Bootstrap Variance Consistency

Nik Julius

July 26, 2017

1 "Proof" of Failure

I have an argument, which may or may not constitute a formal proof, that the proposed bootstrap algorithm fails to reproduce the target variance. The outline of the "proof" is derived from Otsu & Rai's successful proof of bootstrap consistency. Their proof proceeds as follows:

1. Decompose bootstrapped estimator $\tilde{\tau}^{t^*}$ into the original estimator $\tilde{\tau}$, a variance (I think) term $T_N^{t^*}$, and 2 remainder terms $R_{1N}^{t^*}$ and $R_{2N}^{t^*}$.

2. Show the following:

$$\sup_t \mid Pr\left\{\sqrt{N_1} T_N^{t^*} \leq t|\mathbf{Z}\right\} - Pr\left\{\sqrt{N_1}\left(\tilde{\tau}^t - \tau^t\right) \leq t\right\} \mid \to_p 0 \tag{1}$$

$$\sqrt{N_1}\left(R_{1N}^{t^*} + R_{2N}^{t^*}\right) \to_p 0 \tag{2}$$

3. To show (1), since Abadie & Imbens showed that $\sqrt{N_1}\left(\tilde{\tau}^t - \tau^t\right)/\sigma_N^t$ is asymptotically normal, prove the following:

$$Var\left(\sqrt{N_1} T_N^{t^*}|\mathbf{Z}\right) - (\sigma_N^t)^2 \to_p 0 \tag{3}$$

$$\mid Pr\left\{\sqrt{N_1} T_N^{t^*} \leq t|\mathbf{Z}\right\} - \Phi(t) \mid \to_p 0 \ \forall t \in \mathbb{R} \tag{4}$$

What I have done is attempted to follow this roadmap, and what happens is that in trying to show (3), I instead show that $Var(\sqrt{N_1} U_N^{t^*}) - (\sigma_N^t)^2$ (where $U_N^{t^*}$ is my analogue to Otsu & Rai's $T_N^{t^*}$) does not converge to 0. My uncertainty over whether this constitutes a proof is basically uncertainty over the direction of implications in this proof.

Basically, my proof says "Since Abadie & Imbens showed that $\sqrt{N_1}\left(\tilde{\tau}^t - \tau^t\right)/\sigma_N^t$ is asymptotically normal, showing that (3) does not hold is the same as showing that (1) does not hold, and if (1) does not hold, the bootstrap procedure fails to reproduce the target variance".

2 Actual Proof

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

$$\begin{aligned}
\mu(W, X) &= \mathbb{E}\left[Y_i | W_i = W, X_i = X\right] \\
\tau_i &= \mu(1, X_i) - \mu(0, X_i) \\
\hat{\tau}_i &= Y_i(1) - \widehat{Y_i(0)} \\
\varepsilon_i &= \tau_i - \tau
\end{aligned}$$

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 $\hat{\tau}^*$:

$$\begin{aligned}
\hat{\tau}^* &= \hat{\tau} + \frac{1}{N_1} \sum_{W_i=1} e_i \hat{\varepsilon}_i \\
&= \hat{\tau} + \frac{1}{N_1} \sum_{W_i=1} e_i \varepsilon_i + \frac{1}{N_1} \sum_{W_i=1} e_i \left(\hat{\varepsilon}_i - \varepsilon_i\right) \qquad (5)
\end{aligned}$$

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

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

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

$$\sup_t |Pr\left\{\frac{1}{\sqrt{N_1}} \sum_{W_i=1} e_i \hat{\varepsilon}_i \leq t | \mathbf{Z}\right\} - Pr\left\{\sqrt{N_1}(\hat{\tau} - \tau) \leq t\right\}| \to^p 0 \qquad (6)$$

$$\sqrt{N_1} Q_N^{t^*} \to^p 0 \qquad (7)$$

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 (6)):

$$\mathbb{V}(\sqrt{N_1}Q_N^{t^*}|\mathbf{Z}) - (\sigma_N^t)^2 \to^p 0 \tag{8}$$

$$|Pr\left\{\sqrt{N_1}Q_N^{t^*} \leq t|\mathbf{Z}\right\} - \Phi(t)| \to^p 0 \tag{9}$$

and show that (9) holds for all $t \in \mathbb{R}$. To do this, recall the definition of σ_N^t from Abadie & Imbens:

$$(\sigma_N^t)^2 = (\sigma_{1N}^t)^2 + (\sigma_2^t)^2 \tag{10}$$

$$(\sigma_{1N}^t)^2 = \frac{1}{N_1}\sum_{i=1}^{N}(W_i + (1-W_i)K_i)^2\,\sigma^2(W_i, X_i) \tag{11}$$

$$(\sigma_2^t)^2 = \mathbb{E}\left[(\mu(1,X_i) - \mu(0,X_i) - \tau)^2\,|W_i = 1\right] \tag{12}$$

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:

$$\begin{aligned}
\mathbb{V}\left(\sqrt{N_1}Q_N^{t^*}|\mathbf{Z}\right) &= \mathbb{E}\left(N_1(Q_N^{t^*})^2|\mathbf{Z}\right) && (13)\\
&= \mathbb{E}\left(\frac{1}{N_1}\sum_{W_i=1} e_i\varepsilon_i \sum_{j\neq i} e_j\varepsilon_j + \frac{1}{N_1}\sum_{W_i=1} e_i^2\varepsilon_i^2|\mathbf{Z}\right)\\
&= \mathbb{E}\left(\frac{1}{N_1}\sum_{W_i=1}\varepsilon_i^2|\mathbf{Z}\right) && (14)
\end{aligned}$$

Where (13) follows because $\mathbb{E}\left[Q_N^{t^*}\right] = 0$, and (14) 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 (14) 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:

$$\frac{1}{N_1}\sum_{W_i=1}\varepsilon_i^2 = \frac{1}{N_1}\sum_{W_i=1}(\tau_i-\tau)^2 \tag{15}$$

This is equivalent to $(\hat{\sigma}^t_{2N})^2$ in Otsu & Rai, who claim directly that (15) converges in probability to $(\sigma^t_2)^2$ by the law of large numbers.

