Incomplete Data Analysis

Vanda Inácio

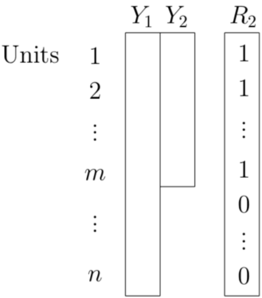
University of Edinburgh

Semester 1, 2020/2021

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

- Consider a medical study where measurements Y_1 are taken at baseline (i.e., at the beginning of the study) and Y_2 at follow-up.
- All individuals have their outcome recorded at baseline but some have their follow-up outcome missing.



Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

$\hookrightarrow$ We consider that $\mathbf{Y} = (Y_1, Y_2)'$ follows a bivariate normal distribution, that is, $\mathbf{Y} \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

$\hookrightarrow$ Remember that the density function of a bivariate normal distribution is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi}(\det\boldsymbol{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\}.$$

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

- ↪ The full data for the ith individual is (y_{1i}, y_{2i}, r_i), where $r_i = 1$ if y_{2i} is observed and $r_i = 0$ if y_{2i} is missing, for $i = 1 \ldots, n$.

- ↪ We assume, without loss of generality, that y_{2i} is observed for $i = 1, \ldots, m$ (i.e., $r_i = 1$, $i = 1, \ldots, m$) and y_{2i} is missing for $i = m+1, \ldots, n$ (i.e., $r_i = 0$, $i = m+1, \ldots, n$).

- ↪ The full data model is

$$f(\mathbf{y}, \mathbf{r} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi) = \prod_{i=1}^{n} f(y_{1i}, y_{2i}, r_i \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi).$$

- ↪ Under ignorability, i.e., assuming that missing in Y_2 is MAR (or MCAR) and that $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{\psi}$ are distinct, we can ‘simply’ work with the likelihood of the observed data in order to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- ↪ Remember that distinctness of the parameters formally implies that the parameter space of $((\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\psi})$ is equal to the Cartesian product of their individual product spaces. Informally stated, this means that the model for the missing data mechanism does not contain information about the parameters of the complete data model.

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

↪ The likelihood of the observed data, which is

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid \mathbf{y}_{\text{obs}}) &= \prod_{i=1}^{m} f(y_{1i}, y_{2i} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{i=m+1}^{n} f(y_{1i} \mid \mu_1, \sigma_1) \\
&= \prod_{i=1}^{m} \left\{ \frac{1}{2\pi} (\det \boldsymbol{\Sigma})^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}) \right] \right\} \\
&\times \prod_{i=m+1}^{n} \left\{ \left(2\pi\sigma_1^2\right)^{-1/2} \exp\left[-\frac{1}{2\sigma_1^2} (y_{1i} - \mu_1)^2 \right] \right\}. \qquad (1)
\end{aligned}
$$

↪ Note that $\mathbf{y}_{\text{obs}} = (y_{1i}, \ldots, y_{1n}, y_{21}, \ldots, y_{2m})$ and $\mathbf{y}_i = (y_{1i}, y_{2i})'$.

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

↪ Maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be obtained by maximising the log likelihood function corresponding to the likelihood in (1) with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

↪ The likelihood equations, however, do not have an obvious solution.

↪ A more convenient parametrisation, for finding the ML estimates analytically in this example, uses a factored likelihood (Little and Rubin, 2002, Chapter 7), in which the several 'factors' or terms appearing in the likelihood have distinct parameters and can be maximised separately.

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

Known facts for a bivariate normal distribution

$\hookrightarrow$ If $\mathbf{Y} = (Y_1, Y_2)' \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$
\begin{aligned}
& Y_1 \sim \mathrm{N}(\mu_1, \sigma_1^2), \\
& Y_2 \mid Y_1 = y_1 \sim \mathrm{N}(\beta_0 + \beta_1 y_1, \sigma_{2|1}^2), \\
& \beta_0 = \mu_2 - \beta_1 \mu_1, \\
& \beta_1 = \frac{\sigma_{12}}{\sigma_1^2}, \\
& \sigma_{2|1}^2 = \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}.
\end{aligned}
$$

$\hookrightarrow$ Proofs of these results can be found in almost every book on multivariate analysis (or somewhere in the web!). I personally like the following book: Rencher, A.C. (2002) *Methods of Multivariate Analysis*, Wiley (see Chapter 4, pp. 82–111, about the multivariate normal distribution).

