

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

Juri Trifonov

February 28, 2024

1 CAR without Clusters

Notation

Let us introduce some additional notations besides the notation used in (Juang et al., 2017).

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

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, in the original expression, the last two terms imply that $\hat{\pi}(s) + (1 - \hat{\pi}(s)) = 1$ and that's why they are not normalized, i.e. $\frac{\hat{\mu}(1, S_i, X_i) - \hat{\mu}(0, S_i, X_i)}{\hat{\pi}(S_i) + (1 - \hat{\pi}(S_i))} = \hat{\mu}(1, S_i, X_i) - \hat{\mu}(0, S_i, X_i)$. Since for the multiple treatment setup the last equality does not hold, we need to divide the last two terms by $\hat{\pi}_a(s) + \hat{\pi}_0(s)$. Thus, the estimator for any treatment $a \in \mathcal{A}$ (relative to control) becomes:

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

where

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

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

$$\begin{aligned} \hat{\Xi}_1(\mathcal{D}_i, s) &= \tilde{\Xi}_1(s) - \frac{1}{N_1(s)} \sum_{j \in I_1(s)} \tilde{\Xi}_{1,j}(\mathcal{D}_i, s), \\ \hat{\Xi}_0(\mathcal{D}_i, s) &= \tilde{\Xi}_0(s) - \frac{1}{N_0(s)} \sum_{j \in I_0(s)} \tilde{\Xi}_{0,j}(\mathcal{D}_i, s), \\ \hat{\Xi}_2 &= \left(\frac{1}{N_1(s)} \sum_{j \in I_1(s)} (Y_j - \hat{\tau} A_j) \right) - \left(\frac{1}{N_0(s)} \sum_{j \in I_0(s)} (Y_j - \hat{\tau} A_j) \right) \\ \tilde{\Xi}_1(\mathcal{D}_i, s) &= \left(1 - \frac{1}{\hat{\pi}(s)} \right) \hat{\mu}(1, s, X_i) - \hat{\mu}(0, s, X_i) + \frac{Y_i}{\hat{\pi}(s)}, \\ \tilde{\Xi}_0(\mathcal{D}_i, s) &= \left(\frac{1}{1 - \hat{\pi}(s)} - 1 \right) \hat{\mu}(0, s, X_i) + \hat{\mu}(1, s, X_i) - \frac{Y_i}{1 - \hat{\pi}(s)}. \end{aligned}$$

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 $\tilde{\Xi}_0(\mathcal{D}_i, s)$ as follows:

$$\begin{aligned} \tilde{\Xi}_1(\mathcal{D}_i, s) &= \hat{\mu}(1, s, X_i) - \hat{\mu}(0, s, X_i) + \frac{Y_i - \hat{\mu}(1, s, X_i)}{\hat{\pi}(s)}, \\ \tilde{\Xi}_0(\mathcal{D}_i, s) &= \hat{\mu}(1, s, X_i) - \hat{\mu}(0, s, X_i) - \frac{Y_i - \hat{\mu}(0, s, X_i)}{1 - \hat{\pi}(s)}. \end{aligned}$$

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.

Second, for the same reason as before, we need to normalize the difference $\hat{\mu}(1, s, X_i) - \hat{\mu}(0, s, X_i)$ in expressions for $\tilde{\Xi}_a$ and $\tilde{\Xi}_0$ by $\hat{\pi}_0(s) + \hat{\pi}_a(s)$.

Further, notice that since the term $\hat{\Xi}_2$ does not involve $\hat{\pi}_a(s)$ or $\hat{\pi}_0(s)$, the estimator for $\mathbb{E}[\Xi_2^2]$ becomes the sample average among the treated and control for every treatment $a \in \mathcal{A}$, i.e. $\frac{1}{N_a + N_0} \sum_{i \in I_{a,0}} \hat{\Xi}_2^2$. In contrast, the terms $\hat{\Xi}_0$ and $\hat{\Xi}_a$ involve $\hat{\pi}_0(s)$ and $\hat{\pi}_a(s)$. Analogously to the case of $\hat{\tau}_a$, the estimators for $\mathbb{E}[\pi_a(S_i) \Xi_a^2]$ and $\mathbb{E}[\pi_0(S_i) \Xi_0^2]$ become $\frac{1}{N} \sum_{i \in I_{a,0}} \tilde{A}_i \hat{\Xi}_a^2(\mathcal{D}_i, S_i)$ and $\frac{1}{N} \sum_{i \in I_{a,0}} (1 - \tilde{A}_i) \hat{\Xi}_0^2(\mathcal{D}_i, S_i)$, respectively¹.

