Chapter 2: Likelihood-based approach (Part 3)

Back to motivating example (Example 2.2)

- Let $t_1, t_2, \cdots, t_n$ be an IID sample from a distribution with density $f_T(t) = f(t; \theta)$.
- 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.
- Marginal density of (y_i, δ_i):

$$f(y_i, \delta_i) = \begin{cases} f(y_i; \theta) & \text{if } \delta_i = 1 \\ P(T > c_i; \theta) & \text{if } \delta_i = 0 \end{cases} \tag{1}$$

Justification for (1): $\delta_i = 1$ case

- For any measurable set B,

$$\begin{aligned} P(Y \in B, \delta_i = 1) &= \int P(Y \in B, \delta_i = 1 \mid t_i) f_T(t_i) d t_i \\ &= \int I(t_i \in B) f_T(t_i) d t_i \\ &= \int_B f_T(t_i) d t_i \end{aligned}$$

- By definition, the density f_1 for $(y_i, \delta_i = 1)$ should satisfy

$$P(Y \in B, \delta_i = 1) = \int_B f_1(y) dy$$

- Therefore, since the two terms are equal for any measurable B, we have

$$f_1(y) = f_T(y)$$

for all y almost everywhere.

Justification for (1): $\delta_i = 0$ case

- For any measurable set B, we have

$$P(Y_i \in B, \delta_i = 0) = \begin{cases} P(\delta_i = 0) & \text{if } c_i \in B \\ 0 & \text{otherwise.} \end{cases}$$

- Thus, the marginal density of (y_i, δ_i) for $\delta_i = 0$ is equal to the marginal density of $\delta_i = 0$.
- Now,

$$\begin{aligned} P(\delta_i = 0) &= \int P(\delta_i = 0 \mid t_i) f_T(t_i) dt_i \\ &= \int I(t_i > c_i) f_T(t_i) dt_i \\ &= P(T > c_i). \end{aligned}$$

- If $f_T(t;\theta) = \theta \exp(-\theta t) I(t > 0)$ for $\theta > 0$, then

$$f(y_i, \delta_i) = \begin{cases} \theta \exp(-\theta y_i) & \text{if } \delta_i = 1 \\ \exp(-\theta y_i) & \text{if } \delta_i = 0 \end{cases}$$

- Observed log-likelihood

$$\ell_{\text{obs}}(\theta) = \sum_{i=1}^{n} \delta_i \log(\theta) - \theta \sum_{i=1}^{n} y_i$$

- Observed score function

$$S_{\text{obs}}(\theta) = \frac{\partial}{\partial \theta} \ell_{\text{obs}}(\theta) = \frac{1}{\theta} \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} y_i$$

- The MLE is obtained by solving $S_{\text{obs}}(\theta) = 0$ for θ.

Alternative approach

- Motivation: Wish to find the MLE without computing the marginal density $f(y, \delta)$.
- Can we directly use the score equation for the original observation?

$$S_{\text{com}}(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} t_i$$

- Idea (by R.A. Fisher): Use the conditional expectation of $S_{\text{com}}(\theta)$ given the observed data. Note that

$$E(t_i \mid y_i, \delta_i) = \begin{cases} y_i & \text{if } \delta_i = 1 \\ y_i + 1/\theta & \text{if } \delta_i = 0 \end{cases}$$

- Check:

$$\frac{n}{\theta} - \sum_{i=1}^{n} E(t_i \mid y_i, \delta_i) = \frac{1}{\theta} \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} y_i$$

Motivation

- The observed likelihood is the marginal density of $(\mathbf{y}_{obs}, \boldsymbol{\delta})$.
- The observed likelihood is

$$L_{obs}(\eta) = \int_{\mathcal{R}(\mathbf{y}_{obs}, \boldsymbol{\delta})} f(\mathbf{y}; \theta) \mathrm{P}(\boldsymbol{\delta} | \mathbf{y}; \phi) d\mu(\mathbf{y}) = \int f(\mathbf{y}; \theta) \mathrm{P}(\boldsymbol{\delta} | \mathbf{y}; \phi) d\mu(\mathbf{y}_{mis})$$

  where $\mathbf{y}_{mis}$ is the missing part of $\mathbf{y}$ and $\eta = (\theta, \phi)$.
- Observed score function:

$$S_{obs}(\eta) \equiv \frac{\partial}{\partial \eta} \log L_{obs}(\eta)$$

- Computing the observed score function can be computationally challenging because the observed likelihood is an integral form.