The difference between $Q_N^{t^*}$ (my algorithm) and $T_N^{t^*}$ (Otsu & Rai's algorithm) appears to be the key:

$$\begin{aligned}
Q_N^{t^*} &= \frac{1}{N_1}\sum_{W_i=1} e_i\varepsilon_i \\
T_N^{t^*} &= \frac{1}{N_1}\sum_{W_i=1} e_i(\varepsilon_i+\xi_i) + \frac{1}{N_1}\sum_{W_i=0} e_i(K_i\varepsilon_i) \\
T_N^{t^*} - Q_N^{t^*} &= \frac{1}{N_1}\sum_{W_i=1} e_i\xi_i + \frac{1}{N_1}\sum_{W_i=0} e_i(K_i\varepsilon_i) \qquad (16)
\end{aligned}$$

By the definitions in Otsu & Rai, (16) is equivalent to:

$$\begin{aligned}
D_N^{t^*} &= \frac{1}{N_1}\sum_{W_i=1} e_i\left(\mu(1,X_i)-\mu(0,X_i)\right) + \frac{1}{N_1}\sum_{W_i=0} e_iK_i(\tau_i-\tau) \\
&= \frac{1}{N_1}\sum_{W_i=1} e_i\tau_i + \frac{1}{N_1}\sum_{W_i=0} e_iK_i(\tau_i-\tau) \qquad (17)
\end{aligned}$$

This makes the problem with my algorithm fairly obvious - the only remaining piece of my bootstrapped estimator is $U_N^{t^*}$, and it seems we should expect this to converge to 0, not to a useful term. Regardless:

$$\begin{aligned}
U_N^{t^*} &= \frac{1}{N_1}\sum_{W_i=1} e_i\left(Y_i(1)-\widehat{Y_i(0)}-\hat{\tau}-\mu(1,X_i)+\mu(0,X_i)\right) \\
&= \frac{1}{N_1}\sum_{W_i=1} e_i\left(\mu(1,X_i)+\epsilon_i-\mu(0,X_{M(i)})-\epsilon_{M(i)}-\hat{\tau}-\mu(1,X_i)+\mu(0,X_i)\right) \\
&= \frac{1}{N_1}\sum_{W_i=1} e_i\left(\epsilon_i+(\mu(0,X_i)-\mu(0,X_{M(i)})-\epsilon_{M(i)}-\hat{\tau}\right) \\
&= \frac{1}{N_1}\sum_{W_i=1} e_i\left(\epsilon_i-\epsilon_{M(i)}+\hat{\tau}_i-\hat{\tau}\right) \qquad (18)
\end{aligned}$$

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.

3 The Intuition for the failure

Otsu & Rai do not provide the exact term that my procedure fails to estimate, but I believe I am correct in saying that it is as follows:

$$\begin{aligned}(\sigma_{1N}^t)^2 &= Var\left(\frac{1}{\sqrt{N_1}}\sum_{W_i=1}\left(Y_i(1)-\mu(1,X_i)+\widehat{Y_i(0)}-\mu(0,X_{M(i)})\right)|\mathbf{W},\mathbf{X}\right)\\ &= Var\left(\frac{1}{\sqrt{N_1}}\sum_{W_i=1}\left(\epsilon_i+\epsilon_{M(i)}\right)|\mathbf{X},\mathbf{W}\right)\end{aligned} \tag{19}$$

This suggests that Kevin's original concern is well-founded - the missing part of the variance appears to be related to the noisiness of $Y_i(1)$ and $\widehat{Y_i(0)}$ as estimates of $\mu(1, X_i)$ and $\mu(0, X_{M(i)})$.

The remaining concern is that (19) is always positive, and therefore does not explain the cases where my procedure overestimates the target variance. This may be down to me misunderstanding the target variance (still), errors in the above, errors in the simulation, or $U_N^{t^*}$ converging so slowly that it dominates. I think the last is the most likely, followed by errors in the simulation. I have not been able to find errors in the above proofs.

On a related note, I suspect it may be that this failure would be present even in the naive bootstrap - I'm not sure if it would be possible to show that, but nothing in Abadie & Imbens' paper on the naive bootstrap suggests (to me, at least) that the ϵ residuals would not cause problems there as well.

I also do not think that this failure is related to the bias correction proposed by Abadie & Imbens - their bias correction is of the form:

$$\hat{B}_N^t = \frac{1}{N_1}\sum_{W_i=1}\left(\hat{\mu}(0,X_i)-\hat{\mu}(0,X_{M(i)})\right) \tag{20}$$

This is clearly a correction for bad matches, not a correction for noisy estimation of $\mu(W_i, X_i)$.