Modelling Missing Data

Complete-data model:
- independent and identically distributed (iid) draws \( Y_1, \ldots, Y_n \)
- from multivariate distribution \( P_\theta \)
\[
Y_i = (Y_{i1}, \ldots, Y_{ip})^T \sim P_\theta
\]

- Data matrix:
\[
Y = (Y_1, \ldots, Y_n)^T = (Y_{ij})_{1 \leq i \leq n, 1 \leq j \leq p}
\]

Inference: estimate \( \theta \)
- likelihood approach → maximize likelihood function
- Bayesian approach → consider posterior distribution

Likelihood function of the complete data:
\[
f_Y(y|\theta) = \prod_{i=1}^{n} f_{Y_i}(y_i|\theta)
\]
where \( f_{Y_i}(.|\theta) \) is the density of \( P_\theta \)

Modelling Missing Data

Idea: describe missingness of \( Y_{ij} \) by indicator variable \( R_{ij} \)
\[
R_{ij} = \begin{cases} 1 & Y_{ij} \text{ has been observed} \\ 0 & Y_{ij} \text{ is missing} \end{cases}
\]

Multivariate dataset with missing values:

\[
\begin{array}{c|c|c|c|c|c|c}
\text{Units} & 1 & 2 & \cdots & n \\
\hline
1 & Y_{11} & Y_{12} & \cdots & Y_{1p} \\
2 & R_{11} & R_{12} & \cdots & R_{1p} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
n & \vdots & \vdots & \ddots & \vdots \\
\end{array}
\]

Observed-data model:
- The observed information consists of
  - observed values \( y_{ij} \)
  - the values \( r_{ij} \) indicating which values are missing
- A statistical model for the observations should make use of all available information

\[
Y = (Y_{obs}, Y_{mis})
\]
where
- \( Y_{obs} \) are the observed variables
- \( Y_{mis} \) are the missing variables

Missing Data Mechanism

Statistical model for missing data:
\[
P(R = r \mid Y = y) = f_{R|Y}(r | y, \xi)
\]
with parameters \( \xi \in \Xi \).

Definition Missing completely at random (MCAR)
The observations are missing completely at random if
\[
P(R = r \mid Y = y) = P(R = r)
\]
or equivalently
\[
f_{R|Y}(r | y, \xi) = f_R(r, \xi),
\]
that is, \( R \) and \( Y \) are independent.

Definition Missing at random (MAR)
The observations are missing at random if
\[
P(R = r \mid Y = y) = P(R = r \mid Y_{obs})
\]
or equivalently
\[
f_{R|Y}(r | y, \xi) = f_{R|Y_{obs}}(r | y_{obs}, \xi),
\]
that is, knowledge about \( Y_{mis} \) does not provide any additional information about \( R \) if \( Y_{obs} \) is already known (\( R \) and \( Y_{mis} \) are conditionally independent given \( Y_{obs} \)).

Definition Not missing at random (NMAR)
Observations are not missing at random if the above assumption for MAR is not fulfilled.
Missing Completely at Random (MCAR)

Example: Bernoulli selection
- Complete data: \( Y_1, \ldots, Y_n \)
- Response indicator: \( R_1, \ldots, R_n \)
- Missingness mechanism: Each unit is observed with probability \( \xi \)
  
  \[ f_R(r|\xi) = (\xi)^r (1-\xi)^{1-r}, \]

Example: Simple random sample (SRS)
- Suppose that only \( m \) observations are available, \( n-m \) values are missing.
- Responding units are simple random sample of all units:
  
  \[ f_R(r|\xi) = \frac{(\frac{m}{n})^r}{(1-\frac{m}{n})^{n-r}}, \quad \text{if } \sum_i R_i = m \]
  
  \[ = 0, \quad \text{otherwise} \]
- Note that \( \xi = m/n \) is a parameter of the distribution of \( R \)

Missing at Random (MAR)

