Simulation data

reference doubleml.datasets.make_irm_data
https://docs.doubleml.org/stable/api/generated/doubleml.datasets.make_irm_data.html#doubleml-datasets-make-irm-data

setting

```
library(mvtnorm)
#sample size, number of covariates, parameters setting
n=7500
p=10
tim=c(0,1,2,3,4)
beta=c(0.5,0.5,1,1/4,1/9,0,0,0,0,0)
G=matrix(0,(p-2),(p-2))
diag(G)=0.09
theta=0
logisticf=function(x){exp(x)/(1+exp(x))}
cd=sqrt(logisticf(tim*0)^2*(pi^2/3)/((1-logisticf(tim*0)^2)*sum(beta*beta)))
cy=sqrt((logisticf((tim*0+1.3)))^2/((1-(logisticf((tim*0+1.3))))^2)*sum(beta*beta)))

#functions and matrices(save value)
X1=matrix(0,n,(p-2))
vaso=matrix(0,n,1)
water=matrix(0,n,1)
Death=matrix(0,n,1)
treat=matrix(0,n,1)
data=matrix(0,n*length(tim),19)
data_D=matrix(0,n,length(tim))
deathp=0
```

generating

```
#-------------------------------------------------------------------------------
set.seed(1219)
v=matrix(c(rnorm(n*(p-2),0,1)),n,(p-2))
X2=matrix(c(rbinom(n,1,0.5),rnorm(n,0,1)),n,2)

set.seed(61916)
for(j in c(1:(length(tim)))){
  X1=rmvnorm(n,rep(0,(p-2)),G)+v-matrix(theta,n,p-2)
  X=cbind(X2,X1)
  term=c(X%*%matrix(beta,p,1)*cd[j])
  g=logisticf(term)
  treat=matrix(rbinom(n,1,g),n,1)
  water=matrix(rgamma(n,exp(term)/2,2),n,1)*5
  water[water<0.01]=0
  vaso=matrix(rgamma(n,exp(term)/2,2),n,1)/50
  vaso[vaso<0.001]=0
  theta=(water+vaso)/(1+water+vaso)
  y=-0.6+0.5*X[,2]+0.5*exp(c(X%*%matrix(beta,p,1)*cy[j]))*(1-c(treat))
  deathp=tanh(y)/2+0.5
```

```
  data_D[,j]=c(rbinom(n,1,deathp))
  Death=data_D[,j]
  Score=c(rexp(n,(1-deathp/1.285)))
  Dsofa=(Score-1)
  Dsofamean=(1-deathp/1.285)^-1
  y0=-0.6+0.5*X[,2]
  y0=tanh(y0)/2+0.5
  y1=-0.6+0.5*X[,2]+0.5*exp(c(X%*%matrix(beta,p,1)*cy[j]))
  y1=tanh(y1)/2+0.5

  #store
  data[c((n*tim[j]+1):(n*tim[j]+n)),1]=c(1:n)
  data[c((n*tim[j]+1):(n*tim[j]+n)),2]=rep(tim[j],n)
  data[c((n*tim[j]+1):(n*tim[j]+n)),3]=y1-y0
  data[c((n*tim[j]+1):(n*tim[j]+n)),4]=Death
  data[c((n*tim[j]+1):(n*tim[j]+n)),5]=(1-y1/1.285)^(-1)-(1-y0/1.285)^(-1)
  data[c((n*tim[j]+1):(n*tim[j]+n)),6]=Dsofa
  data[c((n*tim[j]+1):(n*tim[j]+n)),7]=c(treat)
  data[c((n*tim[j]+1):(n*tim[j]+n)),18]=vaso
  data[c((n*tim[j]+1):(n*tim[j]+n)),19]=water
  data[c((n*tim[j]+1):(n*tim[j]+n)),8:17]=X
}

colnames(data)=c("id","bloc","true_ite.(Death)","Death","true_ite(Dsofa)","Dsofa",
"Treat","gender","age","X1","X2","X3","X4","X5","X6","X7","X8","vaso","water")
data=as.data.frame(data)
data_bloc1=data[data$bloc==0,]
data_bloc1$Death=data_D[,1]
data_bloc1$Treat=1-data_bloc1$Treat

#mark NA at Death=1
for(j in c(1:(length(tim)-1))){
  count=c(1:n)
  deathid=count[c(count*data_D[,j])!=0]
  for(i in c(1:length(deathid))){
    data[(data$id==deathid[i] & data$bloc>=j),c(-1,-2)]=NA
  }
}
data=na.omit(data)
colnames(data_bloc1)=c("id","bloc","true_ite.(Death)","Death","true_ite(Dsofa)",
"Dsofa","Treat","X1","X2","X3","X4","X5","X6","X7","X8","X9","X10","vaso","water")
#-------------------------------------------------------------------------------
```

