ATE & Variance Estimators for the Case with Multiple Treatments and Linear Adjustments

Juri Trifonov

June 12, 2024

1 Individual-Level Treatment Assignment

Notation

Let us introduce some additional notations besides the notation used in (Jiang et al., 2023).

$$
\begin{aligned}
N_a &:= \sum_{i\in[N]} \mathbb{I}\{A_i = a\},\ \forall a \in \mathcal{A} \\
N_0 &:= \sum_{i\in[N]} \mathbb{I}\{A_i = 0\} \\
\hat{\pi}_a(s) &:= \frac{N_a(s)}{N(s)},\ \hat{\pi}_0(s) := \frac{N_0(s)}{N(s)},\ \forall a \in \mathcal{A} \text{ and } \forall s \in \mathcal{S} \\
\tilde{A}_i &= \begin{cases} 1, \text{ if } A_i = a,\ \forall a \in \mathcal{A} \\ 0, \text{ if } A_i = 0. \end{cases}
\end{aligned}
$$

ATE Estimator

Let us recall the version for a binary treatment case:

$$
\hat{\tau} \equiv \frac{1}{N}\sum_{i=1}^{N}\left[\frac{A_i(Y_i - \hat{\mu}(1, S_i, X_i))}{\hat{\pi}(S_i)} - \frac{(1-A_i)(Y_i - \hat{\mu}(0, S_i, X_i))}{1-\hat{\pi}(S_i)} + \hat{\mu}(1, S_i, X_i) - \hat{\mu}(0, S_i, X_i)\right].
$$

Since in the case with multiple treatments $\hat{\pi}_a(s) + \hat{\pi}_0(s) \neq 1$, we should make the following corrections to the expression. First, instead of $1 - \hat{\pi}(s)$ we should directly use $\hat{\pi}_0 = \frac{n_0(s)}{n(s)}$. Second, instead of a binary treatment variable A_i, we shouls use an indicator variable $\tilde{A}_i = \mathbb{I}\{A_i = 1\}$. Thus, the estimator for any treatment $a \in \mathcal{A}$ (relative to control) becomes:

$$
\hat{\tau}_a \equiv \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\tilde{A}_i(Y_i - \hat{\mu}(a, S_i, X_i))}{\hat{\pi}_a(S_i)} - \frac{(1-\tilde{A}_i)(Y_i - \hat{\mu}(0, S_i, X_i))}{\hat{\pi}_0(S_i)} + \hat{\mu}(a, S_i, X_i) - \hat{\mu}(0, S_i, X_i)\right],
$$

where

$$
\tilde{A}_i = \begin{cases} 1, \text{ if } A_i = a,\ \forall a \in \mathcal{A} \\ 0, \text{ if } A_i = 0. \end{cases}
$$

Variance Estimator

Recall that the expression of the variance estimator for the binary treatment case has the following form:

$$
\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\left[A_i\hat{\Xi}_1^2(\mathcal{D}_i, S_i) + (1-A_i)\hat{\Xi}_0^2(\mathcal{D}_i, S_i) + \hat{\Xi}_2^2(\mathcal{D}_i, S_i)\right],
$$

where

$$\hat{\Xi}_1(\mathcal{D}_i, s) = \tilde{\Xi}_1(s) - \frac{1}{N_1(s)} \sum_{j \in I_1(s)} \tilde{\Xi}_{1,j}(\mathcal{D}_i, s),$$

$$\hat{\Xi}_0(\mathcal{D}_i, s) = \tilde{\Xi}_0(s) - \frac{1}{N_0(s)} \sum_{j \in I_0(s)} \tilde{\Xi}_{0,j}(\mathcal{D}_i, s),$$

$$\hat{\Xi}_2 = \left(\frac{1}{N_1(s)} \sum_{j \in I_1(s)} (Y_j - \hat{\tau} A_j) \right) - \left(\frac{1}{N_0(s)} \sum_{j \in I_0(s)} (Y_j - \hat{\tau} A_j) \right)$$

$$\tilde{\Xi}_1(\mathcal{D}_i, s) = \left(1 - \frac{1}{\hat{\pi}(s)} \right) \hat{\mu}(1, s, X_i) - \hat{\mu}(0, s, X_i) + \frac{Y_i}{\hat{\pi}(s)},$$

$$\tilde{\Xi}_0(\mathcal{D}_i, s) = \left(\frac{1}{1 - \hat{\pi}(s)} - 1 \right) \hat{\mu}(0, s, X_i) + \hat{\mu}(1, s, X_i) - \frac{Y_i}{1 - \hat{\pi}(s)}.$$

Following the same reasoning as for the ATE estimator, we need to make the following corrections to get the expression for the multiple treatments case.

First, notice that we can rewrite $\tilde{\Xi}_1(\mathcal{D}_i, s)$ and $\tilde{\Xi}_1(\mathcal{D}_i, s)$ as follows:

$$\tilde{\Xi}_1(\mathcal{D}_i, s) = \hat{\mu}(1, s, X_i) - \hat{\mu}(0, s, X_i) + \frac{Y_i - \hat{\mu}(1, s, X_i)}{\hat{\pi}(s)},$$

$$\tilde{\Xi}_0(\mathcal{D}_i, s) = \hat{\mu}(1, s, X_i) - \hat{\mu}(0, s, X_i) - \frac{Y_i - \hat{\mu}(0, s, X_i)}{1 - \hat{\pi}(s)}.$$

For the same reasoning as before, to generalize these terms for the multiple treatments scenario we need to substitute $\hat{\pi}_a(s)$ and $\hat{\pi}_0(s)$ for $\hat{\pi}(s)$ and $1 - \hat{\pi}(s)$ respectively.

