Taber’s Thoughts on “Randomization Tests under an Approximate Symmetry Assumption” by Canay, Romano, and Shaikh

September 27, 2014
Divide data into $q$ different clusters $j = 1, \ldots, q$ so we have independence with the other groups (and $q$ doesn’t grow with sample size)

For each cluster estimate the parameter $\hat{\theta}_j$

Under the null that $\theta = \theta_0$,

$$
\sqrt{N} \left( \hat{\theta}_1 - \theta_0, \ldots, \hat{\theta}_q - \theta_0 \right) \sim N \left( 0, \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_q \end{bmatrix} \right)
$$

Notice as well that for any $g = (g_1, \ldots, g_q)$ with $g_j \in \{-1, 1\}$,

$$
\sqrt{N} \left( g_1 \left( \hat{\theta}_1 - \theta_0 \right), \ldots, g_q \left( \hat{\theta}_q - \theta_0 \right) \right) \sim N \left( 0, \begin{bmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_q \end{bmatrix} \right)
$$
Thus under the null hypothesis

$$t(1) = T\left( \sqrt{N} \left( \hat{\theta}_1 - \theta_0, ..., \hat{\theta}_q - \theta_0 \right) \right)$$

has the same distribution as

$$t(g) = T\left( \sqrt{N} \left( g_1 \left( \hat{\theta}_1 - \theta_0 \right), ..., g_q \left( \hat{\theta}_q - \theta_0 \right) \right) \right)$$

for any other $g = (g_1, ..., g_q)$

We then compare $t(1)$ to all other possible $t(g)$

1. If it is in the upper $\alpha$ tail we reject the null
2. otherwise we accept it
I have little to add to the formal results—but I guess that is not the point of the interactions conference.

Intuitively this makes a lot of sense and seems like a really cool idea—if \( \hat{\theta} \) is consistently higher than \( \theta_0 \) we will reject the null.

I can see why it might be very convincing.
However, there is still some level at which it's a bit odd and more guidance to empirical researchers on when to use it would be helpful.

It is kind of both asymptotic and small sample at the same time.

- **Small Samples**
  - It feels like a small sample test and works with small samples if we have symmetry.
  - However, if samples aren’t large why would we expect the symmetry assumption to hold?

- **If samples are large**
  - Why not just use standard asymptotic theory?
  - Normals are symmetric but they are also have a lot of other properties. Why focus on symmetry?
To me the nicest examples of when to use this combine ideas of both small and large samples.

There is one point the paper that is in the background that could be clearer.

In general there are two reasons we might expect symmetry—one is if sample sizes are large.

The other is if

- treatment is randomly assigned
- \( i \) is a random person who receives the treatment and \( k(i) \) is a random person that doesn’t who is matched to \( i \).

Then

\[
Y_i - Y_{k(i)}
\]

will be symmetric.
An interesting case is the clustered regression like the Angrist and Lavy paper when we have a finite number of groups and a large number of people per group and treatment is assigned per group.

I want to go a little beyond them and write (and try to stick to their notation)

\[ Y_i = \alpha T_j(i) + \varepsilon_i \]
\[ Y_i = \alpha T_j(i) + \eta_j(i) + \tilde{\varepsilon}_i \]

where

\[ \eta_j(i) = E(\varepsilon_i \mid j(i) = j) \]
\[ \tilde{\varepsilon}_i = \varepsilon_i - \eta_j(i) \]
Then if I form my q clusters by matching one treatment group with one control group. Number the treatments as \( j = 1, \ldots, J_1 \) and let \( k(j) \) be the control matched to them then

\[
\hat{\alpha}_j = \alpha_0 + \eta_j - \eta_{k(j)} + \frac{1}{N_j} \sum_{\{i:j(i)=j\}} \varepsilon_i - \frac{1}{N_{k(j)}} \sum_{\{i:j(i)=k(j)\}} \varepsilon_i
\]

Central limit theorem implies that second two terms are approximately symmetric

Random assignment of T to j implies that \( \eta_j - \eta_{k(j)} \) is symmetric
However this won’t work with “one to many” matching, now $k(j)$ is a set

$$
\hat{\alpha}_j = \alpha_0 + \eta_j - \bar{\eta}_{k(j)} + \frac{1}{N_j} \sum_{\{i:j(i)=j\}} \varepsilon_i - \frac{1}{N_k(j)} \sum_{\{i:j(i)\in k(j)\}} \varepsilon_i
$$

where $\bar{\eta}_{k(j)}$ is weighted average of control clusters

There is no reason to believe $\eta_j - \bar{\eta}_{k(j)}$ will be symmetric

Also if the hypothesis is not $\alpha_0 = 0$ but some other $\alpha_0 = 1$ then its less clear that I believe the symmetry assumption (takes homogeneous treatment effects very seriously)
Cool paper—and Monte Carlo seems to work well

I would still like to get a sense of exactly how to think of the symmetry and when this is a reasonable assumption