Introduction to Dynamic Treatment Regimes
Homework 1, Spring 2019

1. Suppose that there are two treatment options, $\mathcal{A} = \{0, 1\}$, where A is the treatment indicator. Suppose further that there is a single, binary baseline covariate, X, taking values 0 or 1; and the outcome, Y, is also binary, taking values 0 or 1. The joint distribution of the potential outcomes $Y^*(1)$, $Y^*(0)$, X, and A is given by

$Y^*(1)$	$Y^*(0)$	X	A	probability
1	1	1	1	1/6
1	1	1	0	1/12
1	0	1	1	1/9
1	0	1	0	1/18
0	0	1	1	1/18
0	0	1	0	1/36
1	1	0	1	1/36
1	1	0	0	1/18
1	0	0	1	1/18
1	0	0	0	1/9
0	0	0	1	1/12
0	0	0	0	1/6

   (a) Show using direct calculations that the no unmeasured confounders assumption (2.13) holds in this setting.

   (b) Under SUTVA given in (2.3), deduce the distribution of the observed data (X, A, Y); that is, produce a table like that above with all eight possible combinations of values (X, A, Y) can take on together with their associated probabilities.

   (c) Show that $E(Y|A = 1) \neq E\{Y^*(1)\}$ and $E(Y|A = 0) \neq E\{Y^*(0)\}$, and give an intuitive explanation for these results.

   (d) Show that $E\{Y^*(1)\} = E\{E(Y|X, A = 1)\}$ and $E\{Y^*(0)\} = E\{E(Y|X, A = 0)\}$, where as in the notes the outer expectation is with respect to the distribution of X.

2. Assume that we have i.i.d. observed data (X_i, A_i, Y_i), $i = 1, \ldots, n$, where there are two treatment options, $\mathcal{A} = \{0, 1\}$, for which the SUTVA (2.3), the no unmeasured confounders assumption (2.13), and the positivity assumption (2.17) hold. As in the notes, define

$$\pi(X) = P(A = 1|X).$$

   Prove that

$$\frac{\sum_{i=1}^{n} \left[(1 - A_i) \frac{\pi(X_i)}{\{1 - \pi(X_i)\}} Y_i \right]}{\sum_{i=1}^{n} A_i}$$

   converges in probability to $E\{Y^*(0)|A = 1\}$, the expectation of $Y^*(0)$ among those observed to receive treatment 1.

3. In the file `hyper.txt` on the course website you will find data from an observational point exposure study like that described in the notes. The goal of the study was to assess the effectiveness of a medication developed for the treatment of hypertension (systolic blood pressure, SBP, greater than 140mmHg). The outcome of interest is the change in systolic blood pressure after six months of treatment.

   The data are on $n = 1000$ subjects diagnosed as hypertensive. Several baseline patient characteristics are available: $SBP0$, systolic blood pressure (mmHg); W, weight (kg); K, potassium level (mg/dl); Cr, creatinine level (mg/dl); and Ch, total cholesterol (mg/dl). Each subject received either the new medication ($A = 1$) or received no treatment ($A = 0$) based on patient/physician discretion. Six months after the treatment decision, the systolic blood pressure of each patient was remeasured, $SBP6$. For each subject i, the observed data are then

   $$\{SBP0_i, W_i, K_i, Cr_i, Ch_i, A_i, SBP6_i\}, \quad i = 1, \ldots, n.$$

   The outcome of interest is $Y = SBP0 - SBP6$, where larger values are considered better. All participants adhered to the treatment plan to which they were assigned and no participants dropped out of the study.

   Assuming that SUTVA, the no unmeasured confounders assumption, and the positivity assumption all hold, estimate the average causal treatment effect

   $$\delta^* = E\{Y^*(1)\} - E\{Y^*(0)\}$$

   in (2.2) using

   (i) the naive estimator $\widehat{\delta} = \overline{Y}_1 - \overline{Y}_0$ on slide 53

   (ii) the outcome regression estimator $\widehat{\delta}^*_{OR}$ in (2.19)

   (iii) the propensity score stratification estimator $\widehat{\delta}^*_S$ on slide 83

   (iv) the IPW estimator $\widehat{\delta}^*_{IPW}$ in (2.32)

   (v) the optimal, doubly robust AIPW estimator $\widehat{\delta}^*_{DR}$ in (2.34).

   Modeling choices are up to you!

   The data can be read into a dataframe in R using

```
data <- read.csv(file = "hyper.txt", header = TRUE, sep = ",")
```

4. Continue with the study of the antihypertension medication in the previous problem. Here, the observed history at baseline is $H_1 = (SBP0, W, K, Cr, Ch)$. As in the previous problem, the outcome is $Y = SBP0 - SBP6$. Consider the fixed treatment regime

   $$d(h_1) = \mathrm{I}(Ch > 220);$$

   that is, the decision rule is to administer medication if total cholesterol at baseline exceeds 220 mg/dl, otherwise, do not.

   Assume that SUTVA, the no unmeasured confounders assumption, and the positivity assumption all hold. Estimate the value of d, $\mathcal{V}(d) = E\{Y^*(d)\}$, using

   (i) the outcome regression estimator $\widehat{\mathcal{V}}_Q(d)$ in (3.7)

(ii) the IPW estimator $\widehat{\mathcal{V}}_{IPW}(d)$ in (3.13)

(iii) the alternative IPW estimator $\widehat{\mathcal{V}}_{IPW*}(d)$ in (3.14)

(iv) the optimal AIPW estimator $\widehat{\mathcal{V}}_{AIPW}(d)$ in (3.18)

Modeling choices are up to you!

5. Show that the optimal AIPW estimator $\widehat{\mathcal{V}}_{AIPW}(d)$ in (3.18) is doubly robust.