The Average Treated Effect on the Treated

The *average treatment effect on the treated* (ATT) is

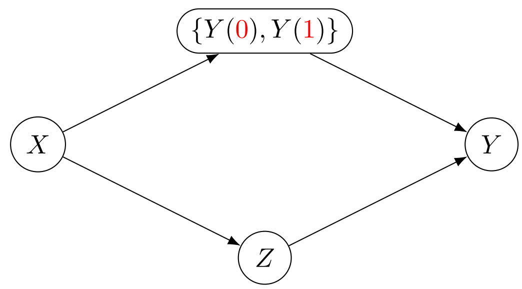
$$\mathbb{E}[Y(1) - Y(0) \mid Z = 1]$$

that is, the ATT aims to identify the causal effect on intervening to change $Z = 0$ to $Z = 1$ but *only* in the subpopulation of individuals who are *observed* to receive treatment. Note that

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y(1) - Y(0) \mid Z = 0]\,\Pr[Z = 0] + \mathbb{E}[Y(1) - Y(0) \mid Z = 1]\,\Pr[Z = 1]$$

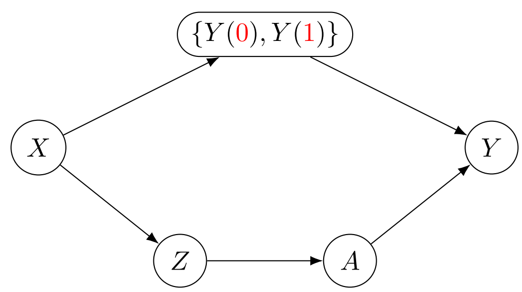
In this calculation, we imagine

- the observational distribution $f^{\mathcal{O}}_{X,Y,Z}$ generating the observed data $\{(x_i, y_i, z_i), i = 1, \ldots, n\}$
- in the subgroup observed to have $Z = 1$, we then consider a second (hypothetical) experimental intervention to change Z to z which over-rides the original Z if $\mathsf{z} = 0$,
- we then consider comparison of outcomes between the two hypothetical subgroups indexed by z.



$$f_X(x)\; f_{Z|X}(z|x)\; f_{Y(0),Y(1)|X}(y_0, y_1|x)\; f_{Y|Z,Y(0),Y(1)}(y|z, y_0, y_1)$$

For the ATT, we can represent the quantity of interest using a modified DAG that proposes a second hypothetical binary treatment, A. We allow Z to cause A, and then allow Z to act as a selection mechanism, but ensure that we have the conditional independence of X and A given Z, $X \perp\!\!\!\perp A \mid Z$,



$$f_X(x)\; f_{Z|X}(z|x)\; f_{Y(0),Y(1)|X}(y_0, y_1|x)\; f_{A|Z}(a|z)\; f_{Y|A,Y(0),Y(1)}(y|a, y_0, y_1)$$

This is the new 'experimental' distribution $\mathcal{E}$.

$$f^{\mathcal{E}}_{Y|A,Y(0),Y(1)}(y|a, y_0, y_1) = \begin{cases} 1 & y = (1-a)y_0 + ay_1 \\ 0 & \text{otherwise} \end{cases}.$$

As A is binary, the model $f_{A|Z}(a|z)$ must take the form

$$f^{\mathcal{E}}_{A|Z}(a|z) = p_z^a (1 - p_z)^{1-a} \qquad a, z \in \{0, 1\}$$

for $0 \leq p_z \leq 1$ for $z = 0, 1$. We can then express the ATT via the new DAG as

$$\mathbb{E}^{\mathcal{E}}_{Y|A,Z}[Y|A = 1, Z = 1] - \mathbb{E}^{\mathcal{E}}_{Y|A,Z}[Y|A = 0, Z = 1].$$

