#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

#

rm(list=ls())

#

#library('Rmpi')

library('MASS')

library('gtools');

library('npmlreg')

library('rootSolve')

#library('locfit')

#library('KernSmooth')

library('sm')

library(sandwich)

#

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

#

#

continuous\_trait<-function(l){

#

set.seed(l)

#

nca<-500

nco<-500

n<-10000

#

b0<-c(0.5,-log(2),log(1.5))

px1<-0.3

se0<-1

sw0<-1

dft<-6

dfc<-4

vt<-dft/(dft-2)

vc<-2\*dfc

c0<-c(0,0.5)

gams<-cbind(-2.8,permutations(3,2,c(0,log(1.5),log(2)),repeats.allowed=T),-0.5)

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

k<-9

g0<-gams[k,]

sumd<-0

while(sumd<nca){

e0<-rnorm(n,0,sqrt(se0))

e1<-sqrt(se0/vc)\*(rchisq(n,dfc,ncp=0)-dfc)

e2<-sqrt(se0/vt)\*(rt(n,dft))

xx<-rbinom(n,2,px1)

ww<-c0[1]+c0[2]\*xx+rnorm(n,0,sqrt(sw0))

yy<-b0[1]+b0[2]\*xx+b0[3]\*ww+e0

prob\_xywc<-(1+exp(-g0[1]-g0[2]\*yy-g0[3]\*xx-g0[4]\*ww))^-1

prob\_xywm<-(1+exp(-g0[1]-g0[2]\*yy-g0[3]\*xx-g0[4]\*ww+log(1.5)\*xx\*yy))^-1

dd<-rbinom(n,1,prob\_xywc)

sumd<-sum(dd)

}

#mean(dd);sum(dd)

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

id<-1:n

data\_pop<-cbind(id,yy,xx,ww,dd)

case\_das<-data\_pop[which(dd==1),]

ncase<-length(case\_das[,1])

cont\_das<-data\_pop[which(dd==0),]

ncont<-length(cont\_das[,1])

case\_sel<-case\_das[sample(1:ncase,nca,replace=FALSE,prob=NULL),]

cont\_sel<-cont\_das[sample(1:ncont,nco,replace=FALSE,prob=NULL),]

stage2\_das<-rbind(case\_sel,cont\_sel)

o<-order(stage2\_das[,1])

stage2\_data<-stage2\_das[o,]

stage2\_sel<-rep(0,n)

for(i in stage2\_data[,1]){stage2\_sel[i]<-1}

#

N1<-sum(dd==1)

N0<-sum(dd==0)

sel<-rep(0,n)

sel[dd==1]<-1

sel[dd==0]<-sample(c(rep(1,N1),rep(0,N0-N1)))

#

#data<-cbind(yy,xx,ww,dd,sel)

data<-cbind(data\_pop[,2:5],stage2\_sel)

od<-order(data[,4],decreasing=TRUE)

data<-data[od,]

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

y<-data[,1]

x<-data[,2]

w<-data[,3]

d<-data[,4]

s<-data[,5]

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

n1<-sum(s)

n<-length(s)

dcat<-unique(d)

ld<-length(dcat)

#

y1<-y[s==1]

x1<-x[s==1]

w1<-w[s==1]

#

h00<-rep(0,ld);for(k in 1:ld){h00[k]<-h.select(y[d==dcat[k]],s[d==dcat[k]],method="cv")}

h01<-rep(0,ld);for(k in 1:ld){h01[k]<-4\*sd(y[d==dcat[k]])\*(length(y[d==dcat[k]])^(-1/3))}

#

dgr<-1

h0<-h01

probs<-rep(0,n);probd<-rep(0,n);for(k in 1:ld){ydk<-y[d==dcat[k]];sdk<-s[d==dcat[k]]

probs[d==dcat[k]]<-sm.regression(ydk,sdk,h=h0[k],eval.points=ydk,poly.index=dgr)$estimate

probd[d==dcat[k]]<-rep(sum(sdk)/length(sdk),length(sdk))}

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

wgts<-s/probs

wgtd<-s/probd

wgts1<-wgts[s==1]

wgtd1<-wgtd[s==1]

exclude<-which((probs<=0.01)|(probs>=0.99)|(as.numeric(is.na(probs))==1))

#length(exclude)

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

# Ideal Case Approach

#

fid<-function(b){apply(cbind(1,x,w)\*(y-b[1]-b[2]\*x-b[3]\*w),2,sum)}

bid<-multiroot(fid,b0,maxiter=200,rtol=1e-6,atol= 1e-8,ctol =1e-8)$root

sid<-sqrt(diag(-solve((gradient(fid,bid,centered =FALSE,pert=1e-8)))))

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

# Complete Case Approach

#

ffc<-function(b){cbind(1,x1,w1)\*(y1-b[1]-b[2]\*x1-b[3]\*w1)}

fcc<-function(b){apply(ffc(b),2,sum)}

bcc<-multiroot(fcc,b0,maxiter=200,rtol=1e-6,atol= 1e-8,ctol =1e-8)$root

GGcc<-solve((gradient(fcc,bcc,centered =FALSE,pert=1e-8)))

