Exercise Sheet 7

Exercise 1 (IPW and M-estimation)**.** In this exercise we will study the asymptotic properties of the IPW estimator. Consider a sample $\mathcal{S} = \{(A_1, L_1, Y_1), \ldots, (A_n, L_n, Y_n)\}$ of iid replicates of (A, L, Y) such that $Y^a \perp\!\!\!\perp A \mid L$ but $Y^a \not\perp\!\!\!\perp A$, with $A \in \{0, 1\}$ and L and Y discrete (with finite support). Hereafter we will assume the propensity score $\pi(a \mid l) = P(A = a|L = l)$ is known. We also assume consistency and positivity.

(a) Write down the expression for the IPW estimator of the ATE of A on Y,

$$\hat{\text{ATE}}_{\text{IPW}} = \hat{\mu}_{\text{IPW}}(1) - \hat{\mu}_{\text{IPW}}(0).$$

(b) Prove that $\hat{\text{ATE}}_{\text{IPW}}$ is a consistent estimator of $E[Y^1 - Y^0]$, i.e.,

$$\hat{\text{ATE}}_{\text{IPW}} \xrightarrow{p} E[Y^1 - Y^0].$$

(c) Define the $\hat{\text{ATE}}_{\text{IPW}}$ estimator as an M-estimator.

(d) Prove that $\hat{\text{ATE}}_{\text{IPW}}$ is a consistent estimator of $E[Y^1 - Y^0]$ without using the same arguments as in point b).

(e) Suppose now the propensity score is unknown and that

$$\pi(l) := \pi(1 \mid l) = \text{expit}(\gamma_0 + l\gamma_1) \text{ for some } \gamma = (\gamma_0, \gamma_1) \in \Gamma \subseteq \mathbb{R}^2.$$

  (i) write down the expression for the IPW estimator of the ATE of A on Y;

  (ii) prove that $\hat{\text{ATE}}_{\text{IPW}}$ is a consistent estimator of $E[Y^1 - Y^0]$ when we posit a correctly specified model for the propensity score and we estimate β via maximum-likelihood estimation. Can you still use the same arguments as in point b)? Is such an IPW estimator a maximum-likelihood estimator?

Solution:

(a) we denote $\pi(a \mid l) := P(A = a \mid L = l)$ and we posit

$$\forall a,\ \hat{\mu}_{\text{IPW}}(a) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i 1(A_i = a)}{\pi(a|L_i)}.$$

(b) for every i define

$$S_i = \frac{Y_i 1(A_i = 1)}{\pi(1 \mid L_i)} - \frac{Y_i 1(A_i = 0)}{\pi(0 \mid L_i)},$$

which is well-defined under positivity, that is,

$$\forall l, a,\ P(L = l) > 0 \Longrightarrow P(A = a \mid L = l)$$

is well defined, with finite mean and variance. Furthermore, S_i is a measurable function of A_i, L_i, Y_i hence

$$\forall i, j, i \neq j,\ (A_i, L_i, Y_i) \perp\!\!\!\perp (A_j, L_j, Y_j) \implies S_i \perp\!\!\!\perp S_j.$$

Under exchangeability, positivity, and consistency, for every i:

$$E[S_i] = E\left[Y^1 - Y^0\right]$$

where the equality follows from the proof at pag 59 of the course and by linearity. Finally, by the weak Law of Large Numbers (WLLN)

$$\frac{\sum_{i=1}^n S_i}{n} \xrightarrow{p} E[Y^1 - Y^0]$$

but

$$\hat{\text{ATE}}_{\text{IPW}} = \frac{\sum_{i=1}^n S_i}{n}$$

and we conclude.