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

- ↪ Remember that we can write the joint density of Y_1 and Y_2 as the product of the conditional density of Y_2 given $Y_1 = y_1$ and the marginal density of Y_1 (and from the results stated in the previous slide, we know what the parameters of these two distributions are)

$$f(y_1, y_2 \mid \phi) = f(y_2 \mid y_1, \beta_0, \beta_1, \sigma_{2|1}) f(y_1 \mid \mu_1, \sigma_1).$$

- ↪ The crux here is that the parameter $\phi \equiv \phi(\boldsymbol{\theta}) = (\mu_1, \sigma_1, \beta_0, \beta_1, \sigma_{2|1})'$ is a one-to-one function of the original parameter $\boldsymbol{\theta} = (\mu_1, \sigma_1, \mu_2, \sigma_2, \sigma_{12})'$ and the parameters of the conditional and marginal densities are distinct since knowledge on (μ_1, σ_1) does not imply any information about $(\beta_0, \beta_1, \sigma_{2|1})$.
- ↪ Saying that ϕ is a one-to-one function of $\boldsymbol{\theta}$ means that all elements of ϕ can be written as a function of elements of $\boldsymbol{\theta}$ and the other way around and that they are uniquely determined.
- ↪ In the previous slide we have already seen how to write the elements of ϕ as a function of the elements of $\boldsymbol{\theta}$.
- ↪ Similarly, the components of $\boldsymbol{\theta}$ other than μ_1 and σ_1 can be expressed as the following functions of the components of ϕ

$$\mu_2 = \beta_0 + \beta_1 \mu_1$$
$$\sigma_{12} = \beta_1 \sigma_1^2$$
$$\sigma_2^2 = \sigma_{2|1}^2 + \beta_1^2 \sigma_1^2$$

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

↪ Finally, the likelihood for the observed data simplifies to

$$\begin{aligned}
L(\boldsymbol{\theta} \mid \mathbf{y}_{\text{obs}}) &= \prod_{i=1}^{m} f(y_{1i}, y_{2i} \mid \boldsymbol{\theta}) \prod_{i=m+1}^{n} f(y_{1i} \mid \boldsymbol{\theta}) \\
&= \left\{ \prod_{i=1}^{m} f\left(y_{1i} \mid \mu_1, \sigma_1\right) f\left(y_{2i} \mid y_{1i}, \beta_0, \beta_1, \sigma_{2|1}\right) \right\} \prod_{i=m+1}^{n} f\left(y_{1i} \mid \mu_1, \sigma_1\right) \\
&= \prod_{i=1}^{m} f\left(y_{2i} \mid y_{1i}, \beta_0, \beta_1, \sigma_{2|1}\right) \prod_{i=1}^{n} f\left(y_{1i} \mid \mu_1, \sigma_1\right) \\
&= \prod_{i=1}^{m} \left\{ (2\pi\sigma_{2|1}^2)^{-1/2} \exp\left[-\frac{1}{2\sigma_{2|1}^2} (y_{2i} - (\beta_0 + \beta_1 y_{1i}))^2 \right] \right\} \\
&\times \prod_{i=1}^{n} \left\{ (2\pi\sigma_1^2)^{-1/2} \exp\left[-\frac{1}{2\sigma_1^2} (y_{1i} - \mu_1)^2 \right] \right\} \qquad (2)
\end{aligned}$$

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

- ↪ The likelihood in (2) is much more tractable, from an analytical point of view, than the one in (1).

- ↪ The first term is the density for m observations from a conditional normal distribution with mean $\beta_0 + \beta_1 y_1$ and variance $\sigma^2_{2|1}$.

- ↪ The second factor is the density of an independent sample of size n from a normal distribution with mean μ_1 and variance σ^2_1.

- ↪ Because we know that the parameters of the two terms are distinct, ML estimates of ϕ can be obtained by independently maximising the likelihoods corresponding to these two components.