We have that from the DAG that

$$f^{\mathcal{E}}_{Y|A,X,Z}(y|a,x,z) \equiv f^{\mathcal{E}}_{Y|A,X}(y|a,x)$$

and hence as before

$$f^{\mathcal{E}}_{Y|A,X}(y|a,x) = \int f^{\mathcal{E}}_{Y|A,Y(0),Y(1)}(y|a,y_0,y_1) f^{\mathcal{E}}_{Y(0),Y(1)|A,X}(y_0,y_1|a,x)\, dy_0\, dy_1 = f^{\mathcal{E}}_{Y(a)|X}(y|x).$$

Also from the DAG, $X \perp\!\!\!\perp A \mid Z$, so for all a, x, z

$$f^{\mathcal{E}}_{X|A,Z}(x|a,z) \equiv f^{\mathcal{E}}_{X|Z}(x|z).$$

For $\mathsf{a} = 0, 1$, we therefore have

$$\mathbb{E}^{\mathcal{E}}_{Y|A,Z}[Y \mid A = \mathsf{a}, Z = z] = \iint y\, f^{\mathcal{E}}_{Y|A,X}(y \mid \mathsf{a}, x) f^{\mathcal{E}}_{X|Z}(x \mid z)\, dy\, dx.$$

Note that in this integral there is a potential *incompatibility* in the conditioning between

$$f^{\mathcal{E}}_{Y|A,X}(y \mid \mathsf{a}, x) \qquad \text{and} \qquad f^{\mathcal{E}}_{X|Z}(x \mid z)$$

when we try to write the integral in terms of the data generating mechanism. As before, choosing the form of the outcome conditional model

$$f^{\mathcal{E}}_{Y|A,X}(y \mid \mathsf{a}, x)$$

is to be avoided if possible due to the danger of mis-specification. If the choice is made correctly, then it can form the basis of a regression estimator.

We seek to resolve the incompatibility using the importance sampling trick, and write the expectation with respect to the observational model

$$f^{\mathcal{O}}_{Y|X,Z}(y|x,z) f^{\mathcal{O}}_{Z|X}(z|x) f^{\mathcal{O}}_{X}(x).$$

First note that

$$f^{\mathcal{E}}_{X|Z}(x \mid z) = \frac{f^{\mathcal{E}}_{Z|X}(z \mid x) f^{\mathcal{E}}_{X}(x)}{f^{\mathcal{E}}_{Z}(z)}$$

so the integral can be rewritten

$$\frac{1}{f^{\mathcal{E}}_{Z}(z)} \iint y\, f^{\mathcal{E}}_{Y|A,X}(y \mid \mathsf{a}, x) f^{\mathcal{E}}_{Z|X}(z \mid x) f^{\mathcal{E}}_{X}(x)\, dy\, dx.$$

For $\mathsf{z} = 0, 1$, we can re-write the integrand using the importance sampling trick as

$$y\, f^{\mathcal{E}}_{Y|A,X}(y \mid \mathsf{a}, x) \frac{f^{\mathcal{E}}_{Z|X}(z \mid x)}{f^{\mathcal{E}}_{Z|X}(\mathsf{a} \mid x)}\, f^{\mathcal{E}}_{Z|X}(\mathsf{a} \mid x) f^{\mathcal{E}}_{X}(x)$$

which can be rearranged to

$$\left\{ y\, \frac{f^{\mathcal{E}}_{Z|X}(z \mid x)}{f^{\mathcal{E}}_{Z|X}(\mathsf{a} \mid x)} \right\} f^{\mathcal{E}}_{Y|A,X}(y \mid \mathsf{a}, x)\, f^{\mathcal{E}}_{Z|X}(\mathsf{a} \mid x) f^{\mathcal{E}}_{X}(x),$$

Comparing the observational and experimental DAGs, we see that for all x and z

$$f^{\mathcal{E}}_{Z|X}(z \mid x) \equiv f^{\mathcal{O}}_{Z|X}(z \mid x) \qquad f^{\mathcal{E}}_{X}(x) \equiv f^{\mathcal{O}}_{X}(x) \qquad f^{\mathcal{E}}_{Z}(z) \equiv f^{\mathcal{O}}_{Z}(z).$$