export

```
data_class=data_bloc1[,c(-5,-6,-18,-19)]
data_cont=data_bloc1[,c(-3,-4,-18,-19)]
write.csv(data_class,"C:/Users/ASUS/Desktop/cg working/CausalML/simulation
data_DTR/for plan/simulation data(binary).csv",row.names = F)
write.csv(data_cont,"C:/Users/ASUS/Desktop/cg working/CausalML/simulation
data_DTR/for plan/simulation data(cont.).csv",row.names = F)
write.csv(data,"C:/Users/ASUS/Desktop/cg working/CausalML/simulation data_DTR/for
plan/simulation data.csv",row.names = F)
```

#==
#--binary case--
#==

```
#data view
options(digits=4)
table_view=as.data.frame(matrix(c(dim(data_class),sum(data_class$Death),
sum(data_class$Treat)),1,4))
colnames(table_view)=c("columns","rows","death","treatments")
head(data_class,10)
```

```
##    id bloc true_ite.(Death) Death Treat X1      X2        X3      X4       X5
## 1   1    0          0.43496     0     1  1  0.6530 -0.004351  0.5868 -2.81646
## 2   2    0          0.13469     0     1  1 -0.3354 -0.689923  0.3452 -0.18485
## 3   3    0          0.28341     1     1  1  0.6364 -0.764598  0.3414  0.11179
## 4   4    0          0.58682     0     0  0 -0.6669  0.507079  2.6162  1.75030
## 5   5    0          0.50444     1     1  0 -0.2783  0.747434  0.4201  0.90179
## 6   6    0          0.04035     0     0  1 -0.5594 -1.434221  0.3674 -0.86756
## 7   7    0          0.12383     0     0  1 -0.1179 -0.499280 -2.0258  1.00651
## 8   8    0          0.37268     0     1  1 -1.8862  1.478520 -0.1303 -0.03908
## 9   9    0          0.15913     1     0  0  1.1972 -1.326675  1.5558 -0.65288
## 10 10    0          0.01654     0     1  0 -0.8247 -1.066176 -1.5254 -0.03394
##           X6      X7      X8      X9     X10
## 1   0.76661 -1.3398  1.0175  0.8941  0.4022
## 2  -1.28426 -0.2198  0.2567  0.4932 -0.5382
## 3  -1.05009  0.9096  0.4286 -0.6812 -1.5088
## 4   0.31831 -1.0196 -0.4373  0.8833  0.5784
## 5   0.68670 -1.4643  2.1852 -0.1837  0.2503
## 6  -1.01713  0.3859  0.2017 -1.3333 -0.5931
## 7  -0.47259  1.0194  1.7846 -1.3659  0.2302
## 8  -0.33175  0.4671 -0.5443  2.4378  0.1546
## 9   0.19008 -1.4914 -0.3171 -2.2652  2.2846
## 10  0.01192 -1.5894  1.9007  1.6740 -0.8389
```

```
table_view
```

```
##   columns rows death treatments
## 1    7500   15  2751       3448
```

