Matching via Kernel Density Estimators

Chi-Shian Dai

10/15/2019

1 Introduction

$\{X_i, Y_i, A_i\}$ are i.i.d random variables. $X_i \in \mathbb{R}^p$ are covariate, $A_i \in \{0, 1\}$ are treatments, $Y_i \in \mathbb{R}$ are responsed variables. Instead of using traditional matching method, [1] propose finding the weighted measure on control via minimizing the operator norm between the weighted measure and the empirical measure on treatment. Their method can make sure the weighted measure on control and the empitical measure on treatment are similar enough to get the consisteny estimator of the treatment effect on treatment $\tau := E[Y(1)|A = 1] - E[Y(0)|A = 1]$; however, they cannot provide the nomality. In this paper, we use Kernel Density Method to estimate the weighted which is consistency and asymptotic normality.

2 Kernel Density Estimator

Kernel density estimator is a well-known method in estimating densities. Having the density estimatos on both treatment and control. Let $\hat{f}_0(\cdot)$ and $\hat{f}_1(\cdot)$ be kernel density estimator.

$$\hat{f}_0(x) = \frac{1}{n_0} \sum_{i \in T_0} \delta_h(x - x_i),$$

$$\hat{f}_1(x) = \frac{1}{n_1} \sum_{i \in T_1} \delta_h(x - x_i).$$

Furthermore, Radon–Nikodym theorem give us an idea to transfer control's measure to treatment's measure, by add a weighted constant $\hat{f}_0/\hat{f}_1$.

Hence, we can defind the estimator

$$\hat{\tau} := \frac{1}{n_1}\sum_{i\in T_1} Y_i - \frac{1}{n_0}\sum_{i\in T_0} Y_i \frac{\hat{f}_1(X_i)}{\hat{f}_0(X_i)}.$$

In the section, we would develop the theoretical property of estimator $\hat{\tau}$.

2.1 Consistency

We asssume the standard assuption in causaul inference.

(A1) SUTVA $Y_j = Y_j(1)A_j + Y_j(0)(1 - A_j)$.

(A2) Ignorability $Y_j(1), Y_j(0) \perp\!\!\!\perp A_i | X_j$

(A3) Positivity $X|A = 1 \sim F_1$ with density f_1, and $X|A = 0 \sim F_0$ with desity f_0, on the support $\mathcal{E}$. Furthermore, $f_0 << f_1$, and $f_1 << f_0$

Assume we know the density f_1 and f_0, the following theorem show that

$$\hat{\tau}^* := \frac{1}{n_1}\sum_{i\in T_1} Y_i - \frac{1}{n_0}\sum_{i\in T_0} Y_i \frac{f_1(X_i)}{f_0(X_i)}.$$

is an unbias estimator of $\tau := E[Y(1)|A = 1] - E[Y(0)|A = 1]$. By law of large number, it is also consistency.

Theorem 2.1. *Under Assumption (A1)-(A3), then*

$$E[Y_i \frac{f_1(X_i)}{f_0(X_i)} | A_i = 0] = E[Y_i(0)|A_i = 1].$$

Proof. From (A1) SUTVA

$$E[Y_i \frac{f_1(X_i)}{f_0(X_i)} | A_i = 0] = E[Y_i(0) \frac{f_1(X_i)}{f_0(X_i)} | A_i = 0]$$

From (A2) Randomness

$$E[Y_i(0) \frac{f_1(X_i)}{f_0(X_i)} | A_i = 0] = E[E[Y_i(0)|X_i] \frac{f_1(X_i)}{f_0(X_i)} | A_i = 0]$$

From (A3)

$$E[E[Y_i(0)|X_i]\frac{f_1(X_i)}{f_0(X_i)}|A_i = 0] = \int E[Y_i(0)|X_i]\frac{f_1(X_i)}{f_0(X_i)}f_0(X_i)dX_i$$
$$= \int E[Y_i(0)|X_i]f_1(X_i)dX_i = E[Y_i(0)|A_i = 1]$$

□

Althogh $\hat{\tau}^*$ is consistency, in general we do not have densities f_1 and f_0. Hence, we use kernel density estimator to plug in f_1 and f_0.

Here is a well-known theorem of kernal density estimator.