(c) we consider (A_i, L_i, Y_i), $i = 0, \ldots, n$ iid replicates of (A, L, Y). The M-estimator $\hat{\theta}$ is defined as the solution of

$$\frac{1}{n}\sum_{i=1}^n M(A_i, L_i, Y_i; \theta) = 0.$$

where we defined

$$M(A_i, L_i, Y_i; \theta) := \frac{A_i Y_i}{\pi(L_i)} - \frac{(1-A_i)Y_i}{1-\pi(L_i)} - \theta$$

(d) In this question, we need to check the regularity conditions given in slide 207 of the lecture slides. First note that : $\theta_0 = ATE$, $M_0(\theta) = ATE - \theta$ (from question (b)), $M_0(\theta_0) = 0$, $M_n(\hat{\theta}) = 0$.
Now let's check the regularity conditions on the slide :
(1) $M_n(\theta) - M_0(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{A_i Y_i}{\pi(L_i)} - \frac{(1-A_i)Y_i}{1-\pi(L_i)} - ATE$ which is constant with θ. So $sup_\theta |M_n(\theta) - M_0(\theta)| = \left|\frac{1}{n}\sum_{i=1}^n \frac{A_i Y_i}{\pi(L_i)} - \frac{(1-A_i)Y_i}{1-\pi(L_i)} - ATE\right| \to^p 0$ by the law of large numbers.
(2) Since Y is categorical, Θ is compact. M is a linear function of θ so it is continuous. Finally, the solution is unique (there is a closed-form solution).
(3) This is trivial because $M_n(\hat{\theta}) = 0$. (In fact, this third regularity condition should be checked carefully only if there is no closed-form solution; and we need to rely on an optimization algorithm to estimate $\hat{\theta}$)

(e) (i)

$$\hat{\text{ATE}}_{\text{IPW}} = \frac{1}{n}\sum_{i=1}^n \frac{Y_i 1(A_i = 1)}{\hat{\pi}(L_i)} - \frac{Y_i 1(A_i = 0)}{1 - \hat{\pi}(L_i)}$$

where $\hat{\pi}(\cdot)$ is an estimator of $\pi(\cdot)$.

(ii) We cannot use the same arguments as in point b). To see this, notice that $\hat{\pi}$ is a solution of

$$\sum_{i=1}^n \begin{pmatrix} 1 \\ L_i \end{pmatrix} \left(A_i - \frac{\exp(\gamma_1 + \gamma_2 L_i)}{1 + \exp(\gamma_1 + \gamma_2 L_i)}\right) = 0, \ (\gamma_1, \gamma_2) \in \Gamma \subseteq \mathbb{R}^2$$

which depends on $((A_1, L_1), \ldots, (A_n, L_n))$ for every every n. The finite sample distribution of $\hat{\pi}$ often cannot be computed in closed form and, when it is the

case, for large n, we rely on its asymptotic distribution. However, this is i) only an approximation when we actually consider finite samples and ii) we would need to clarify the relation between the sample used to fit the propensity score model and the sample used in the IPW estimator. Indeed, we cannot treat samples of $(A_i, \hat{\pi}_i, Y_i)$ as iid (without making additional assumptions) and we cannot use the same arguments as in point b).
To prove that $\hat{\text{ATE}}_{\text{IPW}}$ is a consistent estimator of $E[Y^1 - Y^0]$ we define M and a new parameter space Γ as

$$M(A, L, Y; \gamma) : \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} \longmapsto \begin{pmatrix} \frac{AY}{\text{expit}(\gamma_2+\gamma_3 L)} - \frac{(1-A)Y}{1-\text{expit}(\gamma_2+\gamma_3 L)} - \gamma_1 \\ A - \text{expit}(\gamma_2 + \gamma_3 L) \\ AL - \text{expit}(\gamma_2 + \gamma_3 L)L \end{pmatrix}, \gamma = (\gamma_1, \dots, \gamma_3) \in \Gamma \subseteq \mathbb{R}^3,$$

where the two last equations (M_2 and M_3) correspond to the equations used for estimating the propensity score.
To show the consistency of the estimator, we can again prove the regularity conditions of slide 207.