BBcc<-(n1-1)\*var(ffc(bcc))

scc<-sqrt(diag(GGcc%\*%BBcc%\*%t(GGcc)))

#

npar<-length(bcc)

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

# SIPW Approaches

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

# Simple weights

#

fipws<-function(b){wgtd1\*cbind(1,x1,w1)\*(y1-b[1]-b[2]\*x1-b[3]\*w1)}

ipws<-function(b){apply(fipws(b),2,sum)}

bipws<-multiroot(ipws,b0,maxiter=200,rtol=1e-6,atol= 1e-8,ctol =1e-8)$root

Gipws<-solve((gradient(ipws,bipws,centered =FALSE,pert=1e-8)))

Bipws<-(n1-1)\*var(fipws(bipws))

sipws<-sqrt(diag(Gipws%\*%Bipws%\*%t(Gipws)))

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

# Kernel-assisted weights

#

#

if(length(exclude)>0){

#

cexps<-matrix(0,ncol=npar,nrow=n)

fipwk<-function(b){(wgts\*cbind(1,x,w)\*(y-b[1]-b[2]\*x-b[3]\*w))[-exclude,]}

ipwk<-function(b){apply(fipwk(b),2,sum)}

bipwk<-multiroot(ipwk,b0,maxiter=200,rtol=1e-6,atol= 1e-8,ctol =1e-8)$root

Gipwk<-solve((gradient(ipwk,bipwk,centered =FALSE,pert=1e-8)))

dmat<-cbind(1,x,w)\*(y-bipwk[1]-bipwk[2]\*x-bipwk[3]\*w)

dmas<-s\*dmat;for(k in 1:ld){

cexps[d==dcat[k],]<-sapply(1:ncol(dmat),function(j){

ydk<-y[d==dcat[k]];dmak<-dmas[d==dcat[k],j]

(sm.regression(ydk,dmak,h=h0[k],eval.points=ydk,poly.index=dgr)$estimate)/probs[d==dcat[k]]})}

messipw<-wgts\*dmat+(1-wgts)\*cexps;msipwk<-matrix(0,ncol=npar^2,nrow=n)

for(i in 1:n){msipwk[i,]<-as.vector(messipw[i,]%\*%t(messipw[i,]))}

Bipwk<-matrix(apply(msipwk[-exclude,],2,sum),ncol=npar,nrow=npar)

sipwk<-sqrt(diag(Gipwk%\*%Bipwk%\*%t(Gipwk)))

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

# AIPW Approach

#

aipw\_f<-function(wgta){

#

cexps<-matrix(0,ncol=npar,nrow=n)

faipw<-function(b){

dmat<-cbind(1,x,w)\*(y-b[1]-b[2]\*x-b[3]\*w)

dmas<-s\*dmat

for(k in 1:ld){

cexps[d==dcat[k],]<-sapply(1:npar,function(j){

ydk<-y[d==dcat[k]]

dmak<-dmas[d==dcat[k],j]

(sm.regression(ydk,dmak,h=h0[k],eval.points=ydk,poly.index=dgr)$estimate)/probs[d==dcat[k]]})}

(wgta\*dmat+(1-wgta)\*cexps)[-exclude,]}

aipw<-function(b){apply(faipw(b),2,sum)}

baipw<-multiroot(aipw,b0,maxiter=200,rtol=1e-6,atol= 1e-8,ctol =1e-8)$root

#

GGaipw<-solve((gradient(aipw,baipw,centered =FALSE,pert=1e-8)))

BBaipw<-(n-length(exclude)-1)\*var(faipw(baipw))

saipw<-as.vector(sqrt(diag(GGaipw%\*%BBaipw%\*%t(GGaipw))))

rbind(baipw,saipw)

}

#

aipw1<-aipw\_f(wgtd)

aipw2<-aipw\_f(wgts)

sipw1<-rbind(bipws,sipws)

sipw2<-rbind(bipwk,sipwk)

} else{

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

cexps<-matrix(0,ncol=npar,nrow=n)

fipwk<-function(b){wgts\*cbind(1,x,w)\*(y-b[1]-b[2]\*x-b[3]\*w)}

ipwk<-function(b){apply(fipwk(b),2,sum)}

bipwk<-multiroot(ipwk,b0,maxiter=200,rtol=1e-6,atol= 1e-8,ctol =1e-8)$root

Gipwk<-solve((gradient(ipwk,bipwk,centered =FALSE,pert=1e-8)))

dmat<-cbind(1,x,w)\*(y-bipwk[1]-bipwk[2]\*x-bipwk[3]\*w)

dmas<-s\*dmat;for(k in 1:ld){

cexps[d==dcat[k],]<-sapply(1:npar,function(j){

ydk<-y[d==dcat[k]];dmak<-dmas[d==dcat[k],j]

(sm.regression(ydk,dmak,h=h0[k],eval.points=ydk,poly.index=dgr)$estimate)/probs[d==dcat[k]]})}

messipw<-wgts\*dmat+(1-wgts)\*cexps;msipwk<-matrix(0,ncol=npar^2,nrow=n)

for(i in 1:n){

msipwk[i,]<-as.vector(messipw[i,]%\*%t(messipw[i,]))}

Bipwk<-matrix(apply(msipwk,2,sum),ncol=npar,nrow=npar)

sipwk<-sqrt(diag(Gipwk%\*%Bipwk%\*%t(Gipwk)))