Theorem 2.2 ([2]). *Under some regularity condition,*

$$\|\hat{f}_0 - f_0\|_\infty \xrightarrow{a.e.} 0,$$
$$\|\hat{f}_1 - f_1\|_\infty \xrightarrow{a.e.} 0,$$

where f_0 and f_1 are true density.

From theorem 2.2, and law of large number, $\hat{\tau}$ is a consistency estimator for $\tau := E[Y(1)|A = 1] - E[Y(0)|A = 1]$.

2.2 Normality

In the normality, we need further assumptions. To control the difference between $\hat{\tau}$ and $\hat{\tau}^*$.

(C1) $Y(1)$ and $Y(0)$ are random varaibes bounded by M.

(C2) The densities f_1, f_0 are larger than a positve constant κ.

(C3) There is a constant c such that $\frac{n_0}{n_1} \to c$

(C4) $E_1[f_0(X)]$, and $E_1[f_0(X)^2]$ are integrable.

Lemma 2.3. *Under regularity condition in Theorem 2.2, (C1)-(C4).*

$$\sqrt{n_0}(\hat{\tau} - \hat{\tau}^*) = \frac{1}{\sqrt{n_0}}\sum_{i\in T_0} Y_i\left(\frac{\hat{f}_1(X_i)}{\hat{f}_0(X_i)} - \frac{f_1(X_i)}{f_0(X_i)}\right) = o_p(1)$$

Proof.

$$
\begin{aligned}
&\frac{1}{\sqrt{n_0}}\sum_{i\in T_0} Y_i\left(\frac{\hat{f}_1(X_i)}{\hat{f}_0(X_i)}-\frac{f_1(X_i)}{f_0(X_i)}\right)\\
=&\frac{1}{\sqrt{n_0}}\sum_{i\in T_0}\frac{Y_i}{f_0(X_i)}(\hat{f}_1(X_i)-f_1(X_i)) && (1)\\
+&\frac{1}{\sqrt{n_0}}\sum_{i\in T_0}\frac{Y_i(f_0(X_i)-\hat{f}_0(X_i))}{\hat{f}_0(X_i)f_0(X_i)}\hat{f}_1(X_i) && (2)
\end{aligned}
$$

For every $\epsilon > 0$,

$$
\begin{aligned}
P(|(1)|\geq \epsilon|\boldsymbol{X}_{T_1}) \leq&\frac{Var((1)|\boldsymbol{X}_{T_1})}{\epsilon^2}\\
=&\frac{1}{\epsilon^2}Var\left(\frac{Y_i}{f_0(X_i)}(\hat{f}_1(X_i)-f_1(X_i))|\boldsymbol{X}_{T_1}\right)\\
\leq&\frac{1}{\epsilon^2}E\left[\frac{Y_i^2}{f_0(X_i)^2}(\hat{f}_1(X_i)-f_1(X_i))^2|\boldsymbol{X}_{T_1}\right]\\
\leq&\frac{1}{\epsilon^2}E_0\left[\frac{Y_i^4}{f_0(X_i)^4}\right]^{1/2}E\left[(\hat{f}_1(X_i)-f_1(X_i))^4|\boldsymbol{X}_{Y_1}\right]^{1/2}\\
=&\frac{1}{\epsilon^2}E_0\left[\frac{Y_i^4}{f_0(X_i)^4}\right]^{1/2}\left(\int(\hat{f}_1(X_i)-f_1(X_i))^4 dF_0\right)^{1/2}\\
\leq&\frac{1}{\epsilon^2}E_0\left[\frac{Y_i^4}{f_0(X_i)^4}\right]^{1/2}\|\hat{f}_1-f_1\|_\infty^2.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
P(|(1)|\geq\epsilon) \leq& P(|(1)|\geq\epsilon;\|\hat{f}_1-f_1\|_\infty\leq\epsilon^3)+P(\|\hat{f}_1-f_1\|_\infty>\epsilon^3)\\
\leq& E_0\left[\frac{Y_i^4}{f_0(X_i)^4}\right]\epsilon+P(\|\hat{f}_1-f_1\|_\infty>\epsilon^3)\\
\leq&\frac{M^4}{\kappa^4}\epsilon+P(\|\hat{f}_1-f_1\|_\infty>\epsilon^3)\\
\xrightarrow{n}&\frac{M^4}{\kappa^4}\epsilon.
\end{aligned}
$$