(i) The first regularity assumption (uniform convergence) is the harder to prove. Let's sketch the proof briefly for the first part of the equation for γ_1 only : the idea is to write $\frac{1}{n}\sum \frac{AY}{expit(\gamma_2+\gamma_3 L)}$ as $\frac{1}{n}\sum \frac{AY}{expit(\gamma_2+\gamma_3 L)} - \frac{AY}{\pi(L)} + \frac{1}{n}\sum \frac{AY}{\pi(L)}$. We can easily show that $\frac{1}{n}\sum \frac{AY}{\pi(L)}$ converges uniformly to $E[Y^1]$ with the same arguments we used in question d. We then need to prove that the error term $\frac{1}{n}\sum \frac{AY}{expit(\gamma_2+\gamma_3 L)} - \frac{AY}{\pi(L)}$ converges uniformly to 0. To do so, we use that (i) AY is bounded by $max(|Y|) := M$ (Y is discrete finite), (ii) $expit(\gamma_2+\gamma_3 L)$ converges uniformly to $\pi(L)$ (under some regularity conditions, notably the compactness, which can be proved because the propensity score is bounded away from 0 and 1), and that (iii) the inverse function is $\frac{1}{\epsilon^2}-$Lipschitz continuous on $(\epsilon, 1-\epsilon)$ for any $\epsilon > 0$[1] to prove that

$$sup_\gamma \left| \frac{1}{n}\sum \frac{AY}{expit(\gamma_2+\gamma_3 L)} - \frac{AY}{\pi(L)} \right| < \frac{M}{\epsilon^2} sup_\gamma |expit(\gamma_2+\gamma_3 L) - \pi(L)|$$

where $sup_\gamma|expit(\gamma_2+\gamma_3 L) - \pi(L)|$ converges to 0 by uniform convergence of the logistic regression estimator.

(ii) The regularity conditions 2 and 3 follow easily from the properties the MLE estimator for logistic regression and what was done above.

$\hat{\text{ATE}}_{\text{IPW}}$ is not a maximum-likelihood estimator since it is not the maximizer of a given likelihood function.

Exercise 2 (A comparison of variance)**.** (From [1], Homework 2)
Suppose that the outcome and propensity model are known. Consider two estimators for the average response: $\frac{1}{n}\sum_{i=1}^n Y_i^{a=1}$ and $\frac{1}{n}\sum_{i=1}^n \frac{A_i Y_i^{a=1}}{\pi(A_i|L_i)}$.[2]

[1] we can find one such ϵ such that $\epsilon < \pi(L) < 1-\epsilon$ because of the positivity assumption (the propensity score is bounded away from 0 and 1)

[2] The first estimator is an estimator that is typically impossible to compute because all the counterfactuals are not observed. However, in this exercise we have assumed that $Y_i^{a=1}$ is observed.

(a) By assuming conditional exchangeability $Y_i^a \perp\!\!\!\perp A_i \mid L_i$, show that the first has lower variance than the second (that is, we pay some penalty for not observing all subjects in the data set being treated).

Hint: Show that the second estimator can be written as the first plus something else, and then demonstrate that the two terms are uncorrelated.

(b) Compute the difference in variance between the estimators in (a) if A is randomized with probability $P(A=1) = \frac{1}{2}$ (i.e. $\pi = \frac{1}{2}$)

Solution:

(a) We can write the second estimator as

$$\frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i^{a=1}}{\pi(A_i \mid L_i)} = \frac{1}{n}\sum_{i=1}^{n} Y_i^{a=1} + \frac{1}{n}\sum_{i=1}^{n} Y_i^{a=1}\left(\frac{A_i}{\pi(A_i \mid L_i)} - 1\right) .$$

Next, we want to show that the first sum on the right hand side is not correlated to the second sum, that is:

$$\mathrm{Cov}\left(Y_i^{a=1}, Y_i^{a=1}\left(\frac{A_i}{\pi(A_i \mid L_i)} - 1\right)\right) = 0 .$$