Examples for data missing at random:
- Double sampling: sample survey:
  - observe variables \( Y_{ik} \) for all individuals \( i = 1, \ldots, n \)
  - observe variables \( Y_{ip} \) only for subsample
- Sampling with nonresponse followup: sample survey with nonresponses
  - intensive followup effort for all units not feasible
  - take random sample of all nonresponding units
- Randomized experiment with unequal numbers of cases per treatment group:
  - suppose original design was balanced
  - missingness mechanism is deterministic, thus data are MAR
- Matrix sampling for questionnaire or test items
  - divide test or questionnaire into sections
  - groups of sections are administered to subjects in a randomized fashion

Data: Serial measurements of blood pressure during long-term trials of drugs which reduce blood pressure

Problem: Patients are withdrawn during the course of the study because of inadequate blood pressure control.

- International double-blind randomized study with 429 patients
  - 218 receive \( \beta \)-blocker metoprolol
  - 211 receive serotonergic \( 5-HT_2 \)-receptor blocking agent ketanserin
- Treatment phase lasted 12 weeks
- Scheduled clinic visits for weeks 0, 2, 4, 8, and 12
- Patients with diastolic blood pressure exceeding 110 mmHg in week 4 or week 8 “jump” to an open follow-up phase.
- Missing data pattern for the two groups:

<table>
<thead>
<tr>
<th>Metoprolol group</th>
<th>Ketanserin group</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 2 4 8 12 Number of patients</td>
<td>0 2 4 8 12 Number of patients</td>
</tr>
<tr>
<td>1 1 1 1 1 1 162</td>
<td>1 1 1 1 1 1 136</td>
</tr>
<tr>
<td>1 1 1 0 1 1 1 14 (11 jumps)</td>
<td>1 1 1 0 0 1 0 19 (18 jumps)</td>
</tr>
<tr>
<td>1 1 1 1 1 1 1 14 6 (28 jumps)</td>
<td>1 1 1 1 0 0 1 41</td>
</tr>
<tr>
<td>1 0 1 1 1 1</td>
<td>1 1 0 0 0 0</td>
</tr>
<tr>
<td>1 0 0 1 1 1</td>
<td>1 0 0 1 1 1</td>
</tr>
<tr>
<td>0 1 0 0 0 1</td>
<td>1 0 0 1 0 1</td>
</tr>
<tr>
<td>0 0 0 1 1 1</td>
<td>1 0 0 0 0 1</td>
</tr>
<tr>
<td>1 0 1 1 1 1</td>
<td>1 0 0 0 0 1</td>
</tr>
<tr>
<td>1 0 1 1 0 1</td>
<td>1 0 0 0 0 1</td>
</tr>
<tr>
<td>1 0 1 0 1 1</td>
<td>1 0 1 0 0 1</td>
</tr>
<tr>
<td>1 0 1 0 0 1</td>
<td>1 0 1 0 0 1</td>
</tr>
<tr>
<td>1 0 0 1 1 1</td>
<td>1 0 0 1 1 1</td>
</tr>
<tr>
<td>1 0 0 0 1 1</td>
<td>1 0 0 0 1 1</td>
</tr>
<tr>
<td>1 0 0 0 0 1</td>
<td>1 0 0 0 0 1</td>
</tr>
</tbody>
</table>

- “Jumps” (majority of missing observations) lead to missing data which are missing at random.

Mean diastolic blood pressure based on:
- all available data at each time point
- patients with complete data
- all available data, assuming missing data to be missing at random

\[ \text{Weeks on Treatment} \]
\[ \text{Mean DBP (mmHg)} \]
\[ \begin{array}{cccc}
0 & 4 & 8 & 12 \\
90 & 95 & 100 & 105 & 110
\end{array} \]

\[ \text{Ketanserin} \quad \text{Metoprolol} \]
\[ \begin{array}{cccc}
(a) & (b) & (c)
\end{array} \]
Not Missing at Random (NMAR)

