VARIANCE COMPARISON FOR IPW AND G-ESTIMATION

We can compute an estimate of the variance-covariance matrix using the theory of estimating equations. For the $p \times 1$ system of estimating equations

$$\sum_{i=1}^{n} \mathbf{U}_i(\theta) = \mathbf{0}$$

with $\mathbb{E}[\mathbf{U}(\theta_0)] = \mathbf{0}$, we have that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} Normal_p(\mathbf{0}, \mathbf{V})$$

where

$$\mathbf{V} = \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-\top}$$

with

$$\mathcal{I} = \mathbb{E}[\mathbf{U}(\theta_0)\mathbf{U}(\theta_0)^\top] \qquad \mathcal{J} = \mathbb{E}[\dot{\mathbf{U}}(\theta_0)]$$

both $(p \times p)$ matrices, and

$$\dot{\mathbf{U}}(\theta_0) = -\frac{\partial \mathbf{U}(\theta)}{\partial \theta^\top}\bigg|_{\theta=\theta_0}$$

We proceed by computing estimates, $\widehat{I}_n$ and $\widehat{J}_n$, of the two matrices based on the observed data, and then computing

$$\widehat{\mathbf{V}} = \widehat{J}_n^{-1}\widehat{I}_n\widehat{J}_n^{-\top}$$

to estimate the asymptotic variance; this approach is known as *sandwich* or *robust* variance estimation. In the following analysis for, we denote the ATE by ψ_0.

1. **IPW version 1:** For the elementary application of IPW, there is a single unknown parameter of interest, and the estimator is

$$\widehat{\psi}_0 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Z_i}{e(X)} - \frac{1-Z_i}{1-e(X_i)}\right)Y_i$$

   (a) ***Known propensity score:*** The score is

$$U(\theta) = \left(\frac{Z}{e(X)} - \frac{1-Z}{1-e(X)}\right)Y - \psi_0$$

   and hence $\dot{U}(\theta) = 1$.

   (b) ***Unknown propensity score:*** The score is

$$\mathbf{U}(\theta) = \begin{pmatrix} \left(\frac{Z}{e(X;\alpha)} - \frac{1-Z}{1-e(X;\alpha)}\right)Y - \psi_0 \\ \mathbf{X}^\top(Z - e(X;\alpha)) \end{pmatrix} = \begin{pmatrix} \zeta_Y \\ \mathbf{X}^\top\epsilon_Z \end{pmatrix}.$$

   say. Therefore

$$\mathcal{I} = \mathbb{E}\begin{bmatrix} \zeta_Y^2 & \mathbf{X}\epsilon_Z\zeta_Y \\ \mathbf{X}^\top\epsilon_Z\zeta_Y & \mathbf{X}^\top\mathbf{X}\epsilon_Z^2 \end{bmatrix} \qquad \mathcal{J} = \mathbb{E}\begin{bmatrix} 1 & \left(\frac{Z}{e^2} + \frac{1-Z}{(1-e)^2}\right)Y\mathbf{X}e(1-e) \\ \mathbf{0} & \mathbf{X}^\top\mathbf{X}e(1-e) \end{bmatrix}.$$

2. **IPW version 2:** For the second version of the IPW estimator, we consider weighted least squares and the objective function

$$\sum_{i=1}^{n} R_i(Y_i - \beta_0 - Z_i\psi_0)^2$$

   where

$$R_i = \left(\frac{1-Z}{1-e(X)} + \frac{Z_i}{e(X)}\right).$$

This yields the estimator

$$\widehat{\psi}_0 = \frac{\displaystyle\sum_{i=1}^{n} \frac{Z_i Y_i}{e(X_i)}}{\displaystyle\sum_{i=1}^{n} \frac{Z_i}{e(X_i)}} - \frac{\displaystyle\sum_{i=1}^{n} \frac{(1-Z_i)Y_i}{1-e(X_i)}}{\displaystyle\sum_{i=1}^{n} \frac{(1-Z_i)}{1-e(X_i)}}.$$