Since

$$\begin{aligned}
& E\left[Y_i^{a=1}\left(\frac{I(A_i=1)}{\pi(A_i \mid L_1)} - 1\right) \,\middle|\, Y_i^{a=1}\right] \\
&= Y_i^{a=1}\left(E\left[\frac{I(A_i=1)}{P(A_i=1 \mid L_i)} \,\middle|\, Y_i^{a=1}\right] - 1\right) \\
&= Y_i^{a=1}\left(E\left[E\left[\frac{I(A_i=1)}{P(A_i=1 \mid L_i)} \,\middle|\, L_i, Y_i^{a=1}\right] \,\middle|\, Y_i^{a=1}\right] - 1\right) \\
&= Y_i^{a=1}\left(E\left[\frac{P(A_i=1 \mid L_i, Y_i^{a=1})}{P(A_i=1 \mid L_i)} \,\middle|\, Y_i^{a=1}\right] - 1\right) \\
&\overset{Y_i^a \perp\!\!\!\perp A_i \mid L_i}{=} Y_i^{a=1}\left(E\left[\frac{P(A_i=1 \mid L_i)}{P(A_i=1 \mid L_i)} \,\middle|\, Y_i^{a=1}\right] - 1\right) \\
&= 0 ,
\end{aligned}$$

and

$$E\left[Y_i^{a=1}\left(\frac{I(A_i=1)}{\pi(A_i \mid L_1)} - 1\right)\right] = E[Y^{a=1}] - E[Y^{a=1}] = 0 .$$

Hence

$$\begin{aligned}
& \mathrm{Cov}\left(Y_i^{a=1}, Y_i^{a=1}\left(\frac{A_i}{\pi(A_i \mid L_i)} - 1\right)\right) \\
&= E\left[Y_i^{a=1} \underbrace{E\left[Y_i^{a=1}\left(\frac{A_i}{\pi(A_i \mid L_i)} - 1\right) \mid Y_i^{a=1}\right]}_{=0}\right] - E[Y_i^{a=1}] \underbrace{E\left[Y_i^{a=1}\left(\frac{I(A_i=1)}{\pi(A_i \mid L_1)} - 1\right)\right]}_{=0} = 0
\end{aligned}$$

and therefore

$$\mathrm{Var}\left(Y_i^{a=1} + Y_i^{a=1}\left(\frac{A_i}{\pi(A_i \mid L_i)} - 1\right)\right) = \mathrm{Var}\left(Y_i^{a=1}\right) + \mathrm{Var}\left(Y_i^{a=1}\left(\frac{A_i}{\pi(A_i \mid L_i)} - 1\right)\right).$$

But

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \frac{A_i Y_i^a}{\pi(A_i \mid L_i)}\right) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_i^{a=1}\right) + \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_i^{a=1}\left(\frac{A_i}{\pi(A_i \mid L_i)} - 1\right)\right) \\ &\geq \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_i^{a=1}\right) \quad .\end{aligned}$$

(b)

$$\begin{aligned}\mathrm{Var}\left(Y_i^{a=1}\left(\frac{I(A_i = 1)}{\pi(A_i \mid L_i)} - 1\right)\right) &= E\left[\left(Y_i^{a=1}\left(\frac{I(A_i = 1)}{\pi(A \mid L)} - 1\right)\right)^2\right] \\ &= E\left[\frac{Y_i^2 A_i^2}{\pi^2}\right] - 2\underbrace{E[(Y_i^{a=1})^2]}_{=E\left[\frac{Y_i^2 A_i}{\pi(A_i \mid L_i)}\right]} + E\left[\frac{Y_i^2 A}{\pi}\right] \\ &\overset{A_i^2 = A_i}{=} E\left[\frac{Y_i^2 A_i}{\pi^2(A_i \mid L_i)} - \frac{Y_i^2 A_i}{\pi(A_i \mid L_i)}\right]\end{aligned}$$

and therefore

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n \frac{A_i Y_i^a}{\pi(A_i \mid L_i)}\right) - \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_i^{a=1}\right) &= \frac{1}{n}\left(E\left[\frac{Y_i^2 A(1 - \pi(A_i \mid L_i))}{\pi^2(A_i \mid L_i)}\right]\right) \\ &= \frac{2}{n} \cdot E[Y_i^2 A] \ .\end{aligned}$$