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

$$\hat{\sigma}_a^2 = \frac{1}{N} \sum_{i \in I_{\{a,0\}}} \left[\tilde{A}_i \hat{\Xi}_a^2(\mathcal{D}_i, S_i) + (1 - \tilde{A}_i) \hat{\Xi}_0^2(\mathcal{D}_i, S_i) \right] + \frac{1}{N_a + N_0} \sum_{i \in I_{\{a,0\}}} \hat{\Xi}_2^2(\mathcal{D}_i, S_i),$$

¹Notice that in a binary treatment case $N_a + N_0 = N$, where $a = 1$, so the estimator is simplified to the original version with a binary treatment.

where

$$\begin{aligned}
\hat{\Xi}_a(\mathcal{D}_i, s) &= \tilde{\Xi}_a(s) - \frac{1}{N_a(s)} \sum_{j \in I_a(s)} \tilde{\Xi}_{a,j}(\mathcal{D}_i, s), \\
\hat{\Xi}_0(\mathcal{D}_i, s) &= \tilde{\Xi}_0(s) - \frac{1}{N_0(s)} \sum_{j \in I_0(s)} \tilde{\Xi}_{0,j}(\mathcal{D}_i, s), \\
\hat{\Xi}_2 &= \left(\frac{1}{N_a(s)} \sum_{j \in I_a(s)} (Y_j - \hat{\tau}_a \tilde{A}_j) \right) - \left(\frac{1}{N_0(s)} \sum_{j \in I_0(s)} (Y_j - \hat{\tau}_a \tilde{A}_j) \right), \\
\tilde{\Xi}_a(\mathcal{D}_i, s) &= \frac{\hat{\mu}(a, s, X_i) - \hat{\mu}(0, s, X_i)}{\hat{\pi}_a(s) + \hat{\pi}_0(s)} + \frac{Y_i - \hat{\mu}(a, s, X_i)}{\hat{\pi}_a(s)}, \\
\tilde{\Xi}_0(\mathcal{D}_i, s) &= \frac{\hat{\mu}(a, s, X_i) - \hat{\mu}(0, s, X_i)}{\hat{\pi}_a(s) + \hat{\pi}_0(s)} - \frac{Y_i - \hat{\mu}(0, s, X_i)}{\hat{\pi}_0(s)},
\end{aligned}$$

where

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

2 CAR with Clusters

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

$$\begin{aligned}
\hat{\tau}_a &= \frac{\sum_{g \in I_{\{a,0\}}} \hat{\Xi}_g}{\sum_{j \in I_{\{a,0\}}} N_j} \times \frac{G_a + G_0}{G} = \frac{\frac{1}{G} \sum_{g \in I_{\{a,0\}}} \hat{\Xi}_g}{\frac{1}{G_a + G_0} \sum_{j \in I_{\{a,0\}}} N_j}, \\
\hat{\Xi}_g &= \frac{\tilde{A}_g (N_g \bar{Y}_g - \hat{\mu}(a, S_g, X_g, N_g))}{\hat{\pi}_a(S_g)} - \frac{(1 - \tilde{A}_g) (N_g \bar{Y}_g - \hat{\mu}(0, S_g, X_g, N_g))}{\hat{\pi}_0(S_g)} \\
&\quad + \frac{\hat{\mu}(a, S_g, X_g, N_g) - \hat{\mu}(0, S_g, X_g, N_g)}{\hat{\pi}_a(S_g) + \hat{\pi}_0(S_g)},
\end{aligned}$$

where

$$\tilde{A}_g = \begin{cases} 1, & \text{if } A_g = a, \forall a \in \mathcal{A} \\ 0, & \text{otherwise.} \end{cases}$$

Variance Estimator

$$\hat{\sigma}_a^2 = \frac{\frac{1}{G} \sum_{g \in I_{\{a,0\}}} [\tilde{A}_g \hat{\Xi}_a^2(\mathcal{D}_g, S_g) + (1 - \tilde{A}_g) \hat{\Xi}_0^2(\mathcal{D}_g, S_g)] + \frac{1}{G_a + G_0} \sum_{g \in I_{\{a,0\}}} \hat{\Xi}_2^2(\mathcal{D}_g, S_g)}{\left(\frac{1}{G_a + G_0} \sum_{g \in I_{\{a,0\}}} N_g \right)^2},$$

where

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

where

$$\tilde{A}_g = \begin{cases} 1, & \text{if } A_g = a, \forall a \in \mathcal{A} \\ 0, & \text{otherwise.} \end{cases}$$