Chapter 2: Likelihood-based approach (Part 2)

A motivating example (Example 2.2)

- Let $t_1, t_2, \cdots, t_n$ be an IID sample from a distribution with density $f(t;\theta)$ for $t > 0$.
- Instead of observing t_i, we observe (y_i, δ_i) where

$$y_i = \begin{cases} t_i & \text{if } \delta_i = 1 \\ c_i & \text{if } \delta_i = 0 \end{cases}$$

  and

$$\delta_i = \begin{cases} 1 & \text{if } t_i \le c_i \\ 0 & \text{if } t_i > c_i, \end{cases}$$

  where c_i is a known censoring time for unit i.
- What is the joint density of (y_i, δ_i)?

- Marginal density is derived from the joint density

$$f(y_i, \delta_i) = \int f(y_i, \delta_i \mid t_i) f(t_i) dt_i$$

- For $\delta_i = 1$: $t_i = y_i$ is observed

$$\begin{aligned} f(y_i, \delta_i = 1) &= \int f(y_i \mid \delta_i = 1, t_i) P(\delta_i = 1 \mid t_i) f(t_i; \theta) dt_i \\ &= \int I(t_i = y_i) P(\delta_i = 1 \mid t_i) f(t_i; \theta) dt_i \\ &= P(\delta_i = 1 \mid y_i) f(y_i; \theta) \\ &= f(y_i; \theta) \end{aligned}$$

- For $\delta_i = 0$: $t_i > c_i = y_i$

$$\begin{aligned} f(y_i, \delta_i = 0) &= \int f(y_i, \delta_i = 0 \mid t_i) f(t_i; \theta) dt_i \\ &= \int I(t_i > c_i) f(t_i; \theta) dt_i \\ &= P(T > c_i; \theta) \end{aligned}$$

Section 2: Observed likelihood

Basic Setup

1. Let $\mathbf{y} = (y_1, y_2, \cdots y_n)$ be a realization of random variable Y with density $f(y)$.
2. Let δ_i be an indicator function defined by
$$\delta_i = \begin{cases} 1 & \text{if } y_i \text{ is observed} \\ 0 & \text{otherwise} \end{cases}$$
with density
$$f_2(\delta \mid y) = \{\pi(y)\}^{\delta}\{1 - \pi(y)\}^{1-\delta}$$
for $\delta = 0, 1$.
3. Thus, instead of observing (δ_i, y_i) directly, we observe $(\delta_i, y_{i,obs})$
$$y_{i,obs} = \begin{cases} y_i & \text{if } \delta_i = 1 \\ * & \text{if } \delta_i = 0. \end{cases}$$
4. What is the (marginal) density of $(y_{i,obs}, \delta_i)$?

Motivation (Change of variable technique or induced probability)

1. Suppose that z is a random variable with density $f_Z(z)$.
2. Instead of observing z directly, we observe only $y = y(z)$, where the mapping $z \mapsto y(z)$ is known.
3. The density of Y is

$$f_Y(y) = \int f_{Y|Z}(y \mid z) f_Z(z)\, d\mu(z),$$

where $f_{Y|Z}(y \mid z) = I\{y = y(z)\}$. That is,

$$f_Y(y) = \int_{\mathcal{R}(y)} f_Z(z)\, d\mu(z)$$

with $\mathcal{R}(y) = \{z; y(z) = y\}$.

Remark

- Note that $P(Y \in B) = P(Z \in y^{-1}(B))$ where $y^{-1}(B) = \{z; y(z) \in B\}$
- Thus, the CDF of Y is

$$\begin{aligned} P(Y \leq y) &= P\{z; y(z) \leq y\} \\ &= \int_{\{z; y(z) \leq y\}} f_Z(z) d\mu(z) \end{aligned}$$

- The pdf is

$$f_Y(y) = \frac{d}{dy} P(Y \leq y) = \int_{\{z; y(z) = y\}} f_Z(z) d\mu(z).$$

Example

- Suppose that Z is a discrete random variable with support $S_z = \{1, 2, 3, 4, 5\}$ with $P(Z = z) = 1/5$ for $z \in S_z$.
- Suppose that the mapping $z \mapsto y\,(z)$ is defined as

$$y(z) = \begin{cases} 1 & \text{if } z \in \{1, 2\} \\ 2 & \text{if } z \in \{3, 4, 5\}. \end{cases}$$

- What is the marginal distribution of Y?