Theorem 2.5: Mean Score Theorem (Fisher, 1922)

Under some regularity conditions, the observed score function equals to the mean score function. That is,

$$S_{obs}(\eta) = E_{\eta}\{S_{com}(\eta)|\mathbf{y}_{obs}, \boldsymbol{\delta}\} := \bar{S}(\eta)$$

where

$$\begin{aligned} S_{com}(\eta) &= \frac{\partial}{\partial \eta} \log f(\mathbf{y}, \boldsymbol{\delta}; \eta), \\ f(\mathbf{y}, \boldsymbol{\delta}; \eta) &= f(\mathbf{y}; \theta) P(\boldsymbol{\delta}|\mathbf{y}; \phi). \end{aligned}$$

- The mean score function is computed by taking the conditional expectation of the complete-sample score function given the observation.
- The mean score function is easier to handle than the observed score function.

Proof of Theorem 2.5

Example 2.5

1. Suppose that the study variable y follows from a normal distribution with mean $\mathbf{x}'\boldsymbol{\beta}$ and variance σ^2. The score equations for $\boldsymbol{\beta}$ and σ^2 under complete response are

$$S_1(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\beta})\, \mathbf{x}_i / \sigma^2 = \mathbf{0}$$

and

$$S_2(\boldsymbol{\beta}, \sigma^2) = -n/(2\sigma^2) + \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 / (2\sigma^4) = 0.$$

2. Assume that y_i are observed only for the first r elements and the MAR assumption holds. In this case, the mean score function reduces to

$$\bar{S}_1(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{r} (y_i - \mathbf{x}_i'\boldsymbol{\beta})\, \mathbf{x}_i / \sigma^2$$

and

$$\bar{S}_2(\boldsymbol{\beta}, \sigma^2) = -n/(2\sigma^2) + \sum_{i=1}^{r} (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2 / (2\sigma^4) + (n - r)/(2\sigma^2).$$

3. The maximum likelihood estimator obtained by solving the mean score equations is

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{r} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{r} \mathbf{x}_i y_i$$

and

$$\hat{\sigma}^2 = \frac{1}{r} \sum_{i=1}^{r} \left(y_i - \mathbf{x}_i' \hat{\boldsymbol{\beta}} \right)^2 .$$

Thus, the resulting estimators can be also obtained by simply ignoring the missing part of the sample, which is consistent with the result in Example 2.3.

- We are interested in estimating θ for the conditional density $f(y \mid x; \theta)$.
- Under MAR, the observed likelihood for θ is

$$L_{obs}(\theta) = \prod_{i=1}^{r} f(y_i \mid x_i; \theta) \times \prod_{i=r+1}^{n} \int f(y \mid x_i; \theta) dy = \prod_{i=1}^{r} f(y_i \mid x_i; \theta).$$

- The same conclusion can follow from the mean score theorem. Under MAR, the mean score function is

$$\begin{aligned} \bar{S}(\theta) &= \sum_{i=1}^{r} S(\theta; x_i, y_i) + \sum_{i=r+1}^{n} E_{\theta}\{S(\theta; x_i, Y) \mid x_i\} \\ &= \sum_{i=1}^{r} S(\theta; x_i, y_i) \end{aligned}$$

where $S(\theta; x, y)$ is the score function for θ and the second equality follows from Theorem 2.3 (Bartlett identity).

Remark: Alternative proof for Theorem 2.5

- Since

$$L_{\text{obs}}(\eta) = f(\mathbf{y}, \boldsymbol{\delta}; \eta) / f(\mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \eta),$$

  we have

$$\frac{\partial}{\partial \eta} \ln L_{\text{obs}}(\eta) = \frac{\partial}{\partial \eta} \ln f(\mathbf{y}, \boldsymbol{\delta}; \eta) - \frac{\partial}{\partial \eta} \ln f(\mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \eta). \tag{2}$$

- Taking conditional expectation of the above equation over the conditional distribution of $(\mathbf{y}, \boldsymbol{\delta})$ given $(\mathbf{y}_{\text{obs}}, \boldsymbol{\delta})$, we have

$$\begin{aligned} \frac{\partial}{\partial \eta} \ln L_{\text{obs}}(\eta) &= E\left\{ \frac{\partial}{\partial \eta} \ln L_{\text{obs}}(\eta) \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta} \right\} \\ &= E\left\{ S_{\text{com}}(\eta) \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta} \right\} - E\left\{ \frac{\partial}{\partial \eta} \ln f(\mathbf{y}_{\text{mis}} \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta}; \eta) \mid \mathbf{y}_{\text{obs}}, \boldsymbol{\delta} \right\}. \end{aligned}$$

- The last term is equal to zero by Theorem 2.3 applied to the conditional distribution, and the reference distribution in this case is the conditional distribution of $\mathbf{y}_{\text{mis}}$ given $(\mathbf{y}_{\text{obs}}, \boldsymbol{\delta})$.