Also, we have for any t and y that

$$f^{\mathcal{E}}_{Y|A,X}(y \mid t, x) \equiv f^{\mathcal{O}}_{Y|X,Z}(y \mid x, t).$$

Therefore we have

$$f^{\mathcal{E}}_{Y|A,X}(y \mid \mathsf{a}, x)\, f^{\mathcal{E}}_{Z|X}(\mathsf{a} \mid x) f^{\mathcal{E}}_{X}(x) \equiv f^{\mathcal{O}}_{Y|X,Z}(y \mid x, \mathsf{a})\, f^{\mathcal{O}}_{Z|X}(\mathsf{a} \mid x) f^{\mathcal{O}}_{X}(x).$$

Thus

$$\mathbb{E}^{\mathcal{E}}_{Y|A,Z}[Y \mid A = \mathsf{a}, Z = z] = \frac{1}{f^{\mathcal{O}}_{Z}(z)} \iint \left\{ y \frac{f^{\mathcal{O}}_{Z|X}(z \mid x)}{f^{\mathcal{O}}_{Z|X}(\mathsf{a} \mid x)} \right\} f^{\mathcal{O}}_{X,Y,Z}(x, y, \mathsf{a})\, dy\, dx$$

$$= \frac{1}{f^{\mathcal{O}}_{Z}(z)} \iiint \left\{ \mathbb{1}_{\{\mathsf{a}\}}(t) y \frac{f^{\mathcal{O}}_{Z|X}(z \mid x)}{f^{\mathcal{O}}_{Z|X}(t \mid x)} \right\} f^{\mathcal{O}}_{X,Y,Z}(x, y, t)\, dy\, dx\, dt$$

$$= \frac{1}{f^{\mathcal{O}}_{Z}(z)} \mathbb{E}^{\mathcal{O}}_{X,Y,Z} \left[\mathbb{1}_{\{\mathsf{a}\}}(Z) Y \, \frac{f^{\mathcal{O}}_{Z|X}(z \mid X)}{f^{\mathcal{O}}_{Z|X}(Z \mid X)} \right].$$

For the ATT, we are interested only in $z = 1$. The moment-based estimator is therefore

$$\widehat{\mathbb{E}}^{\mathcal{E}}_{Y|A,Z}[Y \mid A = \mathsf{a}, Z = 1] = \frac{\sum_{i=1}^{n} \mathbb{1}_{\{\mathsf{a}\}}(Z_i) w_1(X_i, Z_i) Y_i}{\sum_{i=1}^{n} \mathbb{1}_{\{1\}}(Z_i)} \qquad \text{where} \qquad w_z(X_i, Z_i) = \frac{f^{\mathcal{O}}_{Z|X}(z \mid X_i)}{f^{\mathcal{O}}_{Z|X}(Z_i \mid X_i)}$$

- When $\mathsf{a} = 1$,

  $$\mathbb{1}_{\{\mathsf{a}\}}(Z_i) w_1(X_i, Z_i) = \mathbb{1}_{\{1\}}(Z_i) = Z_i \quad \text{w.p. } 1$$

  as the weight is identically 1, so therefore

  $$\widehat{\mathbb{E}}^{\mathcal{E}}_{Y|A,Z}[Y \mid A = 1, Z = 1] = \frac{\sum_{i=1}^{n} Z_i Y_i}{\sum_{i=1}^{n} Z_i}$$

  that is, the mean in the treated group.
- When $\mathsf{a} = 0$,

  $$\mathbb{1}_{\{\mathsf{a}\}}(Z_i) w_1(X_i, Z_i) = \mathbb{1}_{\{0\}}(Z_i) \frac{f^{\mathcal{O}}_{Z|X}(1 \mid X_i)}{f^{\mathcal{O}}_{Z|X}(Z_i \mid X_i)} = (1 - Z_i) \frac{f^{\mathcal{O}}_{Z|X}(1 \mid X_i)}{f^{\mathcal{O}}_{Z|X}(0 \mid X_i)}.$$

