

Bootstrap Variance Derivation

Nik Julius

June 23, 2017

Alternate Title: Just fire yourself at the algebra like an idiot for 6 hours and something will happen.

Note that I am defining τ to be the treatment effect *on the treated* for my convenience. We start from (13) in `bootstrap-variance.pdf`:

$$V^*(\hat{\tau}_b|X, W, Y) = \frac{1}{N_1^2} \sum_{W_i=1} \hat{\varepsilon}_i^2 \quad (1)$$

We want to show that $V^*(\hat{\tau}_b|X, W) \rightarrow_p V(\hat{\tau}|X, W)$. To get $V^*(\hat{\tau}_b|X, W)$, we just take the expectation of (1) over all possible Y:

$$\begin{aligned} V^*(\hat{\tau}_b|X, W) &= E \left[\frac{1}{N_1^2} \sum_{W_i=1} \hat{\varepsilon}_i^2 | X, W \right] \\ &= E \left[\frac{1}{N_1^2} \sum_{W_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 | X, W \right] \end{aligned} \quad (2)$$

First, expand the square (and substitute for $\hat{\tau}$):

$$\left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 = \left(Y_i(1) - \widehat{Y_i(0)} - \frac{1}{N_1} \sum_{W_i=1} Y_i(1) + \frac{1}{N_1} \sum_{W_i=1} \widehat{Y_i(0)} \right)^2 \quad (3)$$

Let $\frac{1}{N_1} \sum_{W_i=1} Y_i(1) = \overline{Y_j(1)}$ and $\frac{1}{N_1} \sum_{W_i=1} \widehat{Y_i(0)} = \widehat{\overline{Y_j(0)}}$. Then (3) is:

$$\begin{aligned}
& Y_i(1)^2 - Y_i(1)\widehat{Y_i(0)} - Y_i(1)\overline{Y_j(1)} + Y_i(1)\widehat{Y_j(0)} \\
& - \widehat{Y_i(0)}Y_i(1) + \widehat{Y_i(0)}^2 + \widehat{Y_i(0)}\overline{Y_j(1)} - \widehat{Y_i(0)}\widehat{Y_j(0)} \\
& - \overline{Y_j(1)}Y_i(1) + \overline{Y_j(1)}\widehat{Y_i(0)} + \overline{Y_j(1)}^2 - \overline{Y_j(1)}\widehat{Y_j(0)} \\
& + \widehat{Y_j(0)}Y_i(1) - \widehat{Y_j(0)}\widehat{Y_i(0)} - \widehat{Y_j(0)}\overline{Y_j(1)} + \widehat{Y_j(0)}^2
\end{aligned} \tag{4}$$

Cancelling terms leaves us with:

$$\begin{aligned}
\left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau}\right)^2 &= Y_i(1)^2 + \widehat{Y_i(0)}^2 + \overline{Y_j(1)}^2 + \widehat{Y_j(0)}^2 - 2Y_i(1)\widehat{Y_i(0)} - 2Y_i(1)\overline{Y_j(1)} \\
&+ 2Y_i(1)\widehat{Y_j(0)} + 2\widehat{Y_i(0)}\overline{Y_j(1)} - 2\widehat{Y_i(0)}\widehat{Y_j(0)} - 2\widehat{Y_j(0)}\overline{Y_j(1)}
\end{aligned} \tag{5}$$

Return to (2) and consider the inside of the expectation:

$$\begin{aligned}
\frac{1}{N_1^2} \sum_{w_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau}\right)^2 &= \frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}^2 + \frac{1}{N_1^2} \sum_{w_i=1} \overline{Y_j(1)}^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_j(0)}^2 \\
&- 2\frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)\widehat{Y_i(0)} - 2\frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)\overline{Y_j(1)} + 2\frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)\widehat{Y_j(0)} \\
&+ 2\frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}\overline{Y_j(1)} - 2\frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}\widehat{Y_j(0)} - 2\frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_j(0)}\overline{Y_j(1)}
\end{aligned} \tag{6}$$

Recall the definition of $\overline{Y_j(1)}$ and $\widehat{Y_j(0)}$. Note that by using one of the $\frac{1}{N_1}$ terms that we just distributed, we can make more of these terms (noting that $\overline{Y_i(1)} = \overline{Y_j(1)}$ and likewise for $\widehat{Y_j(0)}$). This gives us:

$$\begin{aligned}
\frac{1}{N_1^2} \sum_{w_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau}\right)^2 &= \frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}^2 + \frac{1}{N_1^2} \sum_{w_i=1} \overline{Y_j(1)}^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_j(0)}^2 \\
&- 2\frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)\widehat{Y_i(0)} - 2\frac{1}{N_1} \overline{Y_j(1)}^2 + 2\frac{1}{N_1} \widehat{Y_j(0)}\overline{Y_j(1)} \\
&+ 2\frac{1}{N_1} \overline{Y_j(1)}\widehat{Y_j(0)} - 2\frac{1}{N_1} \widehat{Y_j(0)}^2 - 2\frac{1}{N_1} \widehat{Y_j(0)}\overline{Y_j(1)}
\end{aligned} \tag{7}$$

