REGRESSION DISCONTINUITY

The regression discontinuity (RD) design arises when treatment depends on thresholding a continuous score. We denote the score Z_i, treatment M_i, and the outcome Y_i.

RD designs were pioneered by Thislethwaite and Campbell nearly 60 years ago in an educational study to evaluate the effect of receiving an award on subsequent outcomes such as receiving scholarships. The receipt of an award was based solely on a test score Z_i.

Thislethwaite and Campbell reasoned that students just above the cutoff and students just below could be compared, and the effect of an award could be estimated around the cutoff value z_0.

Other examples:

1. in medicine, people who have cholesterol levels above a certain number may be prescribed a pill.
2. a surgery may be performed only when the value of a critical indicator has been passed.
3. in urban planning, only streets where one or more fatalities have occurred in the past year may see a change in the speed limit.

RD designs are also popular in the social sciences, and have been used and studied extensively, especially by econometricians. Readers who wish to read about applications in economics may wish to look at the review paper by Lee and Lemieux (2010). The Journal of Econometrics devoted a special issue to this topic in 2008, as did the journal Observational Studies in 2016.

At first, it might seem odd that anything can be learned from a design where there is no overlap in scores between units receiving and not receiving treatment.

We have so far focused on the unconfoundedness condition, noting that problems ensued when it was not possible to find observations in the treatment and control groups with similar values of the covariates or propensity score.

The question is what we can learn when the unconfoundedness condition holds, but the overlap assumption may not. Clearly, one cannot, at least without making a lot of possibly unjustified assumptions, learn about treatment effects far from the cutoff, so we shall focus on the neighborhood around the cutoff.

REGRESSION DISCONTINUITY

The following exposition parallels the treatment in Hahn, Todd and van der Klaauw (2001).

Suppose for all i, the score Z_i has cutoff z_0. Let treatment $M(Z_i)$ takes values 0 or 1, and denote the potential outcomes $Y_i(0)$ and $Y_i(1)$. The observed outcome is $Y_i(M(Z_i))$, and the estimand of interest is $E(Y_i(1) - Y_i(0) \mid Z_i = z_0)$.

We distinguish between two types of RD designs: "sharp" and "fuzzy" designs. In a "sharp"' design, treatment depends deterministically on Z_i. $M(Z_i) = 1$ if $Z_i > z_0$ (or $\geq z_0$), 0 otherwise. Treatment is unconfounded, i.e. $\Pr(M(Z_i) = 1 \mid Y_i(0), Y_i(1), Z_i) = \Pr(M(Z_i) = 1 \mid Z_i)$. However, above the threshold, there are no control units to compare with treated units, and below the cutoff, there are no treated units to compare with the controls.

In a "fuzzy" design, assignment is probabilistic $\Pr(M(Z_i) = 1 \mid Z_i) \in (0, 1)$. An important assumption for fuzzy designs is that there is a discontinuity in the assignment process at the cutoff. i.e. units just above and below the cutoff have different probabilities of receiving treatment; for the sharp design, this is clearly satisfied. Hahn et al. (2001) call this the regression discontinuity

assumption:

$$m^- = \lim_{z \uparrow z_0} \Pr(M(Z_i) = 1 \mid Z_i = z) \neq m^+ = \lim_{z \downarrow z_0} \Pr(M(Z_i) = 1 \mid Z_i = z).$$

Using this assumption, we compare units just above and below the cutoff: $E(Y_i \mid Z_i = z_0 + \varepsilon) - E(Y_i \mid Z_i = z_0 - \varepsilon)$. Letting $Y_i(1) - Y_i(0) = \beta_i$,

$$E(Y_i \mid Z_i = z_0 + \varepsilon) = E(Y_i(1) \mid Z_i = z_0 + \varepsilon) \pm E(Y_i(0) \mid Z_i = z_0 + \varepsilon) =$$
$$E(\beta_i \mid Z_i = z_0 + \varepsilon, M(Z_i) = 1) \Pr(M(Z_i) = 1 \mid Z_i = z_0 + \varepsilon) + E(Y_i(0) \mid Z_i = z_0 + \varepsilon$$