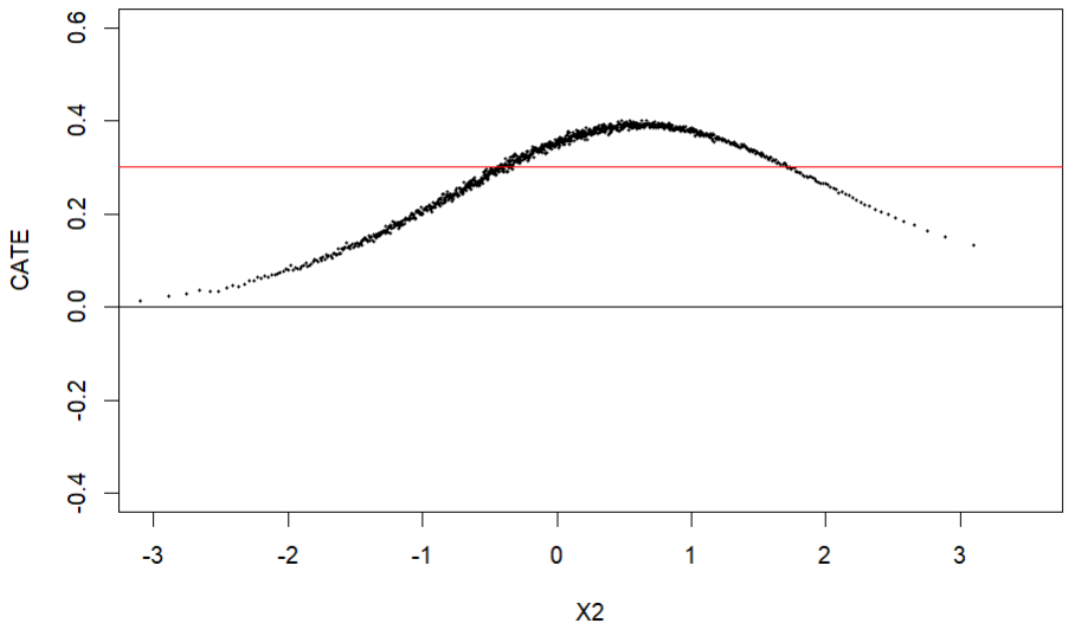
ATE

```
#calculating ATE
n=2000000
p=10
tim=c(0,1,2,3,4)
beta=c(0.5,0.5,1,1/4,1/9,0,0,0,0,0)
G=matrix(0,(p-2),(p-2))
diag(G)=0.09
theta=0

X1=matrix(0,n,(p-2))
vaso=matrix(0,n,1)
water=matrix(0,n,1)
Death=matrix(0,n,1)
```

```
treat=matrix(0,n,1)
logisticf=function(x){exp(x)/(1+exp(x))}
cd=sqrt(logisticf(tim*0)^2*(pi^2/3)/((1-logisticf(tim*0)^2)*sum(beta*beta)))
cy=sqrt((logisticf((tim*0+1.3)))^2/((1-(logisticf((tim*0+1.3)))^2)*sum(beta*beta)))

set.seed(1219)
v=matrix(c(rnorm(n*(p-2),0,1)),n,(p-2))
X2=matrix(c(rbinom(n,1,0.5),rnorm(n,0,1)),n,2)

#generating
j=1
set.seed(61916)
X1=rmvnorm(n,rep(0,(p-2)),G)+v-matrix(theta,n,p-2)
X=cbind(X2,X1)
term=c(X%*%matrix(beta,p,1)*cd[j])
g=logisticf(term)
y=-0.6+0.5*X[,2]+0.5*exp(c(X%*%matrix(beta,p,1)*cy[j]))*(1-c(treat))
deathp=tanh(y)/2+0.5

y0=-0.6+0.5*X[,2]
y0=tanh(y0)/2+0.5
y1=-0.6+0.5*X[,2]+0.5*exp(c(X%*%matrix(beta,p,1)*cy[j]))
y1=tanh(y1)/2+0.5
ATE=mean(y1-y0)
#ATE
ATE

## [1] 0.3016
```

CATE(X2)

```
#calculating CATE(X2)
ft=2
XX=sort(X[,ft])
ITE=c(y1-y0)
ITE=ITE[order(X[,ft])]
qXX=quantile(XX,c(1:1000)/1000)
CATE=rep(0,1000)
CATE[1]=mean(ITE[XX<=qXX[1]])
for(i in c(2:1000)){
  temp=ITE[XX<=qXX[i] & XX>qXX[i-1]]
  CATE[i]=mean(temp)
}
plot(qXX,CATE,pch=20,cex=0.5,xlim=c(-3,3.5),ylim=c(-0.4,0.6),xlab="X2")
abline(ATE,0,col="red")
abline(0,0)
```



