Causal Inference from Epidemiologic Data

Chapter 6. Studies with nonignorable noncompliance: instrumental variables.

1 Introduction: studies with treatment-noncompliance.

(1) Example.

Sommer and Zeger (1991) Vitamin A data.

(2) Studies we consider.

Randomized initial assignment and subsequent noncompliance.

(3) Problems with standard methods.

Assumptions of "no effect of assignment other than through the actual treatment", when expressed in the standard framework, e.g., using,

$$\text{pr}(Y^{obs} \mid Z, D^{obs}) = \text{pr}(Y^{obs} \mid D^{obs}),$$

fail to reproduce the data.

(4) Potential outcomes and potential receipts.

If person i is **assigned** to $z = 0$ or $z = 1$, then the actual treatment he will **take** is not necessarily the same as Z_i, and is denoted as $D_i(0)$ or $D_i(1)$, respectively. Thus we have four groups of patients:

C_i	$D_i(0)$	$D_i(1)$
never takers(n)	0	0
compliers(c)	0	1
always takers(a)	1	1
defiers(d)	1	0

Observed data: $Z_i, D_i(Z_i) = D_i^{obs}, Y_i(Z_i) = Y_i^{obs}$.

2 Assumptions of instrumental variables with potential outcomes (Angrist, Imbens and Rubin 1996).

(A.1) SUTVA (implicitly assumed).

(A.2) Z_i is ignorable (or randomized)

$$Z_i \coprod (Y_i(0), Y_i(1)), (D_i(0), D_i(1)).$$

(A.3) Monotonicity of compliance (no defiers)

$$D_i(1) \geq D_i(0).$$

(A.4) There are some compliers

$$\text{pr}(C_i = c) > 0.$$

(A.5) Exclusion restriction

$$\text{If for person } i, D_i(1) = D_i(0), \text{then } Y_i(1) = Y_i(0).$$

Note Under the above assumptions, Z is called an "**instrumental variable**".

3 Defining and estimating a causal effect of interest.

(1) Define the "**complier average causal effect (CACE)**" to be a comparison between

$$\{Y_i(0) : D_i(0) = 0 \text{ and } D_i(1) = 1\} \quad \text{and} \quad \{Y_i(1) : D_i(0) = 0 \text{ and } D_i(1) = 1\}$$

Why **CACE**?

(a) CACE is a well defined causal effect, because it compares the same people's potential outcomes when assigned to $z = 1$, and having $D_i(1) = 1$, to the potential outcomes when assigned to $z = 0$, and having $D_i(0) = 0$.

(b) CACE informs about biological mechanisms better than ITT effect.

(c) CACE leads to a useful template for observational studies.

(2) Estimation.

Case 1. Y_i is binary. Parameterize the problem as follows

$\pi_c = \text{pr}(C_i = c), \quad \pi_n = \text{pr}(C_i = n), \quad \pi_a = 1 - \pi_c - \pi_n$;

$B_a = \text{pr}(Y_i(z) = 1 \mid C_i = a)$;

$B_n = \text{pr}(Y_i(z) = 1 \mid C_i = n)$;

$B_{c,z} = \text{pr}(Y_i(z) = 1 \mid C_i = c)$.

(By exclusion restriction, B_a, B_n is not a function of z.)

We can directly estimate the above components π_n, π_a, B_a, B_n using the relations

$$\text{pr}(D_i^{obs} = 1 \mid Z_i = 0) = \pi_a, \quad \text{pr}(D_i^{obs} = 0 \mid Z_i = 1) = \pi_n, \quad \pi_c = 1 - \pi_n - \pi_a,$$

$$\text{pr}(Y_i^{obs} = 1 \mid D_i^{obs} = 1, Z_i = 0) = B_a, \quad \text{pr}(Y_i^{obs} = 1 \mid D_i^{obs} = 0, Z_i = 1) = B_n.$$

Now define

$$\tau_{1,1} = \text{pr}(Y_i^{obs} = 1 \mid D_i^{obs} = 1, Z_i = 1), \quad \tau_{0,0} = \text{pr}(Y_i^{obs} = 1 \mid D_i^{obs} = 0, Z_i = 0).$$

Then B_{c1} and B_{c0} can be estimated using the relations

$$B_{c1} = \frac{(\pi_a + \pi_c)\tau_{1,1} - \pi_a B_a}{\pi_c},$$

$$B_{c0} = \frac{(\pi_n + \pi_c)\tau_{0,0} - \pi_n B_n}{\pi_c}.$$

Case 2. Y_i is not binary. Define $Y_i^* = 1$ if $Y_i < c_0$, where c_0 is any value in the range of Y. The full distribution of $Y(z)$ is then recovered by using the same argument as for binary case (case 1) for all values of c_0.

Note 1. The identifiability of causal effects for binary Y extends to any distribution of uncensored Y under the above assumptions.

Note 2. If CACE is $E(Y_i(1) - Y_i(0)) \mid C_i = \text{complier})$, then it equals $\frac{E(Y_i(1) - Y_i(0))}{E(D_i(1) - D_i(0))}$, i.e., ratio of the linear ITT effect on the outcome divided by the linear ITT effect on the taking of the treatment. This is **not true** generally for other contrasts of $E(Y_i(1))$ and $E(Y_i(0))$.

Note 3. The above was first shown for potential outcomes by Angrist, Imbens and Rubin (1996).

(3) Covariate-treatment interaction.

(a) With few covariates we may be able to stratify.

(b) With continuous covariates, likelihood mode forms again a deductive way of modelling

(1) pr$(C_i = s \mid X_i = x) = p(s, x, \gamma), \quad s = c, a,$ or n (e.g., by a multinomial regression);

(2) pr$(Y_i(z) = y \mid C_i = s, X_i = x) = f(y, z, s, x, \beta)$ (e.g., by a logistic regression).

Then,

- The likelihood of all the data is the product of the individual likelihoods.
- Estimation of the parameter β, γ and thus of the causal effect can be either through maximizing the likelihood (e.g., using EM algorithm) or Bayesian methods (e.g., using data augmentation).
- Both of these approaches, in their own way, cycle between estimating (β, γ) given the data, and then imputing the unobserved compliance strata given the estimated parameters. More details are provided in Imbens and Rubin (1997).