Example 2.4

1. Suppose that the study variable y is randomly distributed with Bernoulli distribution with probability of success p_i, where

$$p_i = p_i(\beta) = \frac{\exp(\mathbf{x}_i'\beta)}{1 + \exp(\mathbf{x}_i'\beta)}$$

for some unknown parameter β and $\mathbf{x}_i$ is a vector of the covariates in the logistic regression model for y_i. We assume that 1 is in the column space of $\mathbf{x}_i$.

2. Under complete response, the score function for β is

$$S_1(\beta) = \sum_{i=1}^{n} (y_i - p_i(\beta))\,\mathbf{x}_i.$$

and the score function for ϕ is

$$S_2(\phi) = \sum_{i=1}^{n} (y_i - \pi_i(\phi))\,(\mathbf{x}_i', y_i)'.$$

Example 2.4 (Cont'd)

3. Let δ_i be the response indicator function for y_i with distribution $Bernoulli(\pi_i)$ where
$$\pi_i = \frac{\exp\left(\mathbf{x}_i'\phi_0 + y_i\phi_1\right)}{1 + \exp\left(\mathbf{x}_i'\phi_0 + y_i\phi_1\right)}.$$
We assume that x_i is always observed, but y_i is missing if $\delta_i = 0$.

4. Under missing data, the mean score function for β is
$$\bar{S}_1(\beta, \phi) = \sum_{\delta_i=1} \{y_i - p_i(\beta)\} \mathbf{x}_i + \sum_{\delta_i=0} \sum_{y=0}^{1} w_i(y; \beta, \phi) \{y - p_i(\beta)\} \mathbf{x}_i, \qquad (3)$$
where $w_i(y; \beta, \phi)$ is the conditional probability of $y_i = y$ given $\mathbf{x}_i$ and $\delta_i = 0$:
$$w_i(y; \beta, \phi) = \frac{P_\beta(y_i = y \mid \mathbf{x}_i) P_\phi(\delta_i = 0 \mid y_i = y, \mathbf{x}_i)}{\sum_{z=0}^{1} P_\beta(y_i = z \mid \mathbf{x}_i) P_\phi(\delta_i = 0 \mid y_i = z, \mathbf{x}_i)} \qquad (4)$$
Thus, $\bar{S}_1(\beta, \phi)$ is also a function of ϕ.

5. If the response mechanism is MAR so that $\phi_1 = 0$, then

$$w_i\left(y;\beta,\phi\right) = \frac{P_\beta\left(y_i = y \mid \mathbf{x}_i\right)}{\sum_{z=0}^{1} P_\beta\left(y_i = z \mid \mathbf{x}_i\right)} = P_\beta\left(y_i = y \mid \mathbf{x}_i\right)$$

and so

$$\bar{S}_1\left(\beta,\phi\right) = \sum_{\delta_i=1}\left\{y_i - p_i\left(\beta\right)\right\}\mathbf{x}_i = \bar{S}_1\left(\beta\right).$$

6. If MAR does not hold, then $(\hat{\beta},\hat{\phi})$ can be obtained by solving $\bar{S}_1\left(\beta,\phi\right) = 0$ and $\bar{S}_2(\beta,\phi) = 0$ jointly, where

$$\begin{aligned}\bar{S}_2\left(\beta,\phi\right) &= \sum_{\delta_i=1}\left\{\delta_i - \pi\left(\phi;\mathbf{x}_i,y_i\right)\right\}\left(\mathbf{x}_i,y_i\right) \\ &\quad + \sum_{\delta_i=0}\sum_{y=0}^{1} w_i\left(y;\beta,\phi\right)\left\{\delta_i - \pi_i\left(\phi;\mathbf{x}_i,y\right)\right\}\left(\mathbf{x}_i,y\right).\end{aligned}$$

Remark (on Example 2.4)

- The mean score function $\bar{S}_1(\beta, \phi)$ in (3) can be expressed as a weighted average of the score function for each possible value of y_i for $\delta_i = 0$.
- The weight function in (4) is a function of unknown parameters. If the weights are known, then the solution to $\bar{S}_1(\beta, \phi) = 0$.
- One way to resolve the problem is to update the weights iteratively using the current parameter values. It is closely related to EM by weighting (Ibrahim, 1990).

REFERENCES

Fisher, R. A. (1922), 'On the mathematical foundations of theoretical statistics', *Philosophical Transactions of the Royal Society of London A* **222**, 309–368.

Ibrahim, J. G. (1990), 'Incomplete data in generalized linear models', *Journal of the American Statistical Association* **85**, 765–769.