Example: Medical treatment of blood pressure
n individuals are treated for high blood pressure
- **Observations:** binary responses
  \[ Y_{ij} = \text{blood pressure of } i\text{th individual on } j\text{th day} \]
- **Missing observations:** (not every individual shows up every day)
  \[ R_{ij} = \begin{cases} 1 & \text{if } i\text{th individual appears for measurement on } j\text{th day} \\ 0 & \text{otherwise} \end{cases} \]
- **Statistical model for } Y \text{ and } R: \]
  \[ Y_{ij} \sim N(\mu_i, \sigma^2) \]
  \[ R_{ij} \sim \text{Bin}(1, 1 - \exp(Y_{ij})) \]
  \[
  \text{\(\sim\)} \text{ that is, individuals with high blood pressure are less likely to turn up for a measurement}

Observed-Data Likelihood

Likelihood of complete observations \((Y_{\text{obs}}, R)\)
\[
L_{\text{obs}}(\theta, \xi | Y_{\text{obs}}, R) = f_{Y_{\text{obs}}}(Y_{\text{obs}}, R | \theta, \xi) \\
= \int f_{Y_{\text{obs}}}(Y_{\text{obs}} | \theta, \xi) d_{\text{obs}} \cdot \int f_{R_{\text{obs}}}(R_{\text{obs}} | \theta, \xi) d_{\text{obs}}
\]
If the missing observations are missing at random (MAR), the first factor does not depend on \(y_{\text{mis}}\) and thus is not affected by integration over \(y_{\text{mis}}\):
\[
L_{\text{obs}}(\theta, \xi | Y_{\text{obs}}, R) = \int f_{Y_{\text{obs}}}(Y_{\text{obs}} | \xi) f_{R_{\text{obs}}}(R_{\text{obs}} | \theta, \xi) d_{\text{obs}} \\
= f_{Y_{\text{obs}}}(Y_{\text{obs}} | \xi) \int f_{R_{\text{obs}}}(R_{\text{obs}} | \theta, \xi) d_{\text{obs}} \\
= f_{Y_{\text{obs}}}(Y_{\text{obs}} | \xi) f_{R_{\text{obs}}}(R_{\text{obs}} | \theta) L_{\theta}(\theta | Y_{\text{obs}})
\]
If the parameters \(\xi\) do not depend on \(\theta\), we get
\[
\hat{\theta} = \arg \max_{\theta} L_{\text{obs}}(\theta, \xi | Y_{\text{obs}}, R) \iff \hat{\theta} = \arg \max_{\theta} L_{\theta}(\theta | Y_{\text{obs}}).
\]
that is, we can obtain the maximum likelihood estimator by minimizing \(L_{\theta}(\theta | Y_{\text{obs}})\) with respect to \(\theta\).

**Definition** Observed-data likelihood
The likelihood function
\[
L_{\theta}(\theta | Y_{\text{obs}}) = f_{Y_{\text{obs}}}(Y_{\text{obs}} | \theta)
\]
is called the likelihood ignoring the missing-data mechanism or short observed-data likelihood.

Ignorability

Definition
A missing-data mechanism is **ignorable for likelihood inference** if
- observations are missing at random (MAR) and
- the parameters \(\xi\) (missingness-mechanism) and \(\theta\) (data model) are distinct, in the sense that the joint parameter space of \((\theta, \xi)\) is the product of the parameter spaces \(\Xi\) and \(\Theta\).

