Inverse Probability Weighting

The propensity score for binary exposure, denoted $e(X)$, is defined via the observational conditional model for Z given confounders X, as

$$e(x) = f^{\mathcal{O}}_{Z|X}(1|x) = \Pr^{\mathcal{O}}_{Z|X}[Z = 1|X = x]$$

with $e(X)$ being the corresponding random variable. We have that in the binary treatment case

$$\mu(0) = \frac{\mathbb{E}^{\mathcal{O}}_{X,Y,Z}\left[\dfrac{(1-Z)Y}{1-e(X)}\right]}{\mathbb{E}^{\mathcal{O}}_{X,Y,Z}\left[\dfrac{(1-Z)}{1-e(X)}\right]} \qquad \mu(1) = \frac{\mathbb{E}^{\mathcal{O}}_{X,Y,Z}\left[\dfrac{ZY}{e(X)}\right]}{\mathbb{E}^{\mathcal{O}}_{X,Y,Z}\left[\dfrac{Z}{e(X)}\right]}$$

We deduce that suitable estimators following from the importance sampling results are

$$\widehat{\mu}_{IPW}(0) = \frac{\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{(1-Z_i)Y_i}{1-e(X_i)}}{\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{(1-Z_i)}{1-e(X_i)}} \qquad \widehat{\mu}_{IPW}(1) = \frac{\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{Z_iY_i}{e(X_i)}}{\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{Z_i}{e(X_i)}}.$$

Alternative estimators are

$$\widetilde{\mu}_{IPW}(0) = \frac{1}{n}\sum_{i=1}^{n}\frac{(1-Z_i)Y_i}{1-e(X_i)} \qquad \widetilde{\mu}_{IPW}(1) = \frac{1}{n}\sum_{i=1}^{n}\frac{Z_iY_i}{e(X_i)}.$$

```
#Data simulation
library(mvnfast)
set.seed(23987)
n<-5000
k<-10
Mu<-sample(-5:5,size=k,rep=T)/2
Sigma<-0.1*0.8^abs(outer(1:k,1:k,'-'))
X<-rmvn(n,mu=Mu,Sigma)
al<-c(-4,rep(1,6),rep(-2,4))
Xal<-cbind(1,X)
expit<-function(x){return(1/(1+exp(-x)))}
ps.true<-expit(Xal %*% al)
#hist(ps.true)
Z<-rbinom(n,1,ps.true)
table(Z)

+ Z
+    0    1
+ 2187 2813

be<-rep(0,k+2)
be[1:5]<-c(2,1,-2.0,2.2,3.6)
Xm<-cbind(1,Z,X)
muval<-as.vector(Xm %*% be)
sig<-2
Y<-rnorm(n,muval,sig)
eX<-ps.true
w<-(1-Z)/(1-eX)+Z/eX
```

We may perform direct calculation:

```
w0<-(1-Z)/(1-eX)
W0<-w0/sum(w0)
w1<-Z/eX
```

```
W1<-w1/sum(w1)

#Mu tilde
mean(w1*Y)-mean(w0*Y)

+ [1] 0.7306851

#Mu hat
sum(W1*Y)-sum(W0*Y)

+ [1] 1.024291
```

We may also use regression methods:

```
#Unadjusted
coef(summary(lm(Y~Z)))

+             Estimate Std. Error  t value     Pr(>|t|)
+ (Intercept) 5.406811 0.05091986 106.1828 0.000000e+00
+ Z           1.466253 0.06788712  21.5984 5.509346e-99

#IPW
coef(summary(lm(Y~Z,weights=w)))

+             Estimate Std. Error   t value     Pr(>|t|)
+ (Intercept) 5.720303 0.04691649 121.92523 0.000000e+00
+ Z           1.024291 0.06716312  15.25079 2.296454e-51
```

```
#We may also use a re-weighted pseudo sample
id<-sample(1:n,size=n,prob=w,rep=T) #Resample an unconfounded data set
coef(summary(lm(Y[id]~Z[id])))

+             Estimate Std. Error   t value     Pr(>|t|)
+ (Intercept) 5.709884 0.04715218 121.09480 0.000000e+00
+ Z[id]       1.098114 0.06760908  16.24211 7.527951e-58
```