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

↪ After some calculations, we obtain

$$\widehat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} y_{1i}, \quad \widehat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(y_{1i} - \widehat{\mu}_1)^2$$

$$\widehat{\beta}_0 = \bar{y}_2 - \widehat{\beta}_1 \bar{y}_1, \quad \bar{y}_j = \frac{1}{m}\sum_{i=1}^{m} y_{ji}, \quad j = 1, 2,$$

$$\widehat{\beta}_1 = \frac{s_{12}}{s_1^2}, \quad s_{12} = \frac{1}{m}\sum_{i=1}^{m}(y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2), \quad s_1^2 = \frac{1}{m}\sum_{i=1}^{m}(y_{1i} - \bar{y}_1)^2,$$

$$\widehat{\sigma}_{2|1}^2 = s_2^2 \frac{s_{12}^2}{s_1^2}, \quad s_2^2 = \frac{1}{m}\sum_{i=1}^{m}(y_{2i} - \bar{y}_2)^2.$$

↪ Note that under a complete case analysis, the likelihood would be

$$L(\boldsymbol{\theta} \mid y_{11}, \ldots, y_{1m}, y_{21}, \ldots, y_{2m}) = \prod_{i=1}^{m} f(y_{1i}, y_{2i} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

- ↪ We have conducted a simulation study to ascertain the performance of both the mle and complete cases approach in estimating μ_2. Note that $\widehat{\mu}_2 = \widehat{\beta}_0 + \widehat{\beta}_1\widehat{\mu}_1$.

- ↪ We have that $\widehat{\mu}_2^{\text{CCA}} = \frac{1}{m}\sum_{i=1}^{m} y_{2i}$.

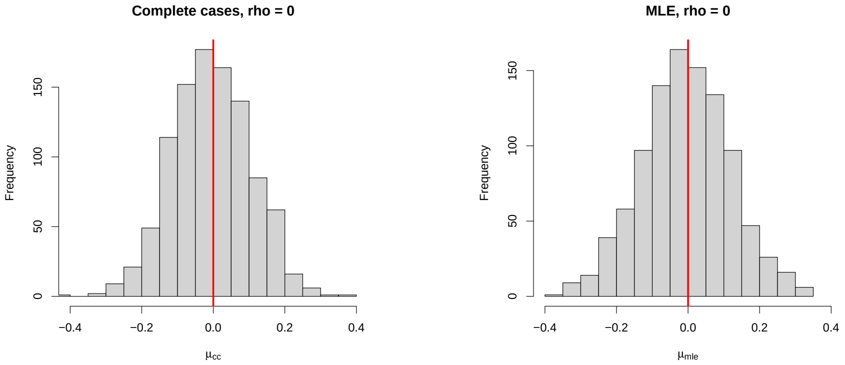
- ↪ We have considered to simulate data that $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and different correlations, namely $\rho \in \{0, 0.5, 0.9\}$ (thus implying $\sigma_{12} \in \{0, 0.5, 0.9\}$).

- ↪ We simulated 1000 datasets, each of size $n = 100$.

- ↪ Induced missingness mechanism is MAR. Specifically, we have considered that $r_i = 0$ (i.e., y_{2i} is missing) if $y_{1i} > z_p$, where $z_p = \Phi^{-1}(a)$, for $a = 0.8$ (distinctness of parameters assumption is also satisfied).

Likelihood based inference with incomplete data

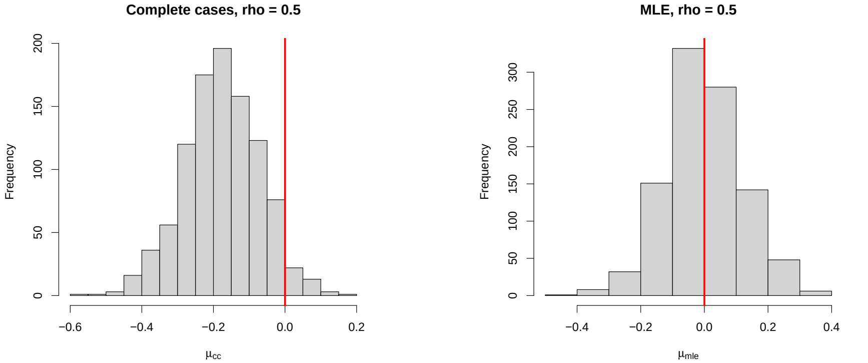
Example: bivariate normal data with one variable subject to missingness



↪ Note that when the correlation between Y_1 and Y_2 is zero, i.e., Y_1 and Y_2 are independent, the mle estimator of μ_2 reduces to the complete cases estimator.

