Kevin's Notes on Matching Estimators for Nik

Kevin D. Duncan

July 13, 2017

Some Assumptions from [3] and [2]

Assumption 0.1. *1. Conditional on $W_i = w$ the sample consistss of independent draws from $Y, X \mid W = w$ for $w \in \{0, 1\}$. For some $r \leq 1$, $N_1^r / N_0 \to \theta \in (0, \infty)$*

2. X is continuously distributed on compact and convex support $X \subset R^k$. The density of X is bounded and bounded away from zero on $\mathbb{X}$.

3. W is independent of $Y_i(0)$ conditional on $X = x$ for almost every x. There exists a positive constant c such that $P(W = 1 \mid X = x) \leq 1 - c$ for almost every x

4. For $w = 0, 1$, $\mu(w, x)$ and $\sigma^2(w, x)$ are Lipschitz in $\mathbb{X}$, $\sigma^2(w, x)$ is bounded away from zero on $\mathbb{X}$, and $\mathrm{E}[Y^4 \mid W = w, X = x]$ is bounded uniformly on $\mathbb{X}$.

Proposed Procedure

1. Estimate $\hat{\tau}^t$
2. Get residuals $\hat{\epsilon}_i = Y_i^{obs} - \hat{Y_i(0)} - \hat{\tau}^t - B_t$ for each treated observation i. (you constantly forget the hat on the epsilons, its important to denote that its a sample residual and not from the population!)
3. For each i, draw $u_i^* = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability} 1/2 \end{cases}$. I.e. the Rademacher distribution
4. Create N_1 bootstrap outcomes, $Y_i^* = \hat{Y_i(0)} + \hat{\tau}^t + u_i^* \hat{\epsilon}_i$
5. Estimate $\hat{\tau}_b^t$ on the bootstrap sample $Z = \{Y_i^*, W, X\}$
6. Repeat the steps 3-5 for $b = 1, ..., B$ bootstraps, and estimate $\hat{\tau}$ as the sample variance of $\hat{\tau}_b$ over the b bootstraps.

For each individual we observe an outcome variable Y_i^{obs}, a set of covariates, X_i, and whether or not an individual was treated, W_i, which is either 1, or 0. For both the treated and untreated groups, we can define the conditional means, $\mu(0, x) = \mathrm{E}[Y_i(0) \mid X_i = x] \quad \mu(1, x) = \mathrm{E}[Y_i(1) \mid X_i = x]$. Then, for each individual i, the Average Treatment Effect on the Treated (ATET) is

$$\tau_i^t = \mu(1, x) - \mu(0, x)$$

For each group, we have $Y_i^{obs} = \mu(W_i, X_i) + e_i$, whether the error structure may dependend on whether or not an individual was treated or not. Assuming both groups have mean zero errors, we have the sample individual ATET,

$$\hat{\tau}_i^t = Y_i^{obs} - \hat{Y}_i(0) = Y_i(1) - \hat{Y}_i(0)$$

For finite samples the matching may not be perfect even under Assumptions (0.1. Let M be the number of matches, such that j is "closer" to i than M, then, $j \in J_M(i)$, being the set of matched pairs. Then, when we have $X_i \neq X_{m_i^c}$

$$\begin{aligned}
\mathrm{E}_{sp}(Y_i^{obs} - Y_{m_i^c} \mid W_i = 1, X_i, X_{m_i^c}) = \mathrm{E}_{sp}(Y_i^{obs} - Y_{m_i^c} \mid X_i, X_{m_i^c}) &= \mu(1, X_i) - \mu(0, X_{m_i^c}) \\
&= \tau(x) + \mu(0, X_i) - \mu(0, X_{m_i^c})
\end{aligned}$$

Thus we see in finite samples, when our matching isn't perfect, that there will be a residual bias term that differs from the ATET. [1] show that the bias term is stochastically bounded, in particular for the ATET it is $O_p(N^{-r/k})$, such that if $k > 2r$ the bias term maw disrupt the convergence in distribution. Regardless, for fixed n, for finite samples by summing across treated individuals,

$$\begin{aligned}
\hat{\tau} = N_1^{-1} \sum_{W_i=1} \hat{\tau}_i^t &= \tau + N_1^{-1} \sum_{W_i=1} \mu(0, X_i) - \mu(0, X_{m_i^c}) \\
&= \tau + N_1^{-1} \sum_{i}^{N} W_i \left(M^{-1} \sum_{j \in J_M(i)} (\mu(0, X_i) - \mu(0, X_j)) \right)
\end{aligned}$$

Now define,

$$B_n^t = N_1^{-1} \sum_{i}^{N} W_i \left(M^{-1} \sum_{j \in J_M(i)} (\mu(0, X_i) - \mu(0, X_j)) \right)$$

