Bootstrap Variance Consistency

Nik Julius

July 14, 2017

This broadly attempts to apply Otsu & Rai's proof to my algorithm. First, some notation. Let:

$$\mu(W,X) = \mathbb{E}\left[Y_i | W_i = W, X_i = X\right]$$

$$\tau_i = \mu(1,X_i) - \mu(0,X_i)$$

$$\hat{\tau}_i = Y_i(1) - \widehat{Y_i(0)}$$

$$\varepsilon_i = \tau_i - \tau$$

Interpreting $\mu(W,X)$ is obvious. τ_i is the *true* treatment effect on the treated conditional on $X = X_i$. $\hat{\tau}_i$ is the estimate of τ_i generated by the matching estimator procedure. Further, let $M(i)$ be a function that gives the index of the control unit matched to unit i (thus, $\widehat{Y_i(0)} = Y_{M(i)}$).

Otsu & Rai first rewrite their bootstrap estimator $\tilde{\tau}^*$ in terms of the original estimator $\tilde{\tau}$ and residuals, so I do the same for the proposed estimator $\hat{\tau}^*$:

$$\begin{aligned}\hat{\tau}^* &= \hat{\tau} + \frac{1}{N_1}\sum_{W_i=1} e_i \hat{\varepsilon}_i \\ &= \hat{\tau} + \frac{1}{N_1}\sum_{W_i=1} e_i \varepsilon_i + \frac{1}{N_1}\sum_{W_i=1} e_i\left(\hat{\varepsilon}_i - \varepsilon_i\right) \qquad (1)\end{aligned}$$

Where e_i is a draw from the Radembacher distribution. Following Otsu & Rai, let:

$$T_N^{t^*} = \frac{1}{N_1}\sum_{W_i=1} e_i \varepsilon_i$$

$$R_N^{t^*} = \frac{1}{N_1}\sum_{W_i=1} e_i\left(\hat{\varepsilon}_i - \varepsilon_i\right)$$

Still following Otsu & Rai, I want to show the following:

$$\sup_t |Pr\left\{\frac{1}{\sqrt{N_1}}\sum_{W_i=1} e_i\hat{\varepsilon}_i \leq t|\mathbf{Z}\right\} - Pr\left\{\sqrt{N_1}(\hat{\tau}-\tau) \leq t\right\}| \to^p 0 \tag{2}$$

$$\sqrt{N_1}R_N^{t^*} \to^p 0 \tag{3}$$

Abadie & Imbens showed that $\sqrt{N_1}(\hat{\tau}-\tau)/\sigma_N^t$ is asymptotically normal, so following their logic it should suffice for me to prove (for proving (2)):

$$\mathbb{V}(\sqrt{N_1}T_N^{t^*}|\mathbf{Z}) - (\sigma_N^t)^2 \to^p 0 \tag{4}$$

$$|Pr\left\{\sqrt{N_1}T_N^{t^*} \leq t|\mathbf{Z}\right\} - \Phi(t)| \to^p 0 \tag{5}$$

and show that (4) holds for all $t \in \mathbb{R}$. To do this, recall the definition of σ_N^t from Abadie & Imbens:

$$(\sigma_N^t)^2 = (\sigma_{1N}^t)^2 + (\sigma_2^t)^2 \tag{6}$$

$$(\sigma_{1N}^t)^2 = \frac{1}{N_1}\sum_{i=1}^{N}(W_i + (1-W_i)K_i)^2\,\sigma^2(W_i, X_i) \tag{7}$$

$$(\sigma_2^t)^2 = \mathbb{E}\left[(\mu(1,X_i) - \mu(0,X_i) - \tau)^2\,|W_i = 1\right] \tag{8}$$

Following Otsu & Rai's proof suggests an immediate problem with my algorithm - Otsu & Rai are able to show that a part of their bootstrap estimator reproduces the distribution of $(\sigma_{1N}^t)^2$, while no part of my proposed estimator appears to satisfy that role. If I can show that the relevant part of my estimator converges in probability to $(\sigma_2^t)^2$, I believe I will have shown that my procedure does not work, because nothing would remain to deal with $(\sigma_{1N}^t)^2$.

Thus, note that:

$$\begin{aligned}
\mathbb{V}\left(\sqrt{N_1}T_N^{t^*}|\mathbf{Z}\right) &= \mathbb{E}\left(N_1(T_N^{t^*})^2|\mathbf{Z}\right) && (9)\\
&= \mathbb{E}\left(\frac{1}{N_1}\sum_{W_i=1} e_i\varepsilon_i \sum_{j\neq i} e_j\varepsilon_j + \frac{1}{N_1}\sum_{W_i=1} e_i^2\varepsilon_i^2|\mathbf{Z}\right)\\
&= \mathbb{E}\left(\frac{1}{N_1}\sum_{W_i=1}\varepsilon_i^2|\mathbf{Z}\right) && (10)
\end{aligned}$$

Where (9) follows because $\mathbb{E}\left[T_N^{t^*}\right] = 0$, (10) follows because e_i and e_j are independent of each other, and because $e_i^2 = 1$.

I'm slightly uncertain here, because Otsu & Rai say that the comparable part of their variance is simply (10) without the expectation, but I don't see what argument allows me to remove the expectation - ε_i are not known conditioned on $\mathbf{Z}$ because they are the true errors, not the residuals. In any case, note the definition of ε_i from above to see that:

$$\frac{1}{N_1} \sum_{W_i=1} \varepsilon_i^2 = \frac{1}{N_1} \sum_{W_i=1} (\tau_i - \tau)^2 \tag{11}$$

This is equivalent to $(\hat{\sigma}^t_{2N})^2$ in Otsu & Rai, who claim directly that (11) converges in probability to $(\sigma^t_2)^2$ by the law of large numbers. This makes the problem with my algorithm fairly obvious - the only remaining piece of my bootstrapped estimator is $R^{t^*}_N$, and it seems we should expect this to converge to 0, not to a useful term. Regardless:

$$\begin{aligned}
R^{t^*}_N &= \frac{1}{N_1} \sum_{W_i=1} e_i \left(Y_i(1) - \widehat{Y_i(0)} - \hat{\tau} - \mu(1, X_i) + \mu(0, X_i) \right) \\
&= \frac{1}{N_1} \sum_{W_i=1} e_i \left(\mu(1, X_i) + \epsilon_i - \mu(0, X_{M(i)}) - \epsilon_{M(i)} - \hat{\tau} - \mu(1, X_i) + \mu(0, X_i) \right) \\
&= \frac{1}{N_1} \sum_{W_i=1} e_i \left(\epsilon_i + (\mu(0, X_i) - \mu(0, X_{M(i)}) - \epsilon_{M(i)} - \hat{\tau} \right) \\
&= \frac{1}{N_1} \sum_{W_i=1} e_i \left(\epsilon_i - \epsilon_{M(i)} + \hat{\tau}_i - \hat{\tau} \right) \qquad (12)
\end{aligned}$$

I think it suffices at this point to note that e_i is independent of $\epsilon_i, \epsilon_{M(i)}, \hat{\tau}_i$, and $\hat{\tau}$, and thus (12) $\to^p 0$ by the law of large numbers.

Thus, the problem is clear. The proposed algorithm reproduces $(\sigma^t_2)^2$, but this is only part of what it needs to reproduce. Nothing in the algorithm reproduces $(\sigma^t_{1N})^2$. This is weird, because its the reverse of what we thought was wrong with the procedure, but I haven't found an error in this so far.