Combining the corrections, we get the following expression of the variance estimator for every treatment $a \in \mathcal{A}$:

$$\hat{\sigma}_a^2 = \frac{1}{N} \sum_{i=1}^{N} \left[\tilde{A}_i \hat{\Xi}_a^2(\mathcal{D}_i, S_i) + (1 - \tilde{A}_i) \hat{\Xi}_0^2(\mathcal{D}_i, S_i) + \hat{\Xi}_2^2(\mathcal{D}_i, S_i) \right],$$

where

$$\hat{\Xi}_a(\mathcal{D}_i, s) = \tilde{\Xi}_a(s) - \frac{1}{N_a(s)} \sum_{j \in I_a(s)} \tilde{\Xi}_{a,j}(\mathcal{D}_i, s),$$

$$\hat{\Xi}_0(\mathcal{D}_i, s) = \tilde{\Xi}_0(s) - \frac{1}{N_0(s)} \sum_{j \in I_0(s)} \tilde{\Xi}_{0,j}(\mathcal{D}_i, s),$$

$$\hat{\Xi}_2 = \left(\frac{1}{N_a(s)} \sum_{j \in I_a(s)} (Y_j - \hat{\tau}_a \tilde{A}_j) \right) - \left(\frac{1}{N_0(s)} \sum_{j \in I_0(s)} (Y_j - \hat{\tau}_a \tilde{A}_j) \right),$$

$$\tilde{\Xi}_a(\mathcal{D}_i, s) = \hat{\mu}(a, s, X_i) - \hat{\mu}(0, s, X_i) + \frac{Y_i - \hat{\mu}(a, s, X_i)}{\hat{\pi}_a(s)},$$

$$\tilde{\Xi}_0(\mathcal{D}_i, s) = \hat{\mu}(a, s, X_i) - \hat{\mu}(0, s, X_i) - \frac{Y_i - \hat{\mu}(0, s, X_i)}{\hat{\pi}_0(s)}.$$

2 Cluster-Level Treatment Assignment

Using the same reasoning as before, we can generalize the ATE and variance estimators for the case with clusters and multiple treatments as follows.

ATE Estimator

$$
\begin{aligned}
\hat{\tau}_a &= \frac{\sum_{g=1}^{G} \hat{\Xi}_g}{\sum_{g=1}^{G} N_g}, \\
\hat{\Xi}_g &= \frac{\tilde{A}_g \left(N_g \bar{Y}_g - \hat{\mu}(a, S_g, X_g, N_g) \right)}{\hat{\pi}_a(S_g)} - \frac{(1 - \tilde{A}_g) \left(N_g \bar{Y}_g - \hat{\mu}(0, S_g, X_g, N_g) \right)}{\hat{\pi}_0(S_g)} \\
&\quad + \hat{\mu}(a, S_g, X_g, N_g) - \hat{\mu}(0, S_g, X_g, N_g),
\end{aligned}
$$

where

$$
\tilde{A}_i = \begin{cases} 1, \text{ if } A_i = a, \ \forall a \in \mathcal{A} \\ 0, \text{ if } A_i = 0. \end{cases}
$$

Variance Estimator

$$
\hat{\sigma}_a^2 = \frac{\frac{1}{G} \sum_{g=1}^{G} \left[\tilde{A}_g \hat{\Xi}_a^2(\mathcal{D}_g, S_g) + (1 - \tilde{A}_g) \hat{\Xi}_0^2(\mathcal{D}_g, S_g) + \hat{\Xi}_2^2(\mathcal{D}_g, S_g) \right]}{\left(\frac{1}{G} \sum_{g=1} N_g \right)^2},
$$

where

$$
\begin{aligned}
\hat{\Xi}_a(\mathcal{D}_g, s) &= \tilde{\Xi}_a(s) - \frac{1}{G_1(s)} \sum_{j \in I_1(s)} \tilde{\Xi}_{1,j}(s) - \hat{\tau}_a \left(N_g - \frac{1}{G(s)} \sum_{j \in I(s)} N_j \right), \\
\hat{\Xi}_0(\mathcal{D}_g, s) &= \tilde{\Xi}_0(s) - \frac{1}{G_0(s)} \sum_{j \in I_0(s)} \tilde{\Xi}_{0,j}(s) - \hat{\tau}_a \left(N_g - \frac{1}{G(s)} \sum_{j \in I(s)} N_j \right), \\
\hat{\Xi}_2(s) &= \left(\frac{1}{G_1(s)} \sum_{j \in I_1(s)} N_j \bar{Y}_j \right) - \left(\frac{1}{G_0(s)} \sum_{j \in I_0(s)} N_j \bar{Y}_j \right) - \hat{\tau}_a \times \left(\frac{1}{G(s)} \sum_{j \in I(s)} N_j \right), \\
\tilde{\Xi}_a(\mathcal{D}_g, s) &= \hat{\mu}(a, s, X_g, N_g) - \hat{\mu}(0, s, X_g, N_g) + \frac{N_g \bar{Y}_g - \hat{\mu}(a, s, X_g, N_g)}{\hat{\pi}_a(s)}, \\
\tilde{\Xi}_0(\mathcal{D}_g, s) &= \hat{\mu}(a, s, X_g, N_g) - \hat{\mu}(0, s, X_g, N_g) - \frac{N_g \bar{Y}_g - \hat{\mu}(0, s, X_g, N_g)}{\hat{\pi}_0(s)}.
\end{aligned}
$$

References

Jiang, L., Linton, O., Tang, H. and Zhang, Y. (2023). Improving estimation efficiency via regression-adjustment in covariate-adaptive randomizations with imperfect compliance. Working paper. Forthcoming in Review of Economics and Statistics. Available at https://arxiv.org/abs/2201.13004.