Example: Non-distinct parameters
Suppose that
- \(Y_i \iid \text{Bin}(1, \theta)\)
- \(R_i \iid \text{Bin}(1, \theta)\)
- \(Y_{\text{mis}} = (Y_1, \ldots, Y_n)^T\) (after reordering)
The joint likelihood of \(Y_{\text{obs}}\) and \(R\) is
\[
L_{\text{obs}}(\theta | Y_{\text{obs}}, R) = \prod_{i=1}^{n} \theta^y_i (1 - \theta)^{1-y_i} \cdot \prod_{i=1}^{m} \theta^m_i (1 - \theta)^{1-m_i}
\]
This is the likelihood of a Bernoulli experiment with \(n + m\) observations. Consequently
\[
\hat{\theta}_{\text{ML}} = \frac{m + Y}{m + n}
\]
**Ignoring the missing-data mechanism:** The observed-data likelihood is
\[
L_{\theta}(\theta | Y_{\text{obs}}) = \prod_{i=1}^{n} \theta^y_i (1 - \theta)^{1-y_i}
\]
which leads to the ML estimator
\[
\hat{\theta}_{\text{ML}} = \frac{Y}{m}
\]
Ignorability

The first condition (MAR) is typically regarded as the more important condition:
- If MAR does not hold, then the maximum likelihood estimator based on the observed-data likelihood can be seriously biased.
- If the data are MAR but distinctness does not hold, inference based on the observed-data likelihood $L_n(\theta | Y_{obs})$ is still valid from the frequentist perspective, but not fully efficient.

Example:
Suppose $Y_i \sim N(\mu, \sigma^2)$
- $R_i \overset{iid}{\sim} \text{Bin}(1, \sigma^2/(1 + \sigma^2))$
- Parameter $\sigma^2$ is not distinct

Frequentist interpretation:
- Parameter $\theta_0 = (\mu_0, \sigma^2_0)$ is fixed.
- If $R_i \overset{iid}{\sim} \text{Bin}(1, \frac{1}{2})$ then:
  - Missingness mechanism ignorable
  - $\hat{\mu}_{obs}$ is the correlation between $Y_1$ and $Y_2$: $Y_2 | Y_1 \sim N(\beta_1 Y_1, \sigma^2_2)$
- However, $\hat{\mu}_{comp}$ is more efficient.

Multivariate Normal Distribution

Let $Y = (Y_1, \ldots, Y_p)^T \sim N(\mu, \Sigma)$

Properties
- If $p = 2$
  \[ \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \]
  where
  \[ \rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2} \]
  is the correlation between $Y_1$ and $Y_2$
- Density of $Y$
  \[ f_Y(y | \mu, \Sigma) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2}(y - \mu)^T \Sigma^{-1} (y - \mu) \right). \]
- Linear transformations
  \[ AY + b \sim N(A\mu + b, A\Sigma A^T). \]
- Marginal distribution of $Y_1$
  \[ Y_1 \sim N(\mu_1, \sigma^2_1) \]
- Conditional distribution of $Y_2$ given $Y_1$
  \[ Y_2 | Y_1 \sim N(\beta_1 Y_1, \sigma^2_2), \]
  with parameters
  \[ \beta_1 = \frac{\mu_2 - \beta_1 \mu_1}{\sigma^2_2 - \sigma_{12}^2 / \sigma_1^2} \]
  \[ \sigma^2_2 = \frac{\sigma^2_1}{\sigma^2_2} - \frac{\sigma^2_{12}^2}{\xi_1^2} \]

Univariate Data with Missing Observations

Example: Incomplete univariate data

<table>
<thead>
<tr>
<th>Unit</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$R$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Suppose that
- Only $Y_{obs} = (Y_1, \ldots, Y_m)^T$ have been observed
- $Y_{mis} = (Y_{m+1}, \ldots, Y_P)^T$ are missing at random

The observed-data likelihood is
\[ L_n(\theta | Y_{obs}) = \prod_{i=1}^{m} f_Y(Y_i | \mu, \Sigma) \prod_{i=m+1}^{n} f_Y(Y_i | \mu_{mis}, \Sigma_{mis}). \]

Bivariate Normal Data

Example: Bivariate data with one variable subject to nonresponse

<table>
<thead>
<tr>
<th>Units</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$R_1$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Suppose that
- $Y_{1}, \ldots, Y_n$ is an iid random sample
- Only $Y_{obs} = (Y_1, \ldots, Y_m)^T$ have been observed
- $Y_{mis} = (Y_{m+1}, \ldots, Y_P)^T$ are missing at random