  Therefore

  $$\widehat{\mathbb{E}}^{\mathcal{E}}_{Y|A,Z}[Y \mid A = 0, Z = 1] = \frac{\sum_{i=1}^{n} (1 - Z_i) w(X_i) Y_i}{\sum_{i=1}^{n} Z_i}$$

  where

  $$w(X_i) = \frac{f^{\mathcal{O}}_{Z|X}(1 \mid X_i)}{f^{\mathcal{O}}_{Z|X}(0 \mid X_i)} = \frac{e(X_i)}{1 - e(X_i)}$$

  That is, this estimator is a weighted sum of contributions from the *untreated* individuals.

Thus the estimator for the ATT is

$$\widehat{\mathbb{E}}^{\mathcal{E}}_{Y|A,Z}[Y \mid A = 1, Z = 1] - \widehat{\mathbb{E}}^{\mathcal{E}}_{Y|A,Z}[Y \mid A = 0, Z = 1] = \frac{\sum_{i=1}^{n} (Z_i - (1 - Z_i) w(X_i)) Y_i}{\sum_{i=1}^{n} Z_i}. \tag{1}$$

Under the standard assumptions, this estimator is consistent for the ATT and asymptotically normally distributed if

$$e(x) = f^{\mathcal{O}}_{Z|X}(1 \mid X_i)$$

is correctly specified; that is, the estimator is *singly robust*.

Example: In this simulation study, we have

- $X \sim Uniform(0, 10)$
- $Z|X = x \sim Bernoulli(e(x))$, with
$$e(x) = \frac{\exp\{-3 + 0.2x\}}{1 + \exp\{-3 + 0.2x\}}$$
- $Y|X = x, Z = z \sim Normal(\mu(x, z), 1)$ with
$$\mu(x, z) = x - \frac{1}{2}x \log x + \frac{5x}{1 + x} + z(2 + 0.25x)$$

Thus the ATT is
$$2 + 0.25\mu_X(1)$$
where
$$\mu_X(1) = \int x f^{\mathcal{O}}_{X|Z}(x|1)\, dx = \frac{\displaystyle\iint \mathbb{1}_{\{1\}}(z) x f^{\mathcal{O}}_{X,Z}(x, z)\, dx\, dz}{\displaystyle\iint \mathbb{1}_{\{1\}}(z) f^{\mathcal{O}}_{X,Z}(x, z)\, dx\, dz}$$
which can be estimated by the sample mean in the subset of observations with $Z = 1$.

```
set.seed(2984)
al<-c(-3,0.2)
psi<-c(2,0.25)
h.func<-function(xv) {return(xv-0.5*xv*log(xv)+5*xv/(1+xv))}
#Large sample Monte Carlo calculation of true ATT
N <- 10000000
X<-runif(N,0,10)
pi.vec<-1/(1+exp(-cbind(1,X) %*% al))
Z<-rbinom(N,1,pi.vec)
att <- psi[1] + psi[2]*mean(X[Z==1])
att

+ [1] 3.591837

Y<-h.func(X)+Z*(psi[1]+psi[2]*X)+rnorm(N)
```

Under correct specification, a regression estimator may be used

```
fit0<-lm(Y~-1+offset(h.func(X))+Z+Z:X)
psi.hat<-coef(fit0)
psi.hat[1]+psi.hat[2]*mean(X[Z==1])

+        Z
+ 3.591741
```

However, under mis-specification, the incorrect answer is obtained:

```
fit1<-lm(Y~X+Z+Z:X)
psi.hat<-coef(fit1)
psi.hat[2]+psi.hat[3]*mean(X[Z==1])

+        X
+ 18.08968
```

