d)		
Parameter	$\psi_2 = -1$	$\psi_2 = 0$	$\psi_2 = 1$
Int.	4.19	4.19	4.19
	.18	.18	.18
PFP	−.08	−.07	−.06
	.12	.12	.12
MW	−.09	−.09	−.09
	.07	.07	.07
SW	−.09	−.09	−.09
	.05	.05	.05
'90	.04	.05	.05
	.03	.03	.03
'95	.13	.15	.17
	.03	.03	.03
%AA	−.003	−.003	−.003
	.001	.001	.001
AGE (years)	.009	.009	.01
	.005	.005	.005
EX	−.37	−.35	−.33
	.20	.20	.20
ψ_{02}	.86	−.87	−2.82
	2.44	2.39	2.37
ψ_{03}	1.67	−.12	−2.16
	2.46	2.41	2.40
ψ_1	.39	−.22	−.78
	.56	.55	.55

Example: Parameter Estimates and Estimated SEs

Notes (because Mike forgets what he wants to say):

- Second column ($\psi_2 = 0$) is MAR
- '95 is the main interest, here: there is a significant change from 90-95 regardless of MAR/NMAR
- Not much difference between columns (coefficients are similar), but there are some differences
- If coefficients don't change much $\rightarrow$ MAR

Random-Coefficient Selection Model

Recall Direct Selection Model for Normal Data:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \boldsymbol{\epsilon_i} \quad (2)$$

$$logit\{Pr(R_i = j | R_i \geq j)\} = \alpha_{j0} + \alpha_1 Y_{i,j-1} + \alpha_2 Y_{ij} \quad (3)$$

where $\boldsymbol{\epsilon_i} \sim N(\mathbf{0}, \mathbf{V})$.

Idea of the random coefficient selection model: Write the Y model as a random effect model and replace the Y in the dropout model by its latent trajectories, e.g. individual intercepts and slopes.

Random-Coefficient Selection Model

Y **model**:

$$Y_{ij} = \beta_{i0} + \beta_{i1} t_j + \mathbf{X}_{ij}^T \boldsymbol{\beta}_x + e_{ij}$$

where $e_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$, $\boldsymbol{\beta}_i \sim N(\beta_t, \mathbf{D})$ are subject-specific intercepts and slopes.

R **Model:**

Replace $Y_{i,j-1}$and Y_{ij} in the dropout model (3) by subject-specific means $E(Y_{i,j-1}|\boldsymbol{\beta}_i, \mathbf{X}_i)$ and $E(Y_{i,j}|\boldsymbol{\beta}_i, \mathbf{X}_i)$.

Random-Coefficient Selection Model

Random-coefficient model:

$$
\begin{aligned}
Y_{ij} &= \beta_{i0} + \beta_{i1} t_j + \mathbf{X}_{ij}^T \boldsymbol{\beta}_x + e_{ij} \\
logit\{Pr(R_i = j | R_i \geq j)\} &= \alpha_{j0} + \alpha_1 \mu_{i,j-1} + \alpha_2 \mu_{ij}
\end{aligned}
$$

where $e_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$ and $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}_{\mathbf{t}}, \mathbf{D})$ and
$\mu_{ij-1} = E(Y_{i,j-1} | \boldsymbol{\beta}_i, \mathbf{X}_{ij-1})$ and $\mu_{ij} = E(Y_{ij} | \boldsymbol{\beta}_i, \mathbf{X}_{ij})$

Remarks:

- The Y model and the R model share common parameters $\boldsymbol{\beta}_t$. This makes calculations of the MLE more challenging.