(a) ***Known propensity score:*** The score vector is

$$\mathbf{U}(\theta) = \begin{pmatrix} R(Y-\beta_0 - Z\psi_0) \\ RZ(Y-\beta_0 - Z\psi_0) \end{pmatrix} = \begin{pmatrix} R\varepsilon_Y \\ RZ\varepsilon_Y \end{pmatrix}$$

say, where $\varepsilon_Y = (Y-\beta_0 - Z\psi_0)$. Therefore

$$\mathcal{I} = \mathbb{E}\begin{bmatrix} R^2\varepsilon_Y^2 & R^2Z\varepsilon_Y^2 \\ R^2Z\varepsilon_Y^2 & R^2Z^2\varepsilon_Y^2 \end{bmatrix} \qquad \mathcal{J} = \mathbb{E}\begin{bmatrix} R & RZ \\ RZ & RZ^2 \end{bmatrix}.$$

(b) ***Unknown propensity score:*** The score vector is

$$\mathbf{U}(\theta) = \begin{pmatrix} R(Y-\beta_0 - Z\psi_0) \\ RZ(Y-\beta_0 - Z\psi_0) \\ \mathbf{X}^\top(Z - e(X;\alpha)) \end{pmatrix} = \begin{pmatrix} R\varepsilon_Y \\ RZ\varepsilon_Y \\ \mathbf{X}^\top\epsilon_Z \end{pmatrix}$$

Therefore

$$\mathcal{I} = \mathbb{E}\begin{bmatrix} R^2\varepsilon_Y^2 & R^2Z\varepsilon_Y^2 & \mathbf{X}R\epsilon_Z\varepsilon_Y \\ R^2Z\varepsilon_Y^2 & R^2Z^2\varepsilon_Y^2 & \mathbf{X}RZ\epsilon_Z\varepsilon_Y \\ \mathbf{X}^\top R\epsilon_Z\varepsilon_Y & \mathbf{X}^\top RZ\epsilon_Z\varepsilon_Y & \mathbf{X}^\top\mathbf{X}\epsilon_Z^2 \end{bmatrix} \qquad \mathcal{J} = \mathbb{E}\begin{bmatrix} R & RZ & -\dot{\mathbf{R}}^\top\varepsilon_Y \\ RZ & RZ^2 & -\dot{\mathbf{R}}^\top Z\varepsilon_Y \\ \mathbf{0} & \mathbf{0} & \mathbf{X}^\top\mathbf{X}e(1-e) \end{bmatrix}.$$

where

$$\dot{\mathbf{R}} = \frac{dR}{d\alpha} = \frac{1-Z}{1-e(X)}e(X)\mathbf{X}^\top - \frac{Z}{e(X)}(1-e(X))\mathbf{X}^\top.$$

3. **G-estimation:** The most elementary form of the G-estimating equation is based on solving

$$\sum_{i=1}^{n}(Z_i - e(X_i))(Y_i - Z_i\psi_0) = 0.$$

(a) ***Known propensity score:*** The score vector is

$$U(\theta) = (Z - e(X))(Y - Z\psi_0)$$

with $\dot{U}(\theta) = Z(Z - e(X))$.

(b) ***Unknown propensity score:*** The score vector is

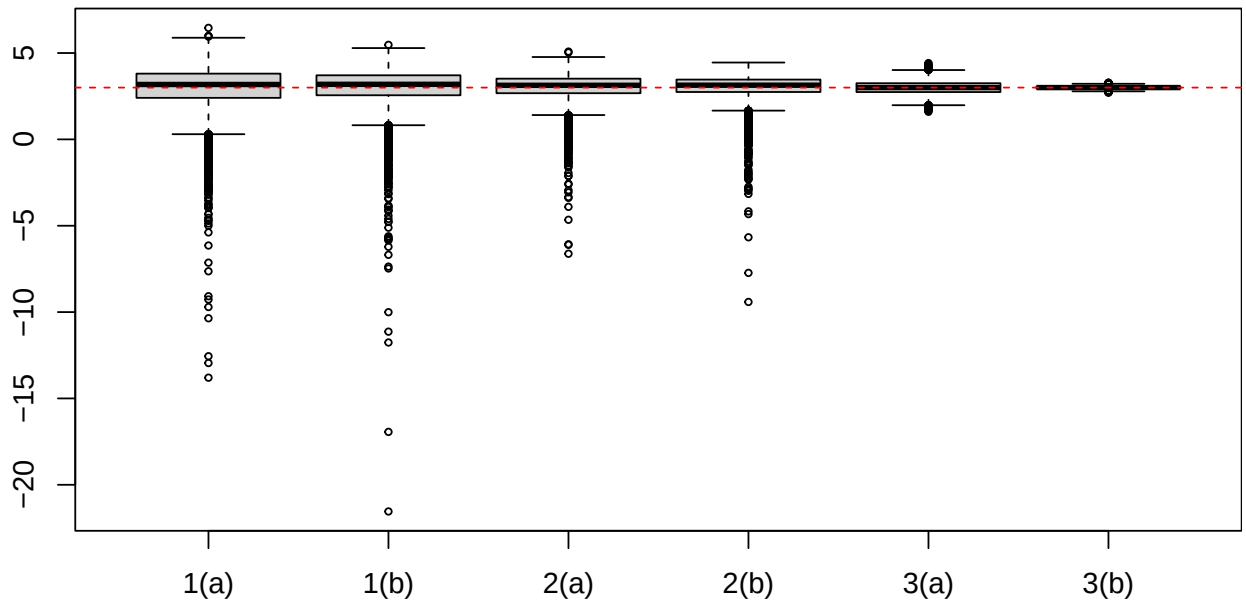
$$\mathbf{U}(\theta) = \begin{pmatrix} (Z - e(X;\alpha))(Y - Z\psi_0) \\ \mathbf{X}^\top(Z - e(X;\alpha)) \end{pmatrix} = \begin{pmatrix} \epsilon_Z\epsilon_Y \\ \mathbf{X}^\top\epsilon_Z \end{pmatrix}$$