Derivation for the marginal density of (y_{obs}, δ)

1. The joint density for $z = (y, \delta)$ is

$$f(y, \delta) = f_1(y) f_2(\delta \mid y)$$

where $f_1(y)$ is the density of y and $f_2(\delta \mid y)$ is the conditional density of δ conditional on y and is given by $f_2(\delta \mid y) = \{\pi(y)\}^{\delta} \{1 - \pi(y)\}^{1-\delta}$, where $\pi(y) = Pr(\delta = 1 \mid y)$.

2. Instead of observing $z = (y, \delta)$ directly, we observe only $(y_{\mathrm{obs}}, \delta)$ where $y_{\mathrm{obs}} = y_{\mathrm{obs}}(y, \delta)$ and the mapping $(y, \delta) \mapsto y_{\mathrm{obs}}$ is known.

3. The (marginal) density of $(y_{\mathrm{obs}}, \delta)$ is

$$f(y_{\mathrm{obs}}, \delta) = \int P(y_{\mathrm{obs}}, \delta \mid y, \delta) f(y, \delta)\, d\mu(y) = \{f(y_{\mathrm{obs}}, \delta)\}^{\delta} \left\{ \int f(y, \delta) dy \right\}^{1-\delta},$$

where

$$P(y_{\mathrm{obs}}, \delta \mid y, \delta) = \begin{cases} I(y_{\mathrm{obs}} = y) & \text{if } \delta = 1 \\ 1 & \text{if } \delta = 0 \end{cases}$$

Likelihood under missing data (=Observed likelihood)

The observed likelihood is the likelihood obtained from the marginal density of $(y_{i,\mathrm{obs}}, \delta_i)$, $i = 1, 2, \cdots, n$, and can be written as, under the IID setup,

$$
\begin{aligned}
L_{\mathrm{obs}}(\theta) &= \prod_{\delta_i=1} \left[f_1(y_i;\theta) f_2(\delta_i \mid y_i) \right] \times \prod_{\delta_i=0} \left[\int f_1(y_i;\theta) f_2(\delta_i \mid y_i)\, dy_i \right] \\
&= \prod_{\delta_i=1} \left[f_1(y_i;\theta) \pi(y_i) \right] \times \prod_{\delta_i=0} \left[\int f_1(y;\theta) \{1-\pi(y)\}\, dy \right] \\
&= C \prod_{\delta_i=1} f_1(y_i;\theta) \times \prod_{\delta_i=0} \left[\int f_1(y;\theta) \{1-\pi(y)\}\, dy \right]
\end{aligned}
$$

If $\pi(y)$ has an unknown parameter such that $\pi(y) = \pi(y;\phi)$ for some ϕ, the observed likelihood is

$$
L_{\mathrm{obs}}(\theta, \phi) = \prod_{\delta_i=1} \left[f_1(y_i;\theta) \pi(y_i;\phi) \right] \times \prod_{\delta_i=0} \left[\int f_1(y;\theta) \{1-\pi(y;\phi)\}\, dy \right].
$$

Example 2.1 (Censored regression model, or Tobit model)

$$z_i = x_i'\beta + \epsilon_i \qquad \epsilon_i \sim N\left(0, \sigma^2\right)$$

$$y_i = \begin{cases} z_i & \text{if } z_i > 0 \\ 0 & \text{if } z_i \leq 0. \end{cases}$$

The observed log-likelihood is

$$\ell_{\text{obs}}\left(\beta, \sigma^2\right) = -\frac{1}{2}\sum_{y_i > 0}\left[\ln 2\pi + \ln \sigma^2 + \frac{\left(y_i - x_i'\beta\right)^2}{\sigma^2}\right] + \sum_{y_i = 0} \ln\left[1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]$$

where $\Phi(x)$ is the cdf of the standard normal distribution.