Further, note that $\overline{Y_j(1)}$ and $\widehat{Y_j(0)}$ are constants, which gives us:

$$\begin{aligned}
\frac{1}{N_1^2} \sum_{w_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 &= \frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}^2 + \frac{1}{N_1} \overline{Y_j(1)}^2 + \frac{1}{N_1} \widehat{Y_j(0)}^2 \\
&\quad - 2 \frac{1}{N_1^2} \sum_{w_i=1} Y_i(1) \widehat{Y_i(0)} - 2 \frac{1}{N_1} \overline{Y_j(1)}^2 + 2 \frac{1}{N_1} \widehat{Y_j(0)} \overline{Y_j(1)} \\
&\quad + 2 \frac{1}{N_1} \overline{Y_j(1)} \widehat{Y_j(0)} - 2 \frac{1}{N_1} \widehat{Y_j(0)}^2 - 2 \frac{1}{N_1} \widehat{Y_j(0)} \overline{Y_j(1)}
\end{aligned} \tag{8}$$

Cancelling terms gives us:

$$\begin{aligned}
\frac{1}{N_1^2} \sum_{w_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 &= \frac{1}{N_1^2} \sum_{w_i=1} Y_i(1)^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}^2 - \frac{2}{N_1^2} \sum_{w_i=1} Y_i(1) \widehat{Y_i(0)} \\
&\quad - \frac{1}{N_1} \overline{Y_j(1)}^2 + \frac{2}{N_1} \overline{Y_j(1)} \widehat{Y_j(0)} - \frac{1}{N_1} \widehat{Y_j(0)}^2
\end{aligned} \tag{9}$$

There is a tantalizing glimpse of further cancellation - if the 5th term in (9) wasn't multiplied by $\frac{1}{N_1}$, it would partially cancel with the 3rd term (it would leave only cross products of $Y_i(1) \widehat{Y_j(0)}$ where $i \neq j$). I have been unable to find a mistake leading to that $\frac{1}{N_1}$ being there, though, so I'm fairly sure this is just reality teasing me.

I don't really know how to move on from here without using the things we know about the DGP, so that's how I move forward. We know that $Y_i(1) = \tau$ always, and $Y_i(0) \sim N(0, 1)$. Also, note that we've been working inside the expectation up to now. I transition to considering the entirety of (2) now:

$$\begin{aligned}
E \left[\frac{1}{N_1^2} \sum_{w_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 | X, W \right] &= E \left[\frac{1}{N_1^2} \tau^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}^2 - \frac{2\tau}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)} \right. \\
&\quad \left. - \frac{1}{N_1} \tau^2 + \frac{2\tau}{N_1} \widehat{Y_j(0)} - \frac{1}{N_1} \widehat{Y_j(0)}^2 | X, W \right]
\end{aligned} \tag{10}$$

Again using the definition of $\widehat{Y_j(0)}$ note that the 3rd term is equivalent to the 5th term. Cancelling leaves us with:

$$E \left[\frac{1}{N_1^2} \sum_{w_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 | X, W \right] = E \left[\frac{1}{N_1^2} \tau^2 + \frac{1}{N_1^2} \sum_{w_i=1} \widehat{Y_i(0)}^2 - \frac{1}{N_1} \tau^2 - \frac{1}{N_1} \widehat{Y_j(0)}^2 | X, W \right] \tag{11}$$

Re-arranging and distributing the expectation, we arrive at:

$$E \left[\frac{1}{N_1^2} \sum_{W_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 | X, W \right] = \frac{1}{N_1} \left(\frac{\tau^2}{N_1} - \tau^2 \right) + \frac{1}{N_1^2} E \left[\sum_{W_i=1} \widehat{Y_i(0)}^2 | X, W \right] - \frac{1}{N_1} E \left[\widehat{\widehat{Y_j(0)}}^2 | X, W \right]$$

At this point it is necessary to change to the $K_i Y_i$ notation used in A& I. Note that $\sum_{W_i=1} \widehat{Y_i}^2 = \sum_{W_i=0} K_i^2 Y_i^2$ (this may be wrong?). In the next step I use this on the 3rd term, use the definition of $\widehat{\widehat{Y_j(0)}}$ on the 4th term, and then use the same equality again on that 4th term:

$$\begin{aligned} E \left[\frac{1}{N_1^2} \sum_{W_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 | X, W \right] &= \frac{1}{N_1} \left(\frac{\tau^2}{N_1} - \tau^2 \right) + \frac{1}{N_1^2} \sum_{W_i=0} K_i^2 E[Y_i^2 | X, W] \\ &\quad - \frac{1}{N_1^3} \sum_{W_i=0} K_i^2 E[Y_i^2 | X, W] \end{aligned} \quad (12)$$

Note that going from the 4th term in the previous equation to the 4th term in (12) is fairly involved, and makes use of the fact that $Y_i(0)$ are iid $N(0, 1)$.

Using that distribution of $Y_i(0)$, we know that $E[Y_i^2 | X, W] = 1$. Thus, (12) becomes:

$$E \left[\frac{1}{N_1^2} \sum_{W_i=1} \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} \right)^2 | X, W \right] = \frac{1}{N_1} \left(\frac{\tau^2}{N_1} - \tau^2 \right) + \left(\frac{1}{N_1^2} - \frac{1}{N_1^3} \right) \sum_{W_i=0} K_i^2 \quad (13)$$

This is as far as I've gone. This is *infuriating* because the lower bound of (13) is easily seen to be negative (indeed, for *any* value of $\sum_{W_i=0} K_i^2$, there is a value of τ that makes (13) negative).

However, if we ignore the first term, the last term looks extremely promising. It differs from the correct variance only in the inclusion of the $-\frac{1}{N_1^3}$ factor, and that would probably explain a lot of the underestimation seen in my simulations.

1 To do:

1. Check if (13) matches the variances from the simulations.
2. The above almost certainly will come out false. Check if (13) *without the term involving τ* matches the variances from the simulations.