say, where $\epsilon_Y = (Y - Z\psi_0)$. Therefore

$$\mathcal{I} = \mathbb{E}\begin{bmatrix} \epsilon_Z^2\epsilon_Y^2 & \mathbf{X}\epsilon_Z^2\epsilon_Y \\ \epsilon_Z^2\epsilon_Y\mathbf{X}^\top & \mathbf{X}^\top\mathbf{X}\epsilon_Z^2 \end{bmatrix} \qquad \mathcal{J} = \mathbb{E}\begin{bmatrix} Z\epsilon_Z & \mathbf{X}e(1-e)\epsilon_Y \\ \mathbf{0} & \mathbf{X}^\top\mathbf{X}e(1-e) \end{bmatrix}.$$

```r
#Monte Carlo study
library(mvnfast)
set.seed(22087)
n<-10000
muX<-2;sigX<-1
al<-c(-3.5,2);be<-c(2,3,4,1);sigY<-3
expit<-function(x){return(1/(1+exp(-x)))}
nreps<-10000
ate.ests<-matrix(0,nrow=nreps,ncol=6)
colnames(ate.ests)<-c('1(a)','1(b)','2(a)','2(b)','3(a)','3(b)')
for(irep in 1:nreps){
    X1<-rnorm(n,muX,sigX)
    Xm<-cbind(1,X1)
    eX0<-expit(Xm %*% al)
    Z<-rbinom(n,1,eX0)
    Xb<-cbind(1,Z,X1,X1^2)
    Y<-Xb %*% be + rnorm(n)*sigY
    eX<-fitted(glm(Z~X1,family=binomial))
    ate.ests[irep,1]<-mean(Z*Y/eX0)-mean((1-Z)*Y/(1-eX0))
    ate.ests[irep,2]<-mean(Z*Y/eX)-mean((1-Z)*Y/(1-eX))
    w<-Z/eX0+(1-Z)/(1-eX0)
    ate.ests[irep,3]<-coef(lm(Y~Z,weights=w))[2]
    w<-Z/eX+(1-Z)/(1-eX)
    ate.ests[irep,4]<-coef(lm(Y~Z,weights=w))[2]
    res2<-Z-eX0
    ate.ests[irep,5]<-sum(res2*Y)/sum(res2*Z)
    res2<-Z-eX
    ate.ests[irep,6]<-sum(res2*Y)/sum(res2*Z)
}
#apply(ate.ests,2,mean)-3
apply(ate.ests,2,var)
```

```
+        1(a)        1(b)        2(a)        2(b)        3(a)        3(b)
+ 1.608020241 1.321286394 0.586662070 0.495380667 0.141293585 0.006275017
```



```
#IPW 1a: PS known
#Direct calculation
vX0<-mean((sigY^2+mu0^2)/(1-eX0))
vX1<-mean((sigY^2+mu1^2)/eX0)
ipw.var<-(vX0+vX1)-be[2]^2
ipw.var/n
+ [1] 1.743757
I.n<-J.n<-matrix(0,1,1)
res1<-(Z/eX0-(1-Z)/(1-eX0))*Y-be[2]
I.n[1,1]<-mean(res1^2)

vx<-eX0*(1-eX0)
J.n[1,1]<-1
I.var<-solve(J.n) %*% (I.n %*% t(solve(J.n)))/n

#This should match the 1(a) variance
I.var[1,1]
+ [1] 1.010634
var(ate.ests[,1])
+ [1] 1.60802
##########################################################
#IPW 1b: PS unknown
I.n<-J.n<-matrix(0,3,3)
res1<-(Z/eX-(1-Z)/(1-eX))*Y-be[2]
res2<-(Z-eX)
I.n[1,1]<-mean(res1^2)
I.n[1,2]<-I.n[2,1]<-mean(res1*res2)
I.n[2,2]<-mean(res2^2)
I.n[1,3]<-I.n[3,1]<-mean(X1*res1*res2)
