DAG structure learning

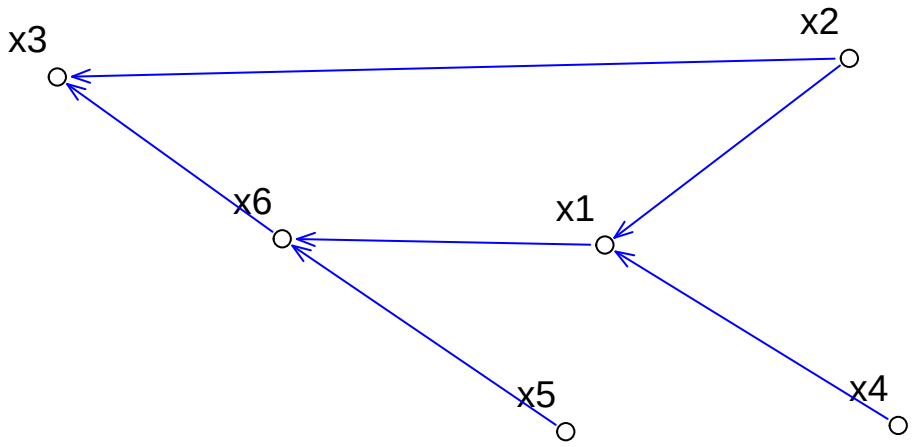
G. Marchetti

2024-10-21

The true generating process

The true DAG

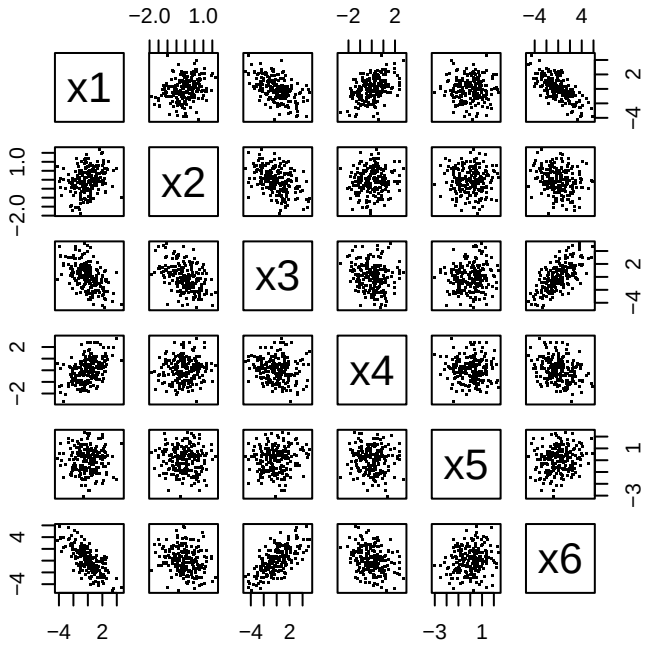
```
D <- matrix(
  c(
    1, 1, 0, 1, 0, 0,
    0, 1, 1, 0, 0, 1,
    0, 0, 1, 1, 1, 0,
    0, 0, 0, 1, 0, 0,
    0, 0, 0, 0, 1, 0,
    0, 0, 0, 0, 0, 1
  ),
  byrow = TRUE, 6, 6
)
D <- t(D- diag(6))
V <- c("x3", "x6", "x1", "x2", "x4", "x5")
dimnames(D) <- list(V,V)
co <- structure(c(9, 32, 65, 90, 95,
                  61, 77, 51, 50, 80,
                  21, 20), dim = c(6L, 2L))
drawGraph(D, coor = co)
```



The data set

```
source("simdat.R")

pairs(simdat,  pch = '.')
```




```
su= apply(simdat, 2, function(x) rbind(mean(x), sd(x)))
rownames(su