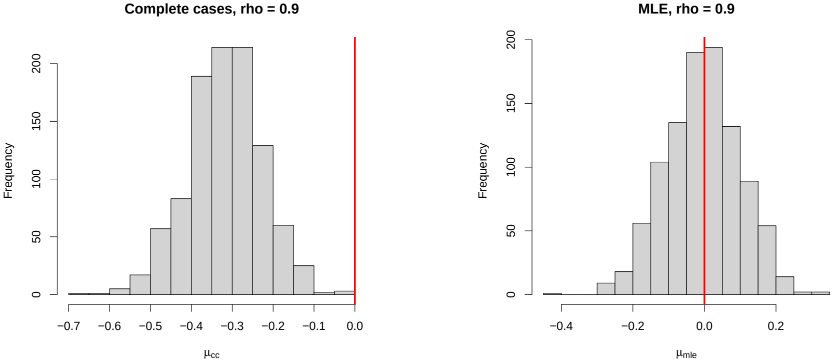
Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness



Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

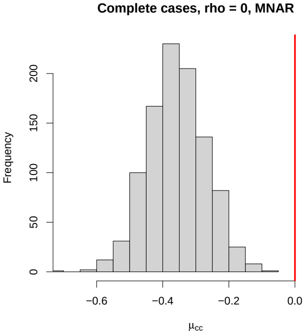


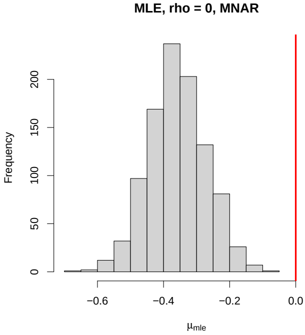
Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

- ↪ We have repeated the same simulation exercise but we now break the assumption of MAR data and simulate MNAR data.

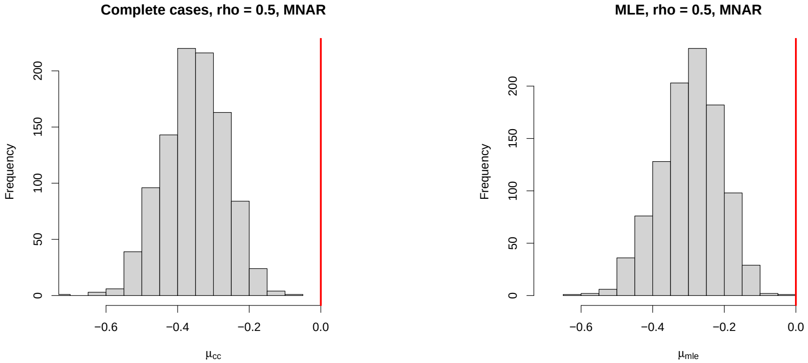
- ↪ Specifically, we have considered that $r_i = 0$ (i.e., y_{2i} is missing) if $y_{2i} > z_p$, where $z_p = \Phi^{-1}(a)$, for $a = 0.8$.





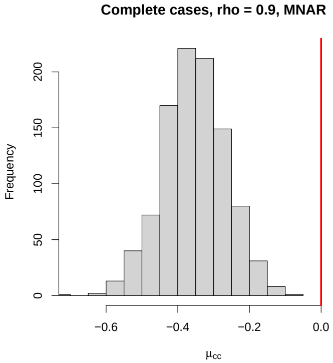
Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness

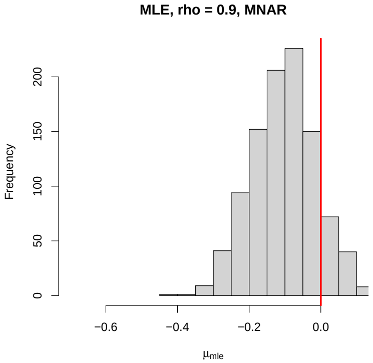


Likelihood based inference with incomplete data

Example: bivariate normal data with one variable subject to missingness



Complete cases, rho = 0.9, MNAR



MLE, rho = 0.9, MNAR

Likelihood based inference with incomplete data

Final remarks

- ↪ Under the ignorability assumption, maximum likelihood estimates can be validly obtained through the observed data likelihood.
- ↪ In this specific example covered here we were able to factorise the observed data likelihood and thus easily obtained the ML estimates.
- ↪ However, general missing data patterns often do not have the particular forms that allow explicit maximum likelihood estimates to be calculated by exploiting factorisations of the likelihood.
- ↪ Furthermore, for some models a factorisation exists, but the parameters in the factorisation are not distinct, and thus maximising the factors separately does not maximise the likelihood.
- ↪ We will now learn about the EM algorithm, an iterative procedure especially designed to deal with incomplete data scenarios.
- ↪ It goes without saying that our ML estimates are only as good as the model we posit for the sampling distribution.