I.n[2,3]<-I.n[3,2]<-mean(X1*res2^2)
I.n[3,3]<-mean(X1^2*res2^2)
I.n
+              [,1]       [,2]       [,3]
+ [1,] 10037.97829 16.1715281 39.1476619
+ [2,]    16.17153  0.1486727  0.2751906
+ [3,]    39.14766  0.2751906  0.5690280
vx<-eX*(1-eX)
J.n[1,1]<-1
J.n[1,2]<-mean(vx*Y*(Z/eX^2+(1-Z)/(1-eX)^2))
J.n[1,3]<-mean(X1*vx*Y*(Z/eX^2+(1-Z)/(1-eX)^2))
J.n[2,2]<-mean(vx)
J.n[2,3]<-J.n[3,2]<-mean(vx*X1)
J.n[3,3]<-mean(vx*X1^2)
J.n
+      [,1]       [,2]       [,3]
+ [1,]    1 16.1715281 39.1476619
+ [2,]    0  0.1486740  0.2755749
+ [3,]    0  0.2755749  0.5703299
I.var<-solve(J.n) %*% (I.n %*% t(solve(J.n)))/n

#This should match the 1b variance
I.var[1,1]
+ [1] 0.6855707
var(ate.ests[,2])
+ [1] 1.321286
```

```
#IPW hat 2a: PS known
w0<-(1-Z)/(1-eX0)
w1<-Z/eX0
w<-w0+w1
ipw.ests<-coef(lm(Y~Z,weights=w))
res1<-Y-ipw.ests[1]-ipw.ests[2]*Z
R<-w

I.n<-matrix(0,2,2)
I.n[1,1]<-mean(R^2*res1^2)
I.n[1,2]<-I.n[2,1]<-mean(R^2*Z*res1^2)
I.n[2,2]<-mean(R^2*Z^2*res1^2)

round(I.n,6)

+           [,1]     [,2]
+ [1,] 3891.1375 916.6819
+ [2,]  916.6819 916.6819

J.n<-matrix(0,2,2)
J.n[1,1]<-mean(R)
J.n[1,2]<-mean(R*Z)
J.n[2,1]<-mean(R*Z)
J.n[2,2]<-mean(R*Z^2)
round(J.n,6)

+          [,1]    [,2]
+ [1,] 2.000105 0.99974
+ [2,] 0.999740 0.99974

#This should match the 2a variance
H.var<-solve(J.n) %*% (I.n %*% t(solve(J.n)))/n
H.var[2,2]

+ [1] 0.3889447

var(ate.ests[,3])

+ [1] 0.5866621

###############################################################
#IPW hat 2b: PS unknown
w0<-(1-Z)/(1-eX)
w1<-Z/eX
w<-w0+w1

ipw.ests<-coef(lm(Y~Z,weights=w))
res1<-Y-ipw.ests[1]-ipw.ests[2]*Z
res2<-Z-eX
R<-w
Rdot1<-((1-Z)/(1-eX))*eX-(Z/eX)*(1-eX)
Rdot2<-X1*Rdot1

I.n<-matrix(0,4,4)
I.n[1,1]<-mean(R^2*res1^2)
I.n[1,2]<-I.n[2,1]<-mean(R^2*Z*res1^2)
I.n[1,3]<-I.n[3,1]<-mean(R*res2*res1)
I.n[1,4]<-I.n[4,1]<-mean(X1*R*res2*res1)
I.n[2,2]<-mean(R^2*Z^2*res1^2)
I.n[2,3]<-I.n[3,2]<-mean(R*Z*res2*res1)
I.n[2,4]<-I.n[4,2]<-mean(X1*R*Z*res2*res1)
I.n[3,3]<-mean(res2^2)
I.n[3,4]<-I.n[4,3]<-mean(X1*res2^2)
I.n[4,4]<-mean(X1^2*res2^2)
round(I.n,6)
```

```
+              [,1]       [,2]      [,3]       [,4]
+ [1,] 3888.372126 932.461267 -4.692068 -10.347673
+ [2,]  932.461267 932.461267 -2.375526  -1.263227