Likelihood for $\beta, \boldsymbol{\alpha}$:

$$
\begin{aligned}
& L_C(\boldsymbol{\beta}, \boldsymbol{\alpha}_0, \alpha_1, \alpha_2) \\
= \; & \prod_i \int L_C(Y_{i1}, \cdots, Y_{iR_i} | \boldsymbol{\beta}_i) L(R_i | \boldsymbol{\beta}_i) f(\boldsymbol{\beta}_i) d\boldsymbol{\beta}_i \\
\propto \; & \prod_i \int \Bigg\{ \left[\frac{e^{\alpha_{R_i,0} + \alpha_1 \mu_{i,R_i-1} + \alpha_2 \mu_{i,R_i}}}{\prod_{j=2}^{R_i} \{1 + e^{\alpha_{j0} + \alpha_1 \mu_{i,j-1} + \alpha_2 \mu_{i,j}}\}} \right]^{I(R_i < J+1)} \\
& \left[\frac{1}{\prod_{j=2}^{J} \{1 + e^{\alpha_{j0} + \alpha_1 \mu_{i,j-1} + \alpha_2 \mu_{i,j}}\}} \right]^{I(R_i = J+1)} \\
& \prod_{j=1}^{J} e^{(Y_{ij} - \beta_{i0} - \beta_{i1} t_j - \mathbf{X}_{ij}^T \beta_x)^2 / 2\sigma^2} \Bigg\} e^{(\boldsymbol{\beta}_i - \boldsymbol{\beta}_t)^T \mathbf{D}^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_t)} d\boldsymbol{\beta}_i
\end{aligned}
$$

- The likelihood is too complicated.
- β is involved in both the complete data model and the missing mechanism model.

Shared Random Effects Model:

Idea: Partition $\boldsymbol{\beta}_i$ into a linear combination of fixed and random effects.

$$\begin{aligned} Y_{ij} &= b_{i0} + b_{i1} t_{ij} + \beta_0 + \beta_1 t_{ij} + \mathbf{X}_{ij}^T \boldsymbol{\beta}_x + e_{ij} \\ logit(\Pr(R_i = j | R_i \geq j)) &= \alpha_{j0} + \alpha_1 b_{i0} + \alpha_2 b_{i1} \end{aligned}$$

where

- $e_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$
- b_{i0}: unknown random intercept departure for subject i.
- b_{i1}: unknown random slope departure for subject i.
- $\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$

Note β is only involved in the complete data model.

This is a common idea to jointly model longitudinal data and survival data. In that case, the logistic model is replaced by Cox model.

Likelihood for $\boldsymbol{\beta}, \boldsymbol{\alpha}$ in the Shared Random Effects Model:

$$
\begin{aligned}
& L_C(\boldsymbol{\beta}, \boldsymbol{\alpha}_0, \alpha_1, \alpha_2) \\
= & \prod_i \int L_C(\beta; \boldsymbol{Y}_{i,obs}, \boldsymbol{Y}_{i,mis} | \mathbf{b}_i) \\
& L(\boldsymbol{\alpha}_0, \alpha_1, \alpha_2, \beta; \mathbf{R}_i | \boldsymbol{Y}_{i,obs}, \boldsymbol{Y}_{i,mis}, \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\
\propto & \prod_i \int \left\{ \left[\frac{e^{\alpha_{R_i,0} + \alpha_1 b_{i0} + \alpha_2 b_{i1}}}{\prod_{j=2}^{R_i} \{1 + exp^{\alpha_{j0} + \alpha_1 b_{i0} + \alpha_2 b_{i1}}\}} \right]^{I(R_i < J+1)} \right. \\
& \left[\frac{1}{\prod_{j=2}^{J} \{1 + e^{\alpha_{j0} + \alpha_1 b_{i0} + \alpha_2 b_{i1}}\}} \right]^{I(R_i = J+1)} \\
& \left. \prod_{j=1}^{J} e^{(Y_{ij} - b_{i0} - b_{i1} t_j - \beta_0 - \beta_1 t_j - \mathbf{X}_{ij}^T \beta_x)^2 / 2\sigma^2} \right\} e^{\mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i} d\mathbf{b}_i
\end{aligned}
$$