A similar decomposition for $E(Y_i \mid Z_i = z_0 - \varepsilon)$ leads to

$$\begin{aligned}
&E(Y_i \mid Z_i = z_0 + \varepsilon) - E(Y_i \mid Z_i = z_0 - \varepsilon) = \\
&\quad \{E(\beta_i \mid Z_i = z_0 + \varepsilon, M(Z_i) = 1) \Pr(M(Z_i) = 1 \mid Z_i = z_0 + \varepsilon) \\
&\qquad\qquad - E(\beta_i \mid Z_i = z_0 - \varepsilon, M(Z_i) = 1) \Pr(M(Z_i) = 1 \mid Z_i = z_0 - \varepsilon)\} \\
&\quad + \{E(Y_i(0) \mid Z_i = z_0 + \varepsilon) - E(Y_i(0) \mid Z_i = z_0 - \varepsilon)\}.
\end{aligned}$$

It seems reasonable to assume that units just to the left and right of the cutoff would have similar outcomes in the absence of treatment:

$$\lim_{\varepsilon \downarrow 0} E(Y_i(0) \mid Z_i = z_0 + \varepsilon) = \lim_{\varepsilon \uparrow 0} E(Y_i(0) \mid Z_i = z_0 - \varepsilon).$$

REGRESSION DISCONTINUITY

For the terms in

$$
\begin{aligned}
&E(Y_i \mid Z_i = z_0 + \varepsilon) - E(Y_i \mid Z_i = z_0 - \varepsilon) = \\
&\quad \{E(\beta_i \mid Z_i = z_0 + \varepsilon, M(Z_i) = 1) \Pr(M(Z_i) = 1 \mid Z_i = z_0 + \varepsilon) \\
&\qquad\qquad - E(\beta_i \mid Z_i = z_0 - \varepsilon, M(Z_i) = 1) \Pr(M(Z_i) = 1 \mid Z_i = z_0 - \varepsilon)\} \\
&\quad + \{E(Y_i(0) \mid Z_i = z_0 + \varepsilon) - E(Y_i(0) \mid Z_i = z_0 - \varepsilon)\}.
\end{aligned}
$$

involving β_i, note that if the treatment effects are constant, i.e. $\beta_i = \beta$ for all i, β can be pulled out of the expectation. β is then identified and equal to $\frac{y^+ - y^-}{m^+ - m^-}$, where $y^+ = \lim_{\varepsilon \downarrow 0} E(Y_i \mid Z_i > z_0 + \varepsilon)$,
$y^- = \lim_{\varepsilon \uparrow 0} E(Y_i \mid Z_i < z_0 - \varepsilon)$.

Although the constant effect assumption is strong, it is worth noting that the result above is obtained without assuming that treatment is unconfounded for the fuzzy design.

However, in the more interesting and realistic case where treatment effects are heterogeneous, assumptions about unconfoundedness are required. If we assume unconfoundedness:

$$Y_i(1) - Y_i(0) \perp\!\!\!\perp M(Z_i) \mid Z_i,$$

$$E(\beta_i \mid Z_i = z_0 \pm \varepsilon, M(Z_i) = 1) = E(\beta_i \mid Z_i = z_0 \pm \varepsilon).$$

Again, the continuity assumption, $E(\beta_i \mid Z_i = z)$ is continuous at z_0, implies the regression discontinuity assumption and the unconfoundedness assumption, $E(Y_i(1) - Y_i(0) \mid Z_i = z_0)$ is identified and equal to $\frac{y^+ - y^-}{m^+ - m^-}$.

Finally, Hahn et al. (2001) consider the case where Z plays the role of an instrument. Unlike Angrist et al. (1996), the instrument is continuous.

Using the continuity assumption, the regression discontinuity assumption, the unconfoundedness assumption,

$$Y_i(1) - Y_i(0), \{M(z) : z \in z_0 \pm \varepsilon\} \perp\!\!\!\perp Z_i,$$

and the additional monotonicity assumption $M_i(z_0 + \varepsilon) \geq M_i(z_0 - \varepsilon)$ in a nieghborhood of z_0, similar to the monotonicity assumption in Angrist et al. (1996), the authors show that $\frac{y^+ - y^-}{m^+ - m^-}$ is the complier average causal effect at the cutoff z_0. i.e. the ATE for subjects who would not take treatment at $z_0 - \varepsilon$, but who would take treatment at $z_0 + \varepsilon$.

To estimate the effects considered above, the authors suggest using a locally linear nonparametric regression estimator. See Hahn et al. (2001).