In practice, we cannot compute $\frac{1}{n}\sum_{i=1}^n Y_i^{a=1}$ because we only observe the counterfactual outcome $Y_i^{a=1}$ in (nearly) half of the individuals.[3]

Exercise 3 (Stabilized IPW estimators)**.** (Technical Points 12.1 and 12.2 in [2]) Let A, L, Y denote treatment, baseline covariates and outcome respectively and suppose the usual assumptions of conditional exchangeability, positivity and consistency hold.

(a) Show that we can identify $E[Y^a]$ from

$$E[Y^a] = \frac{E\left[\frac{I(A=a)Y}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)}{\pi(A|L)}\right]} \ .$$

This form of the identification formula motives a modified version of the IPW estimator called the Hajek estimator (or stabilized IPW estimator):

$$\hat{\mu}_{STIPW}(a) = \frac{\frac{1}{n}\sum_{i=1}^n \frac{I(A_i=a)Y_i}{\pi(A_i|L_i;\gamma)}}{\frac{1}{n}\sum_{i=1}^n \frac{I(A_i=a)}{\pi(A_i|L_i;\gamma)}} \ . \tag{1}$$

[3]This is because the exact number of individuals who receive treatment $A = 1$ is binomially distributed with n trials and $p = \frac{1}{2}$.

(b) Show that

$$E[Y^a] = \frac{E\left[\frac{I(A=a)Yg(A)}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)g(A)}{\pi(A|L)}\right]}$$

and that

$$\hat{\mu}_{STIPW} = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}(A_i)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)Y_i}{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}(A_i)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)},$$

where $g(A)$ is a function of A, and is consistently estimated by $\hat{g}(A)$. We refer to $\frac{g(A)}{\pi(A|L)}$ as stabilized weights because they are, in settings where rely on parametric assumptions, often smaller than the regular IPW weights $\frac{1}{\pi}$, and can thus give rise to estimators with a smaller variance.

Solution:

(a) The expectation in the denominator is 1, because

$$\begin{aligned} E\left[\frac{I(A=a)}{\pi(A \mid L)}\right] &= E\left[\frac{I(A=a)}{P(A=a \mid L)}\right] \\ &= E\left[\frac{1}{P(A=1 \mid L)} E\left[I(A=a) \,\middle|\, L\right]\right] \\ &= E\left[\frac{P(A=1 \mid L)}{P(A=1 \mid L)}\right] \\ &= 1\ . \end{aligned}$$

(b) Then,

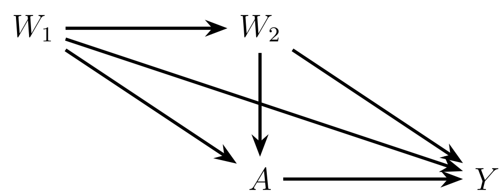
$$\begin{aligned} \frac{E\left[\frac{I(A=a)Yf(A)}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)f(A)}{\pi(A|L)}\right]} &= \frac{E\left[\frac{I(A=a)Yf(a)}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)f(a)}{\pi(A|L)}\right]} \\ &= \frac{f(a)E\left[\frac{I(A=a)Y}{\pi(A|L)}\right]}{f(a)E\left[\frac{I(A=a)}{\pi(A|L)}\right]} \\ &= E[Y^a]\ . \end{aligned}$$

Likewise,

$$\begin{aligned} \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{f}(A_i)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)Y_i}{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{f}(A_i)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)} &= \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{f}(a)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)Y_i}{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{f}(a)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)} \\ &= \frac{\hat{f}(a)\frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=a)Y_i}{\pi(A_i|L_i;\gamma)}}{\hat{f}(a)\frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=a)}{\pi(A_i|L_i;\gamma)}} \\ &= \hat{\mu}_{STIPW}(a)\ . \end{aligned}$$

Exercise 4 (Exploring the IPW estimator)**.** (Based on Lab 4 of [3])