#==========

#----------continuous case----------

#==========

```
#data view
options(digits=4)
table_view=as.data.frame(matrix(c(dim(data_cont),mean(data_cont$Dsofa),sum(data_cont$
Treat)),1,4))
colnames(table_view)=c("columns","rows","mean(Dsofa)","treatments")
head(data_cont,10)
```

```
##    id bloc true_ite(Dsofa)    Dsofa Treat X1      X2        X3      X4       X5
## 1   1    0         1.25872  3.49119     1  1  0.6530 -0.004351  0.5868 -2.81646
## 2   2    0         0.16055 -0.32401     1  1 -0.3354 -0.689923  0.3452 -0.18485
## 3   3    0         0.61809 -0.17642     1  1  0.6364 -0.764598  0.3414  0.11179
## 4   4    0         1.16091  1.19845     0  0 -0.6669  0.507079  2.6162  1.75030
## 5   5    0         0.99122  0.06078     1  0 -0.2783  0.747434  0.4201  0.90179
## 6   6    0         0.04150 -0.27655     0  1 -0.5594 -1.434221  0.3674 -0.86756
## 7   7    0         0.15597 -0.08961     0  1 -0.1179 -0.499280 -2.0258  1.00651
## 8   8    0         0.44414 -0.44984     1  1 -1.8862  1.478520 -0.1303 -0.03908
## 9   9    0         0.41538  0.70374     0  0  1.1972 -1.326675  1.5558 -0.65288
## 10 10    0         0.01579 -0.81310     1  0 -0.8247 -1.066176 -1.5254 -0.03394
##          X6      X7      X8      X9     X10
## 1   0.76661 -1.3398  1.0175  0.8941  0.4022
## 2  -1.28426 -0.2198  0.2567  0.4932 -0.5382
## 3  -1.05009  0.9096  0.4286 -0.6812 -1.5088
## 4   0.31831 -1.0196 -0.4373  0.8833  0.5784
## 5   0.68670 -1.4643  2.1852 -0.1837  0.2503
## 6  -1.01713  0.3859  0.2017 -1.3333 -0.5931
## 7  -0.47259  1.0194  1.7846 -1.3659  0.2302
## 8  -0.33175  0.4671 -0.5443  2.4378  0.1546
## 9   0.19008 -1.4914 -0.3171 -2.2652  2.2846
## 10  0.01192 -1.5894  1.9007  1.6740 -0.8389
```

```
table_view

##   columns rows mean(Dsofa) treatments
## 1    7500   15      0.5941       3448
```

ATE

```
#calculating ATE
n=2000000
p=10
tim=c(0,1,2,3,4)
beta=c(0.5,0.5,1,1/4,1/9,0,0,0,0,0)
G=matrix(0,(p-2),(p-2))
diag(G)=0.09
theta=0

X1=matrix(0,n,(p-2))
vaso=matrix(0,n,1)
water=matrix(0,n,1)
Death=matrix(0,n,1)
treat=matrix(0,n,1)
logisticf=function(x){exp(x)/(1+exp(x))}
cd=sqrt(logisticf(tim*0)^2*(pi^2/3)/((1-logisticf(tim*0)^2)*sum(beta*beta)))
cy=sqrt((logisticf((tim*0+1.3)))^2/((1-(logisticf((tim*0+1.3)))^2)*sum(beta*beta)))

set.seed(1219)
v=matrix(c(rnorm(n*(p-2),0,1)),n,(p-2))
X2=matrix(c(rbinom(n,1,0.5),rnorm(n,0,1)),n,2)

#generating
j=1
set.seed(61916)
X1=rmvnorm(n,rep(0,(p-2)),G)+v-matrix(theta,n,p-2)
X=cbind(X2,X1)
term=c(X%*%matrix(beta,p,1)*cd[j])
g=logisticf(term)
y=-0.6+0.5*X[,2]+0.5*exp(c(X%*%matrix(beta,p,1)*cy[j]))*(1-c(treat))
deathp=tanh(y)/2+0.5

y0=-0.6+0.5*X[,2]
y0=(1-(tanh(y0)/2+0.5)/1.285)^(-1)
y1=-0.6+0.5*X[,2]+0.5*exp(c(X%*%matrix(beta,p,1)*cy[j]))
y1=(1-(tanh(y1)/2+0.5)/1.285)^(-1)
ATE=mean(y1-y0)
#ATE
options(digits = 5)
ATE

## [1] 1.0015
```

CATE(X2)

```
##calculating CATE(X2)
ft=2
XX=sort(X[,ft])
ITE=c(y1-y0)
ITE=ITE[order(X[,ft])]
qXX=quantile(XX,c(1:1000)/1000)
CATE=rep(0,1000)
CATE[1]=mean(ITE[XX<=qXX[1]])
for(i in c(2:1000)){
  temp=ITE[XX<=qXX[i] & XX>qXX[i-1]]
  CATE[i]=mean(temp)
}
plot(qXX,CATE,pch=20,cex=0.5,xlim=c(-3,3.5))
abline(ATE,0,col="red")
abline(0,0)
```