Under assumptions 0.1 we know there are consistent nonparametric estimators $\hat{\mu}(1, x) \to^p \mu(1, x)$ and $\hat{\mu}(0, x) \to^p \mu(0, x))$ for all x. Then, the natural estimator that arises is,

$$\begin{aligned}
\tilde{\tau}^t &= \hat{\tau}^t - \hat{B}_n^t \\
&= N_1^{-1} \sum_{W_i=1} \left((Y_i^{obs} - M^{-1} \sum_{j \in J_M(i)} Y_j^{obs}) - \left(M^{-1} \sum_{j \in J_M(i)} (\hat{\mu}(0, X_i) - \hat{\mu}(0, X_j)) \right) \right) \\
&= N_1^{-1} \sum_{W_i=1} \left((Y_i^{obs} - \hat{\mu}(0, X_i) - M^{-1} \sum_{j \in J_M(i)} Y_j^{obs} - \hat{\mu}(0, X_j)) \right) \\
&= N_1^{-1} \sum_{W_i=1} \left((Y_i^{obs} - \hat{\mu}(1, X_i) - M^{-1} \sum_{j \in J_M(i)} (Y_j^{obs} - \hat{\mu}(0, X_j)) + \hat{\mu}(1, X_i) - \hat{\mu}(0, X_i) \right)
\end{aligned}$$

We want to evaluate the probability limiting behavior of,

$$\begin{aligned}
\mathrm{Var}^*(\hat{\tau}_b^t) &= N_1^{-2}\sum_{W_i=1}\mathrm{E}^*(u_i^*\hat{\epsilon}_i^2) = N_1^{-2}\sum_{W_i=1}\hat{\epsilon}_i^2 \\
&= N_1^{-2}\sum_{W_i=1}(Y_i(1)-\hat{Y_i(0)}-\hat{\tau}^t)^2 \\
&= N_1^{-2}\sum_{W_i=1}(Y_i(1)^2+\hat{Y_i(0)}^2+\hat{\tau}^{t2}-2Y_i(1)\hat{Y_i(0)}-2Y_i(1)\hat{\tau}^t+2\hat{Y_i(0)}\hat{\tau}^t)
\end{aligned}$$

and we know by law of large numbers (this is tentative, and meant to be illustrative, generally we want to expand each term and then apply a weak/strong law of large numbers. Convergence in probability is all that is required). I.e. Under Assumptions 0.1 we know that individuals are *iid*, that $\mathrm{E}(\mu(W_i,X_i))$ and $\mathrm{E}(\mu(W_i,X_i)^2)$ exists for all i, $\mathrm{E}(e_i)$ and $\mathrm{E}(e_i^2)$ exist for all i, then by the strong law of large numbers,

$$\begin{aligned}
N_1^{-2}\sum_{W_i=1}Y_i(1)^2 &= N_1^{-2}\sum_{W_i=1}\mu(1,x_i)^2+\mu(1,x_i)e_i+e_i^2 \\
\implies N_1^{-2}\sum_{W_i=1}\mu(1,x_i)^2 &\to^p \bar{\mu(1,x)}^2 \\
N_1^{-2}\sum_{W_i=1}\mu(1,x_i)e_i &\to^p 0 \\
N_1^{-2}\sum_{W_i=1}e_i &= N_1^{-1}\frac{\sum_{W_i=1}e_i^2}{N_1}\to^p 0
\end{aligned}$$

For our sample estimator for the ATET,

$$\begin{aligned}
\hat{\tau}^t &= \frac{1}{N_t}\left(\sum_{W_i=1}Y_i(1)-\sum_{W_i=0}K_iY_i(0)\right) \\
\implies (\hat{\tau}^t)^2 &= \frac{1}{N_t^2}(\sum_{W_i=1}Y_i(1)^2+\sum_{W_i=1,W_j=1,j\neq i}Y_i(1)Y_j(1) \\
&+\sum_{W_i=0}K_i^2Y_i(0)^2+\sum_{W_i=0,W_j=0,j\neq i}K_iY_i(0)K_jY_j(0) \\
&-2\sum_{W_i=1}Y_i(1)\sum_{W_i=0}K_iY_i(0))
\end{aligned}$$

Similarly,

$$\begin{aligned}
\hat{Y_i(0)} &= \frac{1}{M}\sum_{j\in J_m(i)}Y_j(0) \\
\implies (\hat{Y_i(0)})^2 &= \frac{1}{M^2}(\sum_{j\in J_m(i)}Y_j(0)^2+\sum_{j,k\in J_m(i),j\neq k}Y_j(0)Y_k(0))
\end{aligned}$$

and $\sum_{W_i=1}\sum_{j\in J_m(i)}1=\sum_{W_i=0}K_i$. Then,

$$\sum_{W_i=1} \hat{Y_i(0)} = \frac{1}{M^2} \sum_{W_i=1} (\sum_{j \in J_m(i)} Y_j(0)^2 + \sum_{j,k \in J_m(i), j \neq k} Y_j(0) Y_k(0))$$
$$= \frac{1}{M^2} \sum_{W_i=0} K_i Y_i(0)^2 + \frac{1}{M^2} \sum_{W_j=0, W_k=0, j \neq k} K_j Y_j(0) K_k Y_k(0))$$

Continuing our exansion,

$$2N_1^{-2} \sum_{W_i=1} Y_i(1) \hat{Y_i(0)} = 2N_1^{-2} \sum_{W_i=1} Y_i(1) \sum_{j \in J_M(i)} Y_j(0)$$

$$2N_1^{-2} \sum_{W_i=1} Y_i(1) \hat{\tau}^t =$$

$$2N_1^{2} \sum_{W_i=1} \hat{Y_i(0)} \hat{\tau}^t =$$