IPW intuition – simple example

Suppose there are 10 units where X=1 and 5 where X=0. Assume that conditional on X, treatment assignment is unrelated to potential outcomes, shown below as Y_0 and Y_1. Treatment propensity P(D|X) is strongly related to X. (The X=0 group is a lot more likely to be treated).

	N_0	N_1	N	P(D|X)	Y_1	Y_0
X=1	1	9	10	0.1	45	40
X=0	4	1	5	0.8	38	25

The full dataset is shown below (shaded cells), where Y is the observed outcome:

obs	X	D	Y	wt-att	wt-ate
1	1	0	40	0.111	1.111
2	1	0	40	0.111	1.111
3	1	0	40	0.111	1.111
4	1	0	40	0.111	1.111
5	1	0	40	0.111	1.111
6	1	0	40	0.111	1.111
7	1	0	40	0.111	1.111
8	1	0	40	0.111	1.111
9	1	0	40	0.111	1.111
10	1	1	45	1.000	10.000
11	0	1	38	1.000	1.250
12	0	1	38	1.000	1.250
13	0	1	38	1.000	1.250
14	0	1	38	1.000	1.250
15	0	0	25	4.000	5.000

Consider the simple difference in mean outcomes: $\bar{Y}_{D=1} - \bar{Y}_{D=0} = 39.4 - 38.5 = 0.9$. If we are interested in the ATT, there is selection bias because the likelihood of treatment is related to X (and, in turn, potential outcomes). If we are interested in the ATE, there is both selection bias *and* heterogeneous treatment effect bias, since the treated tend to have different treatment effects (Y_1 – Y_0) than the untreated.

Intuition: re-weighting to estimate the ATT

When using the simple difference in means to estimate the ATT, the untreated group (D=0) is a poor comparison group for the treated (D=1). Why? Because the groups are imbalanced on X (and thus potential outcomes). The untreated cases are much more likely to be X=1s than the treated cases. The untreated group doesn't "look like" the treated group.

Inverse probability weighting is a reweighting procedure that balances the two groups. Consider the following inverse probability weights (IPWs) for the ATT and ATE, respectively:

$$w_{ATT} = D_i + (1 - D_i)\frac{P(D|X)}{1 - P(D|X)}$$

$$w_{ATE} = \frac{D_i}{P(D|X)} + \frac{(1 - D_i)}{1 - P(D|X)}$$

These weights are added to the dataset above as wt-att and wt-ate. What do the ATT weights accomplish? Note there are 5 treated cases: 1 with X=1 and 4 with X=0. We would like the untreated group to "look like" the treated group. If we assign the 9 untreated X=1 cases a weight of 0.111, they will collectively represent 1 "effective" observation that corresponds to the 1 treated X=1 case. If we assign the 1 untreated X=0 case a weight of 4, it will represent 4 "effective" observations that correspond to the 4 treated X=0 cases. When applying these weights, the distribution of X among the untreated looks like that of the treated (the groups are "balanced" on X). Use these weights to calculate a difference in *weighted* means to estimate the ATT:

$$\frac{\sum_{i\in D=1} w_i Y_i}{\sum_{i\in D=1} w_i} - \frac{\sum_{i\in D=0} w_i Y_i}{\sum_{i\in D=0} w_i}$$

$$39.4 - 28 = 11.4$$

Note we are "up-weighting" untreated cases with a high probability of treatment and "down-weighting" untreated cases with a low probability of treatment.

Incidentally, you get the same (correct) answer if you subclassify / stratify by group (X=0, X=1), calculate the average treatment effect within group, and take a weighted average of these treatment effects using the number of treated in each group as weights. You <u>also</u> get the same answer if you exact match each treated observation to untreated observations with the same X.

One final bit of intuition for why the weight for untreated observations is *P(D|X)/(1 – P(D|X))*. For each X we want the effective number of untreated observations to equal the number of treated observations. In other words, we want:

$$w_{ATT,u} * \big(1 - P(D|X)\big)n_x = P(D|X)n_x$$

$$w_{ATT,u} = \frac{P(D|X)}{1 - P(D|X)}$$

The right-hand side of the top equation is the number of treated observations with a given X. The left-hand side of the equation is seeking a weight that—when applied to the untreated observations with a given X—provides an "effective" number of observations equal to the number of treated observations. It turns out this weight is *P(D|X)/(1 – P(D|X))*.

<u>Intuition: re-weighting to estimate the ATE</u>
When estimating the ATT we want our untreated group to "look like" the treated group. For the ATE, we want both groups to look like the full population. The ATE weights w_{ATE} accomplish this.

For the ATE, each treated observation gets a weight of *1/P(D|X)* while each untreated observation gets a weight of *1/(1-P(D|X))*. In the example above, there are 5 treated cases, 4 with X=0 and 1 with X=1. If we assign the 1 treated X=1 case a weight of 10, it collectively represents the 10 total observations with X=1. Likewise, if we assign the 4 treated X=0 cases a weight of 1.25, they collectively represent the 5 total observations with X=0. Along the same lines, there are 10 untreated cases, 9 with X=1 and 1 with X=0. If we assign the 9 untreated X=1 cases a weight of 1.111, they collectively represent the 10 total observations with X=1. Likewise, if we assign the 1 untreated X=0 case a weight of 5, it represents the 5 total observations with X=0. The reweighting has made both the treated and untreated groups "look like" the full population. When applying these weights, the distribution of X in both groups looks like the population. Use these weights to calculate a difference in *weighted* means to estimate the ATE:

$$42.7 - 35 = 7.7$$

One final bit of intuition for why the weight for untreated observations is *1/(1 – P(D|X))* and the weight for treated observations is *1/P(D|X)*. For each X we want the effective number of *untreated* observations to equal the total number of observations with that X. In other words, we want:

$$w_{ATE,u} * \big(1 - P(D|X)\big)n_x = n_x$$

$$w_{ATE,u} = \frac{1}{1 - P(D|X)}$$

The right-hand side of the top equation is the total number of observations with a given X. The left-hand side of the equation is seeking a weight that—when applied to the untreated observations with a given X—provides an "effective" number of observations equal to the total number of observations with that X. It turns out this weight is 1*/(1 – P(D|X))*.

Similarly, for each X we want the effective number of *treated* observations to equal the total number of observations with that X. In other words, we want:

$$w_{ATE,t} * P(D|X)n_x = n_x$$

$$w_{ATE,t} = \frac{1}{P(D|X)}$$

The right-hand side of the top equation is the total number of observations with a given X. The left-hand side of the equation is seeking a weight that—when applied to the treated observations with a given X—provides an "effective" number of observations equal to the total number of observations with that X. It turns out this weight is *1/P(D|X)*.

Connection to propensity score approaches

In the above example P(D|X) is the propensity score—the probability of treatment given X. If we didn't already know the propensity score, we could easily estimate it from these data: P(D|X) = 0.8 when X=0 and P(D|X)=0.1 when X=1. The propensity score model in this example is pretty trivial. If the conditional independence assumption holds, we can estimate treatment effects for observations balanced on X. The example above shows how this is possible using inverse probability weights (or matching, or stratification).

This simple example extends to more X's. While it seems unlikely in practice that conditional independence would hold conditional on one X, it may hold conditional on a combination of X's. Once again, we could use propensity scores P(D|X) to balance the treatment and control groups. The propensity score model likely becomes more complex and difficult to estimate as additional X's are considered—especially if some of these X's are continuous variables—since the functional form of the propensity score model is unknown. The goal of propensity score methods is usually to