The observed-data likelihood is
\[ L_n(\theta | Y_{obs}) = \prod_{i=1}^{m} f_{Y_1}(Y_i | \mu, \Sigma) \prod_{i=m+1}^{n} f_{Y_2}(Y_i | \mu_{mis}, \Sigma_{mis}). \]
Bivariate Normal Data

Example: Bivariate data with one variable subject to nonresponse

Data: \(Y_i = (Y_{i1}, Y_{i2}) \sim N(\mu, \Sigma)\)
- Parameter: \(\mu_1 = \mu_2 = 0, \sigma_1^2 = \sigma_2^2 = 1, \rho\)
- \(Y_{i1}\) completely observed
- \(Y_{i2}\) has missing values

Missing data mechanism: Proportion of missing data \(1 - p = 20\%
- Missing completely at random (MCAR):
  \[R_{i2} \sim \text{Bin}(1, p)\]
- Missing at random (MAR):
  \[R_{i2} = 0 \text{ if } Y_{i1} > z_p\]
- Not missing at random (NMAR):
  \[R_{i2} = 0 \text{ if } Y_{i2} > z_p\]

Simulation study:
- 1000 repetitions for \(\rho = 0, 0.5, 0.9\)
- Simulated means (standard variations) for \(\hat{\mu}_{\text{CC}}\) and \(\hat{\mu}_{\text{ML}}\)

<table>
<thead>
<tr>
<th></th>
<th>MCAR</th>
<th>MAR</th>
<th>NMAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\rho = 0)</td>
<td>(0.00046)</td>
<td>(-0.00095)</td>
<td>(-0.34585)</td>
</tr>
<tr>
<td></td>
<td>((0.10762))</td>
<td>((0.10876))</td>
<td>((0.08760))</td>
</tr>
<tr>
<td>(\rho = 0.5)</td>
<td>(0.00557)</td>
<td>(-0.17073)</td>
<td>(-0.34919)</td>
</tr>
<tr>
<td></td>
<td>((0.11294))</td>
<td>((0.10585))</td>
<td>((0.08406))</td>
</tr>
<tr>
<td>(\rho = 0.9)</td>
<td>(-0.00218)</td>
<td>(-0.31229)</td>
<td>(-0.34964)</td>
</tr>
<tr>
<td></td>
<td>((0.11012))</td>
<td>((0.09034))</td>
<td>((0.08460))</td>
</tr>
</tbody>
</table>

Maximum Likelihood with Missing Data

Maximum Likelihood with Missing Data, Mar 30, 2004 - 17 -

<table>
<thead>
<tr>
<th></th>
<th>Complete-case (MCAR)</th>
<th>MLE (MCAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\rho = 0)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td>(\rho = 0.5)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td>(\rho = 0.9)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
</tbody>
</table>

Maximum Likelihood with Missing Data

Maximum Likelihood with Missing Data, Mar 30, 2004 - 18 -

<table>
<thead>
<tr>
<th></th>
<th>Complete-case (MAR)</th>
<th>MLE (MAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\rho = 0)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td>(\rho = 0.5)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td>(\rho = 0.9)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
</tbody>
</table>

Maximum Likelihood with Missing Data

Maximum Likelihood with Missing Data, Mar 30, 2004 - 19 -

<table>
<thead>
<tr>
<th></th>
<th>Complete-case (NMAR)</th>
<th>MLE (NMAR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\rho = 0)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td>(\rho = 0.5)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td>(\rho = 0.9)</td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
<tr>
<td></td>
<td>(\hat{\mu}_{\text{CC}})</td>
<td>(\hat{\mu}_{\text{ML}})</td>
</tr>
</tbody>
</table>
Maximum Likelihood with Missing Data, Mar 30, 2004 - 21 -

