(1) Likelihood mode of inference: definitions and results

(i) Reference population and model.

(a) Reference population P = a large group of subjects, $j = 1, ..., M$, with $\{Y_j(1), Y_j(0), X_j\}$.

(b) Parameters = characteristics of the reference population P. Examples:

$$\bar{Y}_1 = \frac{1}{M}\sum_{j=1}^{M} Y_j(1), \quad \tilde{Y}_0 := \text{median}\{Y_j(0) - X_j : j = 1, ..., M\}$$

$$\sigma_1^2 = \frac{1}{M-1}\sum_{j=1}^{M}(Y_j(1) - \bar{Y}_1)^2, \quad \rho = \text{cor}(Y(0), X)$$

We denote by θ all the characteristics of the population.

Goal: to estimate causal effects, e.g. $\bar{Y}_1 - \bar{Y}_0$.
Note: the causal effect is some function of θ.

(c) Model = assumption about the histogram (frequency) of $\{Y_j(1), Y_j(0), X_j\}$ in the reference population.
Examples: "model (1): hist$(Y(1), Y(0)|\theta)$ can be any shape"; " model (2): hist$(Y(1), Y(0)|\theta) = N(\bar{Y}_1, \sigma_1^2)$" means that the histogram of $Y(1)$ in P looks like a normal distribution.

(d) Sampling distribution of simple random sampling with replacement= from P. The histogram of $\{Y_i(1), Y_i(0), X_i\}$, where $i = 1, ...$ are indefinitely many draws of subjects, drawn at random and with replacement from the reference population P (or θ). Notation: pr$(Y_i(1), Y_i(0), X_i|\theta)$.

(e) Note: under (a)-(d) (not just under the examples), the sampling distribution is identical to the histogram in P. So, we will use pr$(\cdot|\theta)$ for both.

(ii) Assumptions for this chapter.

A.1 The assignment mechanism for assigning surgical therapy $(Z = 1)$ or medical therapy $(Z = 0)$ is determined by the observed covariates, in the sense:

$$\text{pr}(Z_i = 1|Y_i(1), Y_i(0), X_i = x) = e(x)$$

Note: this assignment is also called strongly ignorable (Rosenbaum and Rubin, 1983). See why in Section (iv).

A.2. The subjects $i = 1, ..., n$ are a simple random sample from a reference population P as in Sec. 2.1. We say $\{Y_i(1), Y_i(0), X_i\}$ are "iid".

(iii) Results connecting potential outcomes and observed data.

Under assumptions A.1 and A.2 we get the following (Rosenbaum and Rubin (1983)).

(1)

a. $\text{pr}(Y_i(1) = y|X_i = x, \theta) = \text{pr}(Y_i^{obs} = y|X_i = x, Z_i = 1, \theta)$.

b. $\text{pr}(Y_i(0) = y|X_i = x, \theta) = \text{pr}(Y_i^{obs} = y|X_i = x, Z_i = 0, \theta)$.

(2)

a. $\text{pr}(Y_i(1) = y|\theta) = \sum_x \text{pr}(Y_i^{obs} = y|X_i = x, Z_i = 1, \theta) * \text{pr}(X_i = x|\theta)$.

(Note: not equal to $\text{pr}(Y_i^{obs} = y|Z_i = 1, \theta)$).

b. $\text{pr}(Y_i(0) = y|\theta) = \sum_x \text{pr}(Y_i^{obs} = y|X_i = x, Z_i = 0, \theta) * \text{pr}(X_i = x|\theta)$.

(iv) Likelihood of observed data and estimation

A causal effect is a comparison of the values $Y_i(0)$ to the values $Y_i(1)$ in the population, or some subset of it. So any causal effect, say Q, is a function of θ, say $Q(\theta)$. To estimate $Q(\theta)$, we need to estimate θ.