The estimator in (1) is computed as follows, and gives the correct answer:

```
eX<-fitted(glm(Z~X,family=binomial))
w.hat<-Z + (1-Z)*eX/(1-eX)
att.hat<-sum((Z-(1-Z)*w.hat)*Y)/sum(Z)
att.hat

+ [1] 3.59099
```

Double robustness: To achieve *double robustness*, we augment the estimand for $a = 0$ as follows:

$$\mathbb{E}[Y(0) \mid Z = 1] = \mathbb{E}[Y(0) - \mu(X, 0)|Z = 1] + \mathbb{E}[\mu(X, 0)|Z = 1]$$

where

$$\mu(x, z) = \mathbb{E}[Y \mid X = x, Z = z]$$

is the modelled conditional mean for Y. To estimate the first term, we use

$$\widehat{\mathbb{E}}[Y(0) - \mu(X, 0) \mid Z = 1] = \frac{\sum_{i=1}^{n} (1 - Z_i) w(X_i)(Y_i - \mu(X_i, 0))}{\sum_{i=1}^{n} Z_i}$$

as in the singly robust case. For the second term, we have

$$\widehat{\mathbb{E}}[\mu(X, 0) \mid Z = 1] = \frac{\sum_{i=1}^{n} Z_i \mu(X_i, 0)}{\sum_{i=1}^{n} Z_i}$$

Therefore

$$\widehat{\mathbb{E}}[Y(0) \mid Z = 1] = \frac{\sum_{i=1}^{n} (1 - Z_i) w(X_i)(Y_i - \mu(X_i, 0)) + Z_i \mu(X_i, 0)}{\sum_{i=1}^{n} Z_i}$$

which yields the augmented ATT estimator

$$\frac{\sum_{i=1}^{n} (Z_i - (1 - Z_i) w(X_i))(Y_i - \mu(X_i, 0))}{\sum_{i=1}^{n} Z_i}.$$

```
#DR estimator with mu(x,z) = beta0 + Z (psi[0]+psi[1]*X)
w.mod <- lm(Y~Z+Z:X,weights=w.hat)
att.dr1<-coef(w.mod)[2]+coef(w.mod)[3]*mean(X[Z==1])
att.dr1

+        Z
+ 3.590976

#DR estimator with mu(x,z) = beta0 + beta1 X + Z (psi[0]+psi[1]*X)
w.mod2 <- lm(Y~Z*X,weights=w.hat)
att.dr2<-coef(w.mod2)[2]+coef(w.mod2)[4]*mean(X[Z==1])
att.dr2

+        Z
+ 3.590978
```