The third inequality comes from Condition (C1) and (C2). The last line comes from the Theorem 2.2. On the orther hand,

$$
\begin{aligned}
P(|(2)| \geq \epsilon | X_{T_0}) &\leq P(\frac{M}{\kappa^2}\|\hat{f}_0 - f_0\|_\infty |\frac{1}{\sqrt{n_0}} \sum_{i\in T_0} \hat{f}_1(X_i)| \geq \epsilon | X_{T_0}) \\
&= P(\frac{M}{\kappa^2}\|\hat{f}_0 - f_0\|_\infty |\frac{1}{\sqrt{n_0}} \sum_{i\in T_0}\sum_{j\in T_1} \frac{1}{n_1}\delta_h(X_i - X_j)| \geq \epsilon | X_{T_0}) \\
&= P(\frac{M}{\kappa^2}\|\hat{f}_0 - f_0\|_\infty \sqrt{\frac{n_0}{n_1}} |\frac{1}{\sqrt{n_1}} \sum_{j\in T_1}\sum_{i\in T_0} \frac{1}{n_0}\delta_h(X_i - X_j)| \geq \epsilon | X_{T_0}) \\
&= P(\frac{M}{\kappa^2}\|\hat{f}_0 - f_0\|_\infty \sqrt{\frac{n_0}{n_1}} |\frac{1}{\sqrt{n_1}} \sum_{i\in T_1} \hat{f}_0(X_i)| \geq \epsilon | X_{T_0}) \\
&\leq \frac{M^2}{\kappa^4}\frac{n_0}{n_1}\|\hat{f}_0 - f_0\|_\infty^2 \frac{Var_1(\hat{f}_0(X)|X_{T_0})}{\epsilon^2} \\
&\leq \frac{M^2}{\kappa^4\epsilon^2}\frac{n_0}{n_1}\|\hat{f}_0 - f_0\|_\infty^2 (\|\hat{f}_0 - f_0\|_\infty^2 + 2\|\hat{f}_0 - f_0\|_\infty E_1[f_0(X)] + E_1[f_0(X)^2]) \\
&= O(\frac{\|\hat{f}_0 - f_0\|_\infty^2}{\epsilon^2})
\end{aligned}
$$

The first line comes from $\hat{f}_1$ are positive. The fiveth line comes from Chebeshev inequality. The last approximation comes from Condition (C3) and (C4). Then,

$$
\begin{aligned}
P(|(2)| \geq \epsilon) =& P(|(2)| \geq \epsilon; \|\hat{f}_0 - f_0\|_\infty \leq \epsilon^2) + P(\|\hat{f}_0 - f_0\|_\infty \geq \epsilon^2) \\
\leq& O(\epsilon^2) + P(\|\hat{f}_0 - f_0\|_\infty \geq \epsilon^2) \\
\to& O(\epsilon^2)
\end{aligned}
$$

The last inequality comes from Theorem 2.2. □

Theorem 2.4. *Under assumption in Lemma 2.3,*

$$\sqrt{n_0}(\hat{\tau} - \tau) \xrightarrow{d} N(0, V),$$

where

$$V = c^2 Var(Y(1)|C = 1) + Var(Y(0)\frac{f_1(X)}{f_0(X)}|A = 0).$$

Proof.

$$\sqrt{n_0}(\hat{\tau} - \tau) = \sqrt{n_0}\left(\sum_{i\in T_1}\frac{1}{n_1}Y_i - \sum_{i\in T_0}\frac{1}{n_0}\frac{f_1(X_i)}{f_0(X_i)} - \tau\right) + \frac{1}{\sqrt{n_0}}\sum_{i\in T_0} Y_i \left(\frac{\hat{f}_1(X_i)}{\hat{f}_0(X_i)} - \frac{f_1(X_i)}{f_0(X_i)}\right)$$

From Lemma 2.3, the second term convergnece in probability to zero. And from Central Limit Theorem, the first term convergence in distribution to $N(0, V)$. Then, Slutcky theorem complete the argument. $\square$

References

[1] Nathan Kallus. Generalized optimal matching methods for causal inference. 2016.

[2] Bernard W. Silverman. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, 6(1):177–184, 01 1978.