Proof

Multivariate extension

Basic Setup

- Let $\mathbf{y} = (y_1, \ldots, y_p)$ be a p-dimensional random vector with probability density function $f(\mathbf{y}; \theta)$.
- Let $\mathbf{y}_1, \cdots, \mathbf{y}_n$ be n independent realizations of $\mathbf{y}$ from $f(\mathbf{y}; \theta)$. (IID sample)
- Let δ_{ij} be the response indicator function of y_{ij} with
$$\delta_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed} \\ 0 & \text{otherwise.} \end{cases}$$
- $\boldsymbol{\delta}_i = (\delta_{i1}, \cdots, \delta_{ip})$: p-dimensional random vector with density $P(\boldsymbol{\delta} \mid \mathbf{y})$ assuming $P(\boldsymbol{\delta}|\mathbf{y}) = P(\boldsymbol{\delta}|\mathbf{y}; \phi)$ for some ϕ.
- Let $(\mathbf{y}_{i,\mathrm{obs}}, \mathbf{y}_{i,\mathrm{mis}})$ be the observed part and missing part of $\mathbf{y}_i$, respectively.

Observed Likelihood for multivariate Y case

- Under IID setup: The observed likelihood is
$$L_{\rm obs}(\theta, \phi) = \prod_{i=1}^{n} \left[\int f(\mathbf{y}_i; \theta) \mathrm{P}(\boldsymbol{\delta}_i | \mathbf{y}_i; \phi) d\mathbf{y}_{i,\rm mis} \right],$$
where it is understood that, if $\mathbf{y}_i = \mathbf{y}_{i,\rm obs}$ and $\mathbf{y}_{i,\rm mis}$ is empty then there is nothing to integrate out.
- In the special case of scalar y, the observed likelihood is
$$L_{\rm obs}(\theta, \phi) = \prod_{\delta_i=1} \left[f(y_i; \theta) \pi(y_i; \phi) \right] \times \prod_{\delta_i=0} \left[\int f(y; \theta) \left\{ 1 - \pi(y; \phi) \right\} dy \right],$$
where $\pi(y; \phi) = \mathrm{P}(\delta = 1 | y; \phi)$.

Definition: Missing At Random (MAR)

$P(\boldsymbol{\delta} \mid \mathbf{y})$ is the density of the conditional distribution of $\boldsymbol{\delta}$ given $\mathbf{y}$. Let $\mathbf{y}_{obs} = \mathbf{y}_{obs}(\mathbf{y}, \boldsymbol{\delta})$ where

$$y_{i,obs} = \begin{cases} y_i & \text{if } \delta_i = 1 \\ * & \text{if } \delta_i = 0. \end{cases}$$

The response mechanism is MAR if $P(\boldsymbol{\delta} \mid \mathbf{y}_1) = P(\boldsymbol{\delta} \mid \mathbf{y}_2)$ $\{$ or $P(\boldsymbol{\delta}|\mathbf{y}) = P(\boldsymbol{\delta} \mid \mathbf{y}_{obs})\}$ for all $\mathbf{y}_1$ and $\mathbf{y}_2$ satisfying $\mathbf{y}_{obs}(\mathbf{y}_1, \boldsymbol{\delta}) = \mathbf{y}_{obs}(\mathbf{y}_2, \boldsymbol{\delta})$.

- MAR: the response mechanism $P(\boldsymbol{\delta} \mid \mathbf{y})$ depends on $\mathbf{y}$ only through $\mathbf{y}_{obs}$.
- Let $\mathbf{y} = (\mathbf{y}_{obs}, \mathbf{y}_{mis})$. By Bayes theorem,

$$P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \boldsymbol{\delta}) = \frac{P(\boldsymbol{\delta} \mid \mathbf{y}_{mis}, \mathbf{y}_{obs})}{P(\boldsymbol{\delta} \mid \mathbf{y}_{obs})} P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}).$$

- MAR: $P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs}, \boldsymbol{\delta}) = P(\mathbf{y}_{mis} \mid \mathbf{y}_{obs})$. That is, $\mathbf{y}_{mis} \perp \boldsymbol{\delta} \mid \mathbf{y}_{obs}$.
- MAR: the conditional independence of $\boldsymbol{\delta}$ and $\mathbf{y}_{mis}$ given $\mathbf{y}_{obs}$.

Remark

- MCAR (Missing Completely at random): $P(\boldsymbol{\delta} \mid \mathbf{y})$ does not depend on $\mathbf{y}$.
- MAR (Missing at random): $P(\boldsymbol{\delta} \mid \mathbf{y}) = P(\boldsymbol{\delta} \mid \mathbf{y}_{obs})$
- NMAR (Not Missing at random): $P(\boldsymbol{\delta} \mid \mathbf{y}) \neq P(\boldsymbol{\delta} \mid \mathbf{y}_{obs})$
- Thus, MCAR is a special case of MAR.