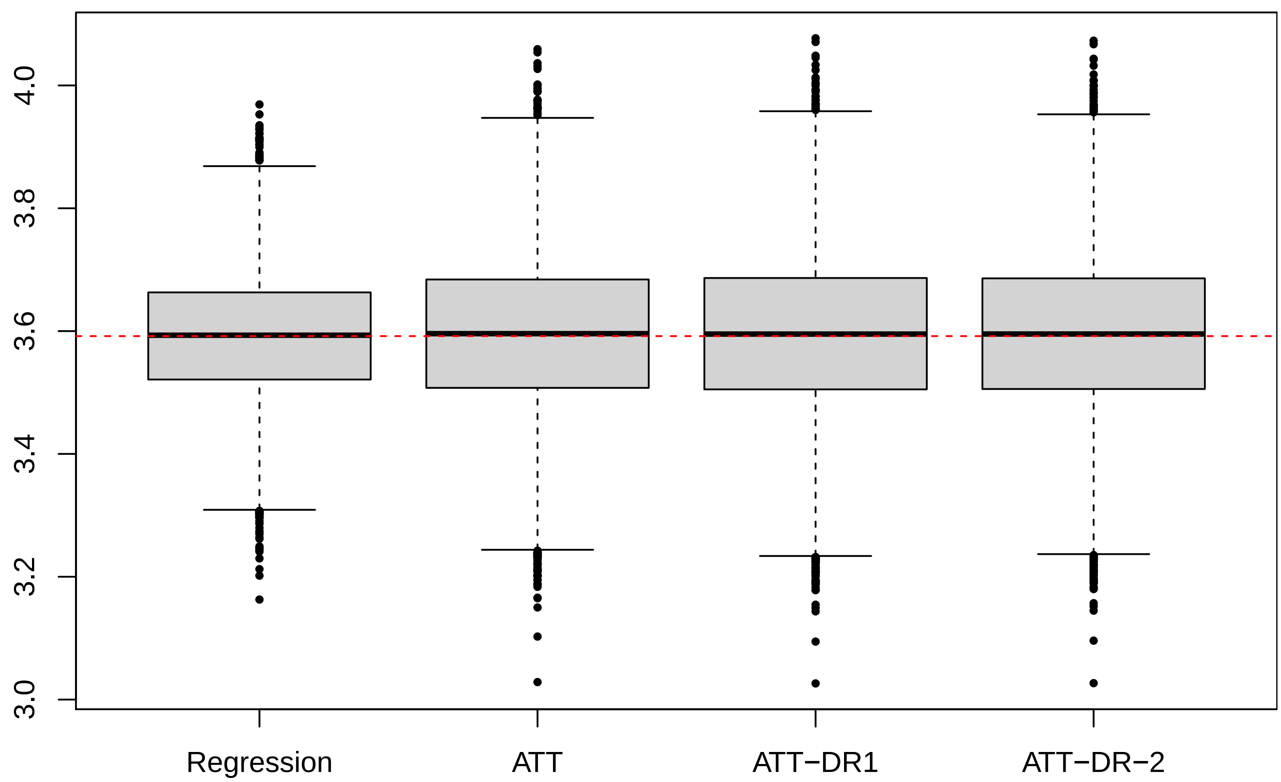
Monte Carlo study: we carry out a Monte Carlo study with 10000 replicates with $n = 1000$:

```
nreps<-10000
ests.mat<-matrix(0,nrow=nreps,ncol=4)
n<-1000
for(irep in 1:nreps) {
    X<-runif(n,0,10)
    eX0<-1/(1+exp(-cbind(1,X) %*% al))
    Z<-rbinom(n,1,eX0)
```

```
    Y<-psi[1]*Z+psi[2]*Z*X+h.func(X)+rnorm(n)
    fit0<-lm(Y~-1+offset(h.func(X))+Z+Z:X)
    psi.hat<-coef(fit0)
    ests.mat[irep,1]<-psi.hat[1]+psi.hat[2]*mean(X[Z==1])          #Regression estimator
    eX<-fitted(glm(Z~X,family=binomial))
    w.hat<-Z + (1-Z)*eX/(1-eX)
    ests.mat[irep,2]<-sum((Z-(1-Z)*w.hat)*Y)/sum(Z)                #Singly robust ATT estimator
    w.mod <- lm(Y~Z+Z:X,weights=w.hat)
    ests.mat[irep,3]<-coef(w.mod)[2]+coef(w.mod)[3]*mean(X[Z==1])   #Doubly robust ATT estimator 1
    w.mod2 <- lm(Y~Z*X,weights=w.hat)
    ests.mat[irep,4]<-coef(w.mod2)[2]+coef(w.mod2)[4]*mean(X[Z==1]) #Doubly robust ATT estimator 2

}
```

```
nv<-c('Regression','ATT','ATT-DR1','ATT-DR-2')
par(mar=c(3,3,2,0))
boxplot(ests.mat,names=nv,pch=19,cex=0.5)
abline(h=att,col='red',lty=2)
```




```
apply(ests.mat,2,var)*n                              #Variance

+ [1] 10.98275 17.02629 17.96629 17.74381
```