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

# AIPW Approach

#

aipw\_f<-function(wgta){

#

cexps<-matrix(0,ncol=npar,nrow=n)

faipw<-function(b){

dmat<-cbind(1,x,w)\*(y-b[1]-b[2]\*x-b[3]\*w)

dmas<-s\*dmat;for(k in 1:ld){

cexps[d==dcat[k],]<-sapply(1:npar,function(j){

ydk<-y[d==dcat[k]];dmak<-dmas[d==dcat[k],j]

(sm.regression(ydk,dmak,h=h0[k],eval.points=ydk,poly.index=dgr)$estimate)/probs[d==dcat[k]]})}

wgta\*dmat+(1-wgta)\*cexps}

aipw<-function(b){apply(faipw(b),2,sum)}

baipw<-multiroot(aipw,b0,maxiter=200,rtol=1e-6,atol= 1e-8,ctol =1e-8)$root

#

GGaipw<-solve((gradient(aipw,baipw,centered =FALSE,pert=1e-8)))

BBaipw<-(n-1)\*var(faipw(baipw))

saipw<-as.vector(sqrt(diag(GGaipw%\*%BBaipw%\*%t(GGaipw))))

rbind(baipw,saipw)

}

#

aipw1<-aipw\_f(wgtd)

aipw2<-aipw\_f(wgts)

sipw1<-rbind(bipws,sipws)

sipw2<-rbind(bipwk,sipwk)

}

#

#round(cbind(bid,bcc,sipw1[1,],sipw2[1,],aipw1[1,],aipw2[1,])-b0,3)

#round(cbind(sid,scc,sipw1[2,],sipw2[2,],aipw1[2,],aipw2[2,]),3)

#range(probs);length(exclude)

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

# Full cohort MLE

#

xa<-c(0,1,2)

ngq<-10

ng2<-t(rep(1,ngq))

pxa<-dbinom(xa,2,px1,log=FALSE)

xxs<-gqz(numnodes=ngq,minweight=0.000001)

wei<-xxs$weight

loc<-xxs$location

yo2<-t(y%\*%ng2)

do2<-t(d%\*%ng2)

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

#

mle\_full<-function(a){

lmis<-apply(sapply(1:3,function(k){

zloc<-sqrt(sw0)\*loc+0.5\*xa[k]

py\_xw0<-exp(-0.5\*(yo2-a[1]-a[2]\*xa[k]-a[3]\*zloc)^2/a[4])/sqrt(2\*pi\*a[4])

hxyw0<-(1+exp(-a[5]-a[6]\*yo2-a[7]\*xa[k]-a[8]\*zloc-a[9]\*xa[k]\*yo2))^-1

pd\_xyw0<-(hxyw0^do2)\*((1-hxyw0)^(1-do2))

apply(wei\*pd\_xyw0\*py\_xw0\*pxa[k],2,sum)}),1,sum)

#

py\_xw<-exp(-0.5\*(y-a[1]-a[2]\*x-a[3]\*w)^2/a[4])/sqrt(2\*pi\*a[4])

pdyxw1<-(1+exp(-a[5]-a[6]\*y-a[7]\*x-a[8]\*w-a[9]\*x\*y))^-1

pd\_yxw<-(pdyxw1^d)\*((1-pdyxw1)^(1-d))

pw\_x<-exp(-0.5\*(w-0.5\*x)^2/sw0)/sqrt(2\*pi\*sw0)

p\_x<-dbinom(x,2,px1,log=FALSE)

lobs<-pd\_yxw\*py\_xw\*pw\_x\*p\_x

-sum(log(lobs[s==1]))-sum(log(lmis[s==0]))

}

#

mle\_fb<-optim(c(b0,se0,g0,-log(1.5)),mle\_full,hessian=TRUE)

bmle<-as.vector(mle\_fb$par)[1:3]

smle<-sqrt(diag(solve(mle\_fb$hessian)))[1:3]

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

#

#\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

#

bhat<-as.vector(cbind(bmle,bcc,sipw1[1,],sipw2[1,],aipw1[1,],aipw2[1,]))

sder<-as.vector(cbind(smle,scc,sipw1[2,],sipw2[2,],aipw1[2,],aipw2[2,]))

tval<-rep(b0,6)

clb<-bhat-1.96\*sder

cub<-bhat+1.96\*sder

cps<-as.numeric(tval-clb>=0)\*as.numeric(cub-tval>=0)

#

#round(cbind(bmle,bcc,sipw1[1,],sipw2[1,],aipw1[1,],aipw2[1,])-b0,3)

#round(cbind(smle,scc,sipw1[2,],sipw2[2,],aipw1[2,],aipw2[2,]),3)

#cps

#

c(as.vector(rbind(bhat,sder,cps)),mean(d),mean(s),sum(s),sum(d[s==1]))

#

}

#

#

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

#

#

#%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%