Likelihood factorization theorem

Theorem 2.4 (Rubin, 1976)

$\mathrm{P}_{\phi}(\boldsymbol{\delta}|\mathbf{y})$ is the joint density of $\boldsymbol{\delta}$ given $\mathbf{y}$ and $f_{\theta}(\mathbf{y})$ is the joint density of $\mathbf{y}$. Under conditions

1. the parameters θ and ϕ are distinct and
2. MAR condition holds,

the observed likelihood can be written as

$$L_{obs}(\theta, \phi) = L_1(\theta) L_2(\phi),$$

and the MLE of θ can be obtained by maximizing $L_1(\theta)$.

Thus, we do not have to specify the model for response mechanism. The response mechanism is called ignorable if the above likelihood factorization holds.

Proof

- Bivariate data (x_i, y_i) with pdf $f(x, y) = f_1(y \mid x) f_2(x)$.
- x_i is always observed and y_i is subject to missingness.
- Assume that the response status variable δ_i of y_i satisfies

$$P(\delta_i = 1 \mid x_i, y_i) = \pi(x_i, y_i; \phi)$$

  for some function $\pi(\cdot)$ of known form.
- Let θ be the parameter of interest in the regression model $f_1(y \mid x; \theta)$. Let α be the parameter in the marginal distribution of x, denoted by $f_2(x_i; \alpha)$.
- Three parameters
  - θ: parameter of interest
  - α and ϕ: nuisance parameter

Example 2.3 (Cont'd)

- Observed likelihood

$$
\begin{aligned}
L_{obs}(\theta, \alpha, \phi) &= \left[\prod_{\delta_i = 1} f_1(y_i \mid x_i; \theta) f_2(x_i; \alpha) \pi(x_i, y_i; \phi) \right] \\
&\times \left[\prod_{\delta_i = 0} \int f_1(y \mid x_i; \theta) f_2(x_i; \alpha) \{1 - \pi(x_i, y; \phi)\} dy \right] \\
&= L_1(\theta, \phi) \times L_2(\alpha)
\end{aligned}
$$

  where $L_2(\alpha) = \prod_{i=1}^{n} f_2(x_i; \alpha)$.

- Thus, we can safely ignore the marginal distribution of x if x is completely observed.

Example 2.3 (Cont'd)

- If $\pi(x, y; \phi) = \pi(x; \phi)$ does not depend on y, then MAR holds and

$$L_1(\theta, \phi) = L_{1a}(\theta) \times L_{1b}(\phi)$$

where

$$L_{1a}(\theta) = \prod_{\delta_i = 1} f_1(y_i \mid x_i; \theta)$$

and

$$L_{1b}(\phi) = \prod_{\delta_i = 1} \pi(x_i; \phi) \times \prod_{\delta_i = 0} \{1 - \pi(x_i; \phi)\}$$

- Thus, under MAR, the MLE of θ can be obtained by maximizing $L_{1a}(\theta)$, which is obtained by ignoring the missing part of the data.

Remark (on Example 2.3)

- Instead of y_i subject to missingness, if x_i is subject to missingness, then the observed likelihood becomes

$$\begin{aligned} L_{\text{obs}}(\theta, \phi, \alpha) &= \left[\prod_{\delta_i=1} f_1(y_i \mid x_i; \theta) f_2(x_i; \alpha) \pi(x_i, y_i; \phi)\right] \\ &\times \left[\prod_{\delta_i=0} \int f_1(y_i \mid x; \theta) f_2(x; \alpha) \{1 - \pi(x, y_i; \phi)\} dx\right] \\ &\neq L_1(\theta, \phi) \times L_2(\alpha). \end{aligned}$$

- If $\pi(x, y; \phi)$ does not depend on x, then

$$L_{\text{obs}}(\theta, \alpha, \phi) = L_1(\theta, \alpha) \times L_2(\phi)$$

and MAR holds. Although we are not interested in the marginal distribution of x, we still need to specify the model for the marginal distribution of x.

REFERENCES

Rubin, D. B. (1976), 'Inference and missing data', *Biometrika* **63**, 581–590.