(a) Suppose we assume models (see Sec. (i) part (c)), for example:

$$\text{pr}(Y_i(0) = y|X_i = x, \theta) = f(y, x, \theta_0)$$

$$\text{pr}(Y_i(1) = y|X_i = x, \theta) = g(y, x, \theta_1)$$

(b) Write likelihood of observed data, $\text{pr}(\{Y_i^{obs}, Z_i\}|\{X_i\}, \theta)$.

$$
\begin{aligned}
\prod_i \text{pr}(Y_i^{obs}, Z_i | X_i, \theta) &= \prod_i \text{pr}(Y_i^{obs} | X_i, Z_i, \theta) * \text{pr}(Z_i | X_i, \theta) \\
= \prod_{i:Z_i=1} \text{pr}(Y_i^{obs} | X_i, Z_i = 1, \theta) & \prod_{i:Z_i=0} \text{pr}(Y_i^{obs} | X_i, Z_i = 0, \theta) \prod_i \text{pr}(Z_i | X_i, \theta) \\
= \prod_{i:Z_i=1} \text{pr}(Y_i(1) = y | X_i = x, \theta)_{y=Y_i^{obs}} & \prod_{i:Z_i=0} \text{pr}(Y_i(0) = y | X_i = x, \theta)_{y=Y_i^{obs}} \prod_i \text{pr}(Z_i | X_i, \theta) \\
= \prod_{i:Z_i=1} g(Y_i^{obs}, X_i, \theta_1) & \prod_{i:Z_i=0} f(Y_i^{obs}, X_i, \theta_0) \prod_i e(X_i)^{Z_i} (1 - e(X_i))^{(1-Z_i)}
\end{aligned}
$$

(c) Estimation

We can use the last line to estimate θ, e.g. with $\hat{\theta} =$ as the value of θ that maximizes the likelihood. Then, estimate of causal effect $Q(\theta)$ is $Q(\hat{\theta})$. Note that for such assignment and such mode of inference, the values of the probabilities of assignment can be ignored (hence the term "ignorable assignment").

(v) Examples

for estimating two causal effects of interest: $E(Y_i(1) - Y_i(0)|\theta)$, $E(Y_i(1) - Y_i(0)|X_i = x, \theta)$, with likelihood methods under certain models.

(1) Normal linear regression model with additive effect.

$$
\begin{aligned}
\text{pr}(Y_i(0) = y | X_i = x, \theta) &= N(\alpha_0 + x\beta_0, \sigma^2) \\
\text{pr}(Y_i(1) = y | X_i = x, \theta) &= N(\alpha_0 + x\beta_0 + Q, \sigma^2)
\end{aligned}
$$

Then,

$$
E(Y_i(1) - Y_i(0) | X_i = x) = Q, \text{ and } E(Y_i(1) - Y_i(0)) = Q
$$

and maximum likelihood estimates are obtained from usual least squares.

(2) Normal linear regression model with non additive effect.

$$
\begin{aligned}
\text{pr}(Y_i(0) = y | X_i = x, \theta) &= N(\alpha_0 + x\beta_0, \sigma^2) \\
\text{pr}(Y_i(1) = y | X_i = x, \theta) &= N(\alpha_1 + x\beta_1, \sigma^2)
\end{aligned}
$$

Then

$$E(Y_i(1) - Y_i(0)|X_i = x) = \alpha_1 - \alpha_0 + x(\beta_1 - \beta_0)$$
$$E(Y_i(1) - Y_i(0)) = \alpha_1 - \alpha_0 + (\beta_1 - \beta_0)E(X_i)$$

and maximum likelihood estimates are obtained from usual least squares with an interaction.
Note: a causal effect is not generally a parameter in the model, but a function of the parameters.

Question:
Suppose $Y(z)$ is binary (survival). Write down how to estimate $pr(Y(1) = 1)/pr(Y(0) = 1)$ if we posit a model

$$\text{logitpr}(Y_i(z) = 1|X_i = x, \theta) = \alpha_0 + x\beta_0 + z\,\delta_0. \tag{3}$$

(vi) Comments.
(1) The mode of inference with a likelihood model on the potential outcomes is a suggestive ("deductive") mode of inference: it suggests how to estimate the causal effects (i.e., using the likelihood).

(2) With a small number of covariates, this approach is reliable. With large number of covariates, the consequences of mispecification of the outcome model become more severe.