Tools and Techniques for ML Homework 1

Zian Jiang (zj444)

February 23, 2021

1 Estimators for missing at random (MAR)

1.1 Total inverse propensity weight for observations has expectation n

1.1.1

$$
\begin{aligned}
E(\sum_{i=1}^{n} W_i R_i) &= \sum_{i=1}^{n=1} E(W_i R_i) \\
&= \sum_{i=1}^{n=1} E(\frac{R_i}{\pi(X_i)}) \\
&= \sum_{i=1}^{n=1} E(E(\frac{R_i}{\pi(X_i)}|X_i)) \text{ (by Adam's law)} \\
&= \sum_{i=1}^{n=1} E(\frac{1}{\pi(X_i)} E(R_i|X_i)) \\
&= \sum_{i=1}^{n=1} E(\frac{1}{\pi(X_i)} \pi(X_i)) \text{ (by definition)} \\
&= \sum_{i=1}^{n=1} 1 = n.
\end{aligned}
$$

■

1.2 Complete case estimator is not consistent for MAR setting

1.2.1

$$
\begin{aligned}
\hat{u}_{cc} &= \frac{\frac{1}{n}\sum_{i=1}^{n} R_i Y_i}{\frac{1}{n}\sum_{i=1}^{n} R_i} \\
&\to \frac{E(RY)}{E(R)} \text{ (by the weak law of large numbers)} \\
&= \frac{E(E(RY|X))}{E(E(R|X))} \text{ (by Adam's law)} \\
&= \frac{E(E(R|X)E(Y|X))}{E(E(R|X))} \text{ (by the conditional independence of R and Y given X)} \\
&= \frac{E(\pi(X)\mu(X))}{E(\pi(X))} \text{ (by definition).}
\end{aligned}
$$

■

1.2.2

Since $Y|X = x \sim \mathcal{N}(x, 1)$,

$$\mu(X) = E(Y|X = x) = x,$$

and

$$\pi(X) = P(R = 1|X = x) = \frac{1}{1 + e^{4x-4}}.$$

Thus,

$$E(\pi(X)) = (\frac{1}{1 + e^{-4}} + \frac{1}{1 + e^{0}} + \frac{1}{1 + e^{4}})/3 = 0.5.$$

On the other hand,

$$\pi(X)\mu(X) = \frac{x}{1 + e^{4x-4}}$$

and thus

$$E(\pi(X)\mu(X)) = (\frac{0}{1 + e^{-4}} + \frac{1}{1 + e^{0}} + \frac{2}{1 + e^{4}})/3 = 0.18.$$

It follows that the complete case mean converges to

$$\frac{E(\pi(X)\mu(X))}{E(\pi(X))} = \frac{0.18}{0.5} = 0.36.$$

However,

$$E(Y) = E(E(Y|X = x)) = E(X) = 1.$$

■

1.3 IPW estimator is not equivariant

1.3.1

$$
\begin{aligned}
\hat{u}_{sn-ipw}(\mathcal{D}+a) &= \frac{1}{n}\sum_{i=1}^{n}\frac{W_iR_i(Y_i+a)}{W_iR_i} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{W_iR_iY_i}{W_iR_i} + \frac{1}{n}\sum_{i=1}^{n}\frac{aW_iR_i}{W_iR_i} \\
&= \hat{u}_{sn-ipw}(\mathcal{D}) + \frac{1}{n}\sum_{i=1}^{n} a \\
&= \hat{u}_{sn-ipw}(\mathcal{D}) + a.
\end{aligned}
$$

■

This also shows that the complete case estimator $\hat{u}_{cc}$ is equivariant because $\hat{u}_{cc}$ is a special case of $\hat{u}_{sn-ipw}$ with $W_i \equiv k, \forall i$, where k is some constant.

1.3.2

$$
\begin{aligned}
\hat{u}_{ipw}(\mathcal{D}+a) &= \frac{1}{n}\sum_{i=1}^{n}\frac{R_i(Y_i+a)}{\pi(X_i)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{R_iY_i}{\pi(X_i)} + \frac{1}{n}\sum_{i=1}^{n}\frac{aR_i}{\pi(X_i)} \\
&= \hat{u}_{ipw}(\mathcal{D}) + \frac{a}{n}\sum_{i=1}^{n}\frac{R_i}{\pi(X_i)} \\
&\neq \hat{u}_{ipw}(\mathcal{D}) + a, \ \text{(in most cases not this equality does not hold)}
\end{aligned}
$$

unless

$$
\frac{R_i}{\pi(X_i)} = 1, \forall i
$$

which means everyone is responding, or in other words $\pi(X_i) = 1, \forall i$. Since when $\pi(X_i) = 1$, R_i has to be 1 as well because $\pi(X_i) = 1$ means an 100% response rate. ■

1.3.3

Firstly, let $f(x) = E(R|X = x)$, then $f(x) = 1 \times p(R = 1|X = x) + 0 \times p(R = 0|X = x) = \pi(X)$, thus

$$
E(R|X) = \pi(X) \tag{1}
$$

Also,

$$
\hat{u}_{ipw-a}(\mathcal{D}+a) = \frac{1}{n}\sum_{i=1}^{n}\frac{R_iY_i}{\pi(X_i)} + \frac{1}{n}\sum_{i=1}^{n}\frac{aR_i}{\pi(X_i)} - a.
$$

Then,

$$\begin{aligned}
E(\frac{RY}{\pi(X)}) &= E(E(\frac{RY}{\pi(X)}|X)) \\
&= E(\frac{1}{\pi(X)}E(RY|X)) \\
&= E(\frac{1}{\pi(X)}E(R|X)E(Y|X)) \\
&= E(E(Y|X)) \text{ (from Eq.(1))} \\
&= E(Y). \text{ (from Adam's law)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(\frac{R}{\pi(X)}) &= E(E(\frac{R}{\pi(X)}|X)) \\
&= E(\frac{1}{\pi(X)}E(R|X)) \\
&= E(\frac{1}{\pi(X)}E(R|X) \\
&= E(1) \text{ (from Eq.(1))} \\
&= 1.
\end{aligned}$$

Combining the above results, we have finally

$$\begin{aligned}
E(\hat{u}_{ipw-a}(\mathcal{D}+a)) &= E(\frac{1}{n}\sum_{i=1}^{n}\frac{R_iY_i}{\pi(X_i)} + \frac{1}{n}\sum_{i=1}^{n}\frac{aR_i}{\pi(X_i)} - a) \\
&= \frac{1}{n}\sum_{i=1}^{n}E(\frac{R_iY_i}{\pi(X_i)}) + \frac{a}{n}\sum_{i=1}^{n}E(\frac{aR_i}{\pi(X_i)}) - a \\
&= E(Y) + \frac{a}{n} \times n - a \\
&= E(Y)
\end{aligned}$$

■