In this exercise we will implement the IPW and Hajek estimators numerically in R in order to explore their efficiency in cases with near violations of positivity. Consider treatment A and outcome Y with baseline covariates W_1, W_2 in the dataset `stabilized_weights.csv`, and suppose these satisfy the causal model below: The data was generated by drawing



$n = 5000$ i.i.d. samples from the distributions

$$\begin{aligned} W_1 &\sim Ber\left(p = \frac{1}{2}\right) \\ W_2 &\sim Multinom\,(1; (0.125, 0.375, 0.375, 0.125)) \\ A &\sim Ber\left(p = \text{logit}^{-1}(-1.3 - 3W_1 + 3W_2)\right) \\ Y &\sim Ber\left(p = \text{logit}^{-1}(-2 - 2W_1 + 3W_2 + 3A + 2AW_2)\right) \\ Y^{a=1} &\sim Ber\left(p = \text{logit}^{-1}(-2 - 2W_1 + 3W_2 + 3 \cdot 1 + 2 \cdot 1 \cdot W_2)\right) \\ Y^{a=0} &\sim Ber\left(p = \text{logit}^{-1}(-2 - 2W_1 + 3W_2 + 3 \cdot 0 + 2 \cdot 0 \cdot W_2)\right), \end{aligned}$$

subject to the constraint

$$Y = Y^{a=1} I(A = 1) + Y^{a=0} I(A = 0)\ .$$

The true effect is given by $E[Y^{a=1} - Y^{a=0}] \approx 0.26$ (computed by evaluating $\frac{1}{n'} \sum_{i=1}^{n'} (Y_i^1 - Y_i^0)$ in a larger realization of the data with $n' = 100000$) .

(a) Import the dataset `stabilized_weights.csv` into R and use the `glm` command to perform the following logistic regression for the treatment mechanism $\pi(A \mid L)$:

$$\text{logit}\ \pi(A \mid L; \gamma) = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2\ .$$

Plot the empirical cumulative distribution function of the IPW weights $\frac{1}{\pi(A_i \mid W_{1,i}, W_{2,i})}$ and use the weights to evaluate the IPW estimator

$$\hat{\mu}_{IPW} = \frac{1}{n} \sum_{i=1}^{n} \frac{I(A_1 = a) Y_i}{\pi(A_i \mid W_{1,i}, W_{2,i}; \gamma)}\ .$$

(b) Compute $\hat{\mu}_{IPW}$ with truncated weights $\frac{I(\pi \leq 10)}{\pi} + 10 \cdot I(\pi > 10)$ instead of the weights $\frac{1}{\pi}$ in part (a).

(c) Evaluate the stabilized IPW estimator given by Eq. 1 using the weights as in part (a).

(d) Estimate the variance of the estimators in parts (a)-(d) by drawing $R = 5000$ different realizations of a population with $n = 5000$ i.i.d. individuals from the data generating mechanism outlined above.

Solution:

The IPW weights are computed as follows:

FIGURE 1. CDF of IPW weights

```
ObsData<- read.csv('R/stabilized_weights.csv') # Importing data
n<- nrow(ObsData) # Defining the size of the population

# Performing logistic regression
pi.reg<- glm(A ~ W1 +W2, family="binomial", data=ObsData)
pi.reg$coef # Displaying the regression coefficients
## (Intercept)          W1          W2
##   -1.332659   -3.224437    3.184136
# Predicting \pi(A|W1,W2; \gamma) using the regression model
pred.pi1 <- predict(pi.reg, type= "response")
pred.pi0 <- 1 - pred.pi1
pi <- rep(NA, n)

# Evaluating \pi(A|W1,W2; \gamma) for each individual
pi[ObsData$A==1] <- pred.pi1[ObsData$A==1]
pi[ObsData$A==0] <- pred.pi0[ObsData$A==0]

# Computing the weights
wt<- 1/pi
```