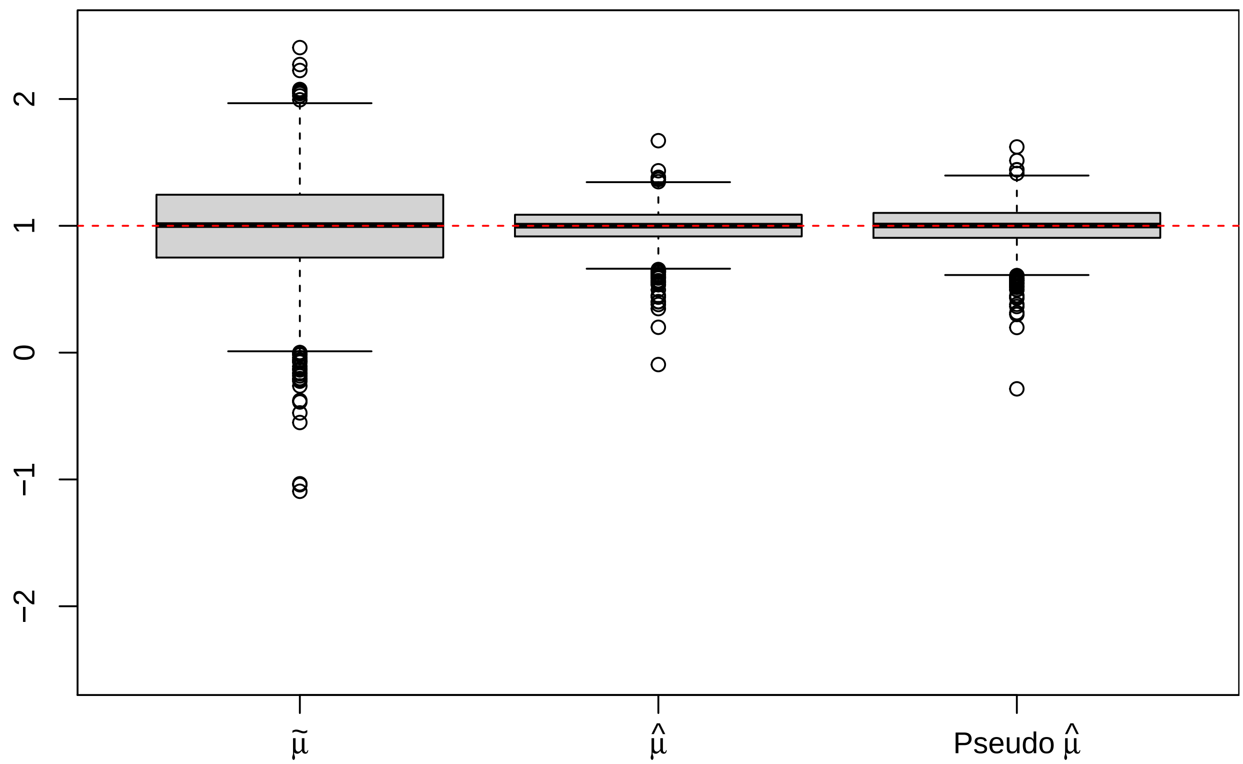
```
#Monte Carlo study
nreps<-2000
ests<-matrix(0,nrow=nreps,ncol=3)
for(irep in 1:nreps){

    X<-rmvn(n,mu=Mu,Sigma)
    Xal<-cbind(1,X)
    ps.true<-expit(Xal %*% al)
    Z<-rbinom(n,1,ps.true)
    Xm<-cbind(1,Z,X)
    muval<-as.vector(Xm %*% be)
    Y<-rnorm(n,muval,sig)
    eX<-ps.true
    w<-(1-Z)/(1-eX)+Z/eX

    w0<-(1-Z)/(1-eX)
    W0<-w0/sum(w0)
    w1<-Z/eX
    W1<-w1/sum(w1)
    ests[irep,1]<-mean(w1*Y)-mean(w0*Y)
    ests[irep,2]<-sum(W1*Y)-sum(W0*Y)

    id<-sample(1:n,size=n,prob=w,rep=T)
    ests[irep,3]<-coef(lm(Y[id]~Z[id]))[2]
}
```

```r
par(mar=c(4,4,2,0))
boxplot(ests,ylim=range(-2.5,2.5),
    names=c(expression(tilde(mu)),expression(hat(mu)),expression(paste('Pseudo ',hat(mu)))))
abline(h=1,lty=2,col='red')
```




```r
#Bias
sqrt(n)*apply(ests-1,2,mean)

+ [1] -1.4325601 -0.2229784 -0.3145703

#Variance
n*apply(ests,2,var)

+ [1] 935.8373 103.3114 128.0590
```