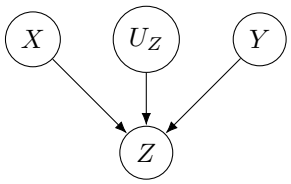
CONDITIONING ON A COLLIDER

The graph encodes the dependencies between the four random variables (X, Y, Z, U_Z): the joint density can be



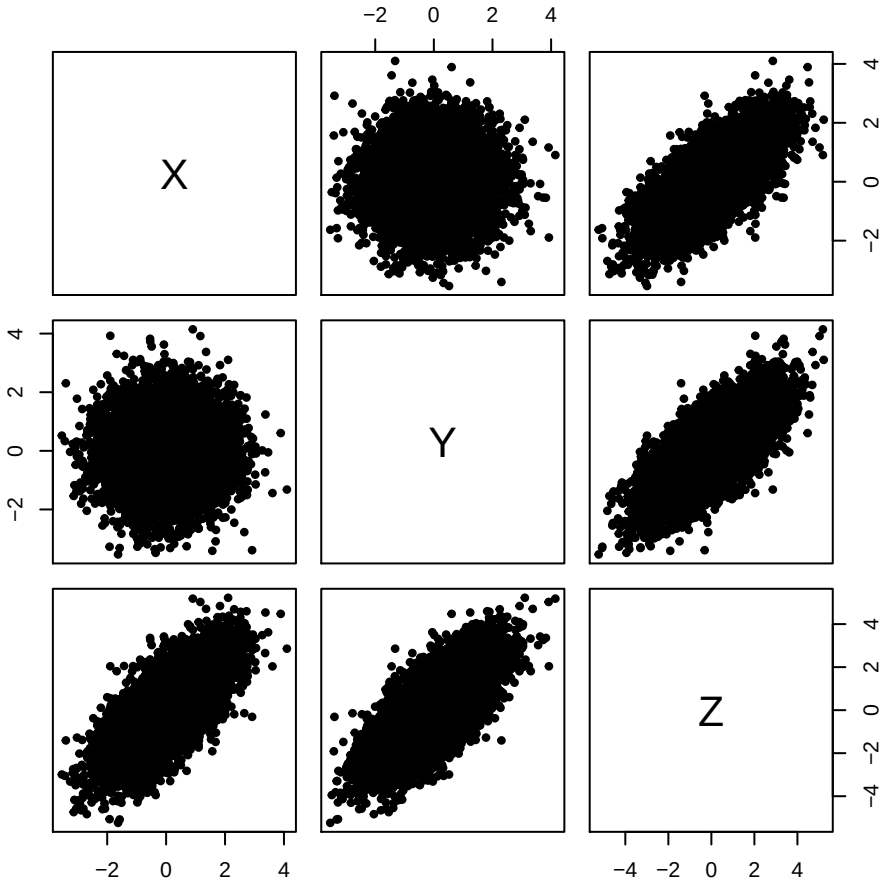
represented

$$f_{X,Y,Z,U_Z}(x, y, z, u) = f_X(x) f_Y(y) f_{U_Z}(u) f_{Z|X,Y,U_Z}(z|x, y, u)$$

that is, X, Y, U_Z are mutually independent. Suppose that X and Y are distributed as standard Normal random variables, $U_Z \sim \text{Normal}(0, 0.1^2)$, and

$$Z = X + Y + U_Z$$

```r
set.seed(2101)                          #Set the random number generator seed value
n<-10000                                #Set the sample size
X<-rnorm(n,0,1)                         #Generate the X random variables
Y<-rnorm(n,0,1)                         #Generate the Y random variables
UZ<-rnorm(n,0,0.1)                      #Generate the UZ random variables
Z<-X+Y+UZ
par(mar=c(3,2,1,0))                     #Set up the plotting margins
pairs(cbind(X,Y,Z),pch=19,cex=0.7)
```



If we condition on the value of Z, and inspect the joint density of X and Y given Z, we see that X and Y are conditionally **dependent**.

```
par(mar=c(3,2,1,0))                                    #Set up the plotting margins
X1<-X[Z>-2.5 & Z < -1.5];Y1<-Y[Z>-2.5 & Z < -1.5];   #First subset analysis
X2<-X[Z>-0.5 & Z < 0.5];Y2<-Y[Z>-0.5 & Z < 0.5];     #Second subset analysis
X3<-X[Z>0.5 & Z < 1.5];Y3<-Y[Z>0.5 & Z < 1.5];       #Third subset analysis
par(mar=c(4,3,1,0),pty='s',mfrow=c(2,2))             #Set up the plotting margins
plot(X1,Y1,pch=19,cex=0.7)
plot(X2,Y2,pch=19,cex=0.7)
plot(X3,Y3,pch=19,cex=0.7)
cor(X1,Y1)

+ [1] -0.9142475

cor(X2,Y2)

+ [1] -0.9100298

cor(X3,Y3)

+ [1] -0.9109185
```