Censored Variables

Example: Randomly censored data

- Objective: Analysis of data

\[ Y_1, \ldots, Y_n \sim \mathcal{N}(\mu, \sigma^2) \]

- Fixed censoring at \( c \)

\[ R_i = \begin{cases} 1 & Y_i \leq c \\ 0 & \text{otherwise} \end{cases} \]

The likelihood of the complete observations is

\[
L_\theta(Y_{obs}, R) = \prod_{i=1}^{m} f_i(Y_i; \mu, \sigma^2) \prod_{i=m+1}^{n} S_i(c)
\]

\[ = \prod_{i=1}^{m} f_i(Y_i; \mu, \sigma^2) \prod_{i=m+1}^{n} \left[ 1 - \Phi\left( \frac{c - \mu}{\sigma} \right) \right] \]

where

\[ S_i(y) = 1 - \Phi\left( \frac{y - \mu}{\sigma} \right) \]

is the survivor function of \( Y_i \).

Notation:

- \( \varphi(y) \) is the density of the standard normal distribution,

\[ \varphi(y) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} y^2 \right) \]

- \( \Phi(y) \) is the corresponding cumulative distribution function (CDF),

\[ \Phi(y) = \int_{-\infty}^{y} \varphi(z) \, dz. \]

Maximum Likelihood with Missing Data, Mar 30, 2004 - 22 -

Censored Variables

Example: Randomly censored data

Derivation of the likelihood:

We have for the missing data mechanism

\[ P(R_i = r_i | Y) = P(R_i = r_i | Y) = 1_{-\infty < Y_i} Y_i (1 - 1_{-\infty < Y_i})^{1 - r_i} \]

Hence the likelihood of the complete data (including \( R \)) is

\[
L_\theta(Y, R) = \prod_{i=1}^{m} f_i(Y_i; \theta) \prod_{i=m+1}^{n} 1_{-\infty < Y_i} Y_i (1 - 1_{-\infty < Y_i})^{1 - R_i}
\]

Since \( R_i = 1 \) and thus \( 1_{-\infty < Y_i} Y_i = 1 \) for uncensored observations \( Y_i \), we have

\[
L_\theta(Y_{obs}, R) = \int \prod_{i=1}^{m} f_i(Y_i; \theta) \prod_{i=m+1}^{n} f_i(Y_i; \theta) 1_{-\infty < Y_i} Y_i (1 - 1_{-\infty < Y_i})^{1 - R_i} \, dy_{m+1} \cdots dy_n
\]

\[ = \prod_{i=1}^{m} f_i(Y_i; \theta) \int \prod_{i=m+1}^{n} f_i(Y_i; \theta) 1_{-\infty < Y_i} Y_i \, dy_{m+1} \cdots dy_n
\]

\[ = \prod_{i=1}^{m} f_i(Y_i; \theta) \int \prod_{i=m+1}^{n} f_i(Y_i; \theta) dy_{m+1} \cdots dy_n
\]

Maximum Likelihood with Missing Data, Mar 30, 2004 - 23 -

Bivariate Normal Data

Example: Bivariate data with both variables subject to nonresponse

Units
\[
\begin{array}{cccc}
\text{Units} & 1 & 1 & 1 \\
1 & 1 & l & 1 \\
1 & l + 1 & 1 & 0 \\
\vdots & \vdots & \vdots & \vdots \\
m & 1 & 0 & 0 \end{array}
\]

Data: \( Y = (Y_1, Y_2)^T \sim \mathcal{N}(\mu, \Sigma) \)

- Both variables subject to nonresponse

The observed-data likelihood is given by

\[
L_\theta(Y_{obs}) = \prod_{i=1}^{m} f_i(Y_i; \theta) \prod_{i=m+1}^{n} f_i(Y_i; \theta) 1_{-\infty < Y_i} Y_i (1 - 1_{-\infty < Y_i})^{1 - R_i}
\]

In this case, no appropriate parametrization which leads to a factorization into likelihoods with distinct parameters can be found.