+ [3,]   -4.692068  -2.375526  0.148673   0.275191
+ [4,]  -10.347673  -1.263227  0.275191   0.569028

J.n<-matrix(0,4,4)
J.n[1,1]<-mean(R)
J.n[1,2]<-mean(R*Z)
J.n[1,3]<-mean(-Rdot1*res1)
J.n[1,4]<-mean(-Rdot2*res1)
J.n[2,1]<-mean(R*Z)
J.n[2,2]<-mean(R*Z^2)
J.n[2,3]<-mean(-Rdot1*Z*res1)
J.n[2,4]<-mean(-Rdot2*Z*res1)
J.n[3,3]<-mean(res2^2)
J.n[3,4]<-J.n[4,3]<-mean(X1*res2^2)
J.n[4,4]<-mean(X1^2*res2^2)
round(J.n,6)

+          [,1]     [,2]      [,3]       [,4]
+ [1,] 2.000016 1.003235 -4.692068 -10.347673
+ [2,] 1.003235 1.003235 -2.375526  -1.263227
+ [3,] 0.000000 0.000000  0.148673   0.275191
+ [4,] 0.000000 0.000000  0.275191   0.569028

#This should match the 2c variance
H.var<-solve(J.n) %*% (I.n %*% t(solve(J.n)))/n
H.var[2,2]

+ [1] 0.2845798

var(ate.ests[,4])

+ [1] 0.4953807
```

```
#G-estimation variance 3a: known PS
I.n<-J.n<-matrix(0,1,1)
res1<-(Y-be[2]*Z)

I.n[1,1]<-mean(res1^2*res2^2)
J.n[1,1]<-mean(Z*res2)

#This should match the 3a variance
G.var<-solve(J.n) %*% (I.n %*% t(solve(J.n)))/n
G.var

+           [,1]
+ [1,] 0.1403005

var(ate.ests[,5])

+ [1] 0.1412936

################################################################
#G-estimation variance 3b: unknown PS
I.n<-J.n<-matrix(0,3,3)
res1<-(Y-be[2]*Z)
res2<-(Z-eX)

I.n[1,1]<-mean(res1^2*res2^2)
I.n[1,2]<-I.n[2,1]<-mean(res2^2*res1)
I.n[2,2]<-mean(res2^2)
I.n[1,3]<-I.n[3,1]<-mean(X1*res2^2*res1)
I.n[2,3]<-I.n[3,2]<-mean(X1*res2^2)
I.n[3,3]<-mean(X1^2*res2^2)
I.n

+           [,1]      [,2]      [,3]
+ [1,] 31.011689 1.9662120 4.1041356
+ [2,]  1.966212 0.1486727 0.2751906
+ [3,]  4.104136 0.2751906 0.5690280

vx<-eX*(1-eX)
J.n[1,1]<-mean(Z*res2)
J.n[1,2]<-mean(vx*res1)
J.n[2,2]<-mean(vx)
J.n[1,3]<-mean(vx*X1*res1)
J.n[2,3]<-J.n[3,2]<-mean(vx*X1)
J.n[3,3]<-mean(vx*X1^2)
J.n

+           [,1]      [,2]      [,3]
+ [1,] 0.1486734 1.9675813 4.1076827
+ [2,] 0.0000000 0.1486740 0.2755749
+ [3,] 0.0000000 0.2755749 0.5703299

#This should match the 3b variance
G.var<-solve(J.n) %*% (I.n %*% t(solve(J.n)))/n
G.var[1,1]

+ [1] 0.00628136

G.var[1,1]-G.var[1,-1] %*% solve(G.var[-1,-1]) %*% G.var[-1,1]

+             [,1]
+ [1,] 0.006280537

var(ate.ests[,6])

+ [1] 0.006275017
```