We also plot the CDF of the weights (displayed in Fig. 1):

```
library(latex2exp)
# Plotting the CDF of the IPW weights
plot(ecdf(wt), main='', ylab=TeX('$P(W\\leq w)$'), xlab=TeX('$w$'))
```

Computing μ_{IPW}:

```
IPW <- mean( wt*as.numeric(ObsData$A==1)*ObsData$Y) -
  +   mean( wt*as.numeric(ObsData$A==0)*ObsData$Y)
IPW
## [1] 0.1974928
```

Computing μ_{IPW} with truncated weights:

```
wt.trunc<- wt
wt.trunc[ wt.trunc>10] =10
mean( wt.trunc*as.numeric(ObsData$A==1)*ObsData$Y) -
  +   mean( wt.trunc*as.numeric(ObsData$A==0)*ObsData$Y)
## [1] 0.5437832
```

Computing μ_{STIPW}:

```
mean( wt*as.numeric(ObsData$A==1)*ObsData$Y)/mean( wt*
  +  as.numeric(ObsData$A==1)) - mean( wt*
```

```
  +  as.numeric(ObsData$A==0)*ObsData$Y)/mean( wt*
  +  as.numeric(ObsData$A==0))
## [1] 0.2783772
```

The mean and variance of the estimators can be estimated using the following code:

```
genData<- function(n){
     W1 <- rbinom(n, size=1, prob=.5)
     W2 <- rbinom(n, size=3, prob=.5)
     A <- rbinom(n, size =1, prob=plogis(-1.3 - 3*W1 +3*W2))
     Y<- rbinom(n, size=1, prob= plogis(-2 - 2*W1 +3*W2 +3*A+ 2*A*W2 ))
     Y.1<- rbinom(n, size=1, prob= plogis(-2 - 2*W1 +3*W2 +3*1+ 2*1*W2 ))
     Y.0<- rbinom(n, size=1, prob= plogis(-2 - 2*W1 +3*W2 +3*0+ 2*0*W2 ))
data.frame(W1,W2, A, Y, Y.1,Y.0) }

# number of iterations
set.seed(259)
R<- 5000
# matrix for estimates from IPW
estimates<- matrix(, nrow = R, ncol = 3)
for(r in 1:R){
  # 0. redraw the data
  NewData<- genData(n)
  # 1.  Estimate the propensity \pi(A | W1,W2)
  pi.reg<- glm(A ~ W1 +W2, family="binomial", data=NewData)

  # 2. # predicted probability of treatment
   pred.pi1 <- predict(pi.reg, type= "response")
  # predicted probability of treatment
  pred.pi0 <- 1 - pred.pi1
  # we need the predicted prob of the observed treatment,
  # given covariates.

  # create an empty vector
  pi <- rep(NA, n)
  # for individuals with A=1, \pi = P(A=1 | W1,W2)
  pi[NewData$A==1] <- pred.pi1[NewData$A==1]
  # for individuals with A=0, gAW = P(A=0 | W1,W2)
  pi[NewData$A==0] <- pred.pi0[NewData$A==0]

  # 3. Each subject gets a weight equal to 1/pi
  wt<- 1/pi
  # 4. # IPW estimator
  IPW<- mean( wt*as.numeric(NewData$A==1)*NewData$Y) -
    +  mean( wt*as.numeric(NewData$A==0)*NewData$Y)
```

```
  # 6.  truncate  weights at 10
  wt.trunc<- wt
  wt.trunc[ wt.trunc>10] =10
  # evaluate the IPW estimand with the truncated weights
  IPW.tr<- mean( wt.trunc*as.numeric(NewData$A==1)*NewData$Y) -
  + mean( wt.trunc*as.numeric(NewData$A==0)*NewData$Y)

  # 7. Stabilized IPW estimator
  STIPW<-mean(wt*as.numeric(NewData$A==1)*NewData$Y)/
  +  mean(wt*as.numeric(NewData$A==1))-
  + mean( wt*as.numeric(NewData$A==0)*NewData$Y)/
  + mean( wt*as.numeric(NewData$A==0))

  estimates[r,]<-c(IPW, IPW.tr, STIPW)
}

# Average value of the estimates over R repetitions
colMeans(estimates)

# Variance
diag(var(estimates))
```

which yields the following estimates:

	IPW	Truncated IPW	Stabilized IPW
Mean	0.2625547	0.535427	0.2713136
Variance	0.03286365	0.000130988	0.002740077

Exercise 5. (Logistic regression model) We would like to estimate the effects of a pesticide on the statue of stink bugs in a farm. We observe the statue of n stink bugs, and let Z_i be the binary outcome of the experiment for the stink bug i. Y is the sum of Z_i and corresponds to the number of stink bugs that are observed to be alive after the termination of experiment.

(a) What distribution is reasonable to assume for Y if each stink bug is given the same dosage of pesticide? What assumption does that require making on the Z_i?

(b) Now assume stink bug i is given a specific dosage of pesticide, namely $x_i > 0$. Using logistic model, state the probability that a bug survives in terms of the constant β_0 and linear coefficient β_1.

(c) Describe how to fit the parameters of the linear model given data Z_i.

(d) Recall from the statistics course that for large sample size n, the variance of the MLE estimator is given by the inverse of the Fisher information (In other words, the variance achieves Cramer-Rao bound asymptotically). Assume $\beta_0 = 0$ and calculate the Fisher information and find an asymptotic estimate for the variance of $\hat{\beta}_1$.

(e) What assumptions were required to write down the likelihood function?

Solution:

(a) We can assume Z_is are i.i.d. Bernoulli variables and thus Y have Binomial distribution.

(b) We can assume Z_i are independent Bernoulli variables such that

$$\mathbb{E}(Z_i) = g^{-1}(\beta_0 + x_i\beta_1), \quad g(\mu) = \log(\frac{\mu}{1-\mu}).$$

Because $x_i > 0$, and we can expect that increasing the dosage of pesticide, would decrease the probability that fruit flies survive, thus β_1 should be negative.

(c) The likelihood for the model is

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} z_i \log p(x_i, \beta) + (1 - z_i) \log(1 - p(x_i, \beta)) \\ &= \sum_{i=1}^{n} \{z_i(\beta_0 + \beta_1 x_i) - \log(1 + e^{\beta_0 + \beta_1 x_i})\} \end{aligned}$$

where $p(x_i, \beta) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$. Thus we want to solve the equation $\frac{\partial l(\beta)}{\partial \beta} = 0$. We can calculate $\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T}$, for $\beta^T = (\beta_0, \beta_1)$, and starting from some initial β, use the Newton-Raphson method with the following iterations:

$$\beta^{new} = \beta^{old} - (\frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^T})^{-1} \frac{\partial l(\beta)}{\partial\beta}.$$

(d)

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^{n} x_i\big(z_i - \frac{e^{\beta_0+\beta_1 x_i}}{1 + e^{\beta_0+\beta_1 x_i}}\big).$$

We compute the Fisher information $I(\beta) = -\mathbb{E}[\frac{\partial^2}{\partial\beta^2} l(\beta)]$ by calculating the second derivatives of l:

$$\frac{\partial^2}{\partial\beta^2} l(\beta) = -\sum_{i=1}^{n} x_i^2 \frac{e^{\beta_0+\beta_1 x_i}}{(1 + e^{\beta_0+\beta_1 x_i})^2}.$$

Thus for large sample size n, one can estimate the variance of $\hat{\beta}$ by the inverse of the calculated Fisher information above.

(e) The logistic regression model is correctly specified that is, when the Y_i's are truly independent random variables with distribution Bernoulli(p_i), where the logit(p_i) is the same linear combination of the covariates x_i.

References

[1] David M. Vock. PubH 7485 & 8485: Methods for Causal Inference (University of Minnesota School of Public Health).

[2] Miguel Hernan and James M. Robins. *Causal Inference*. Chapman & Hall, 2018.

[3] Maya L. Petersen and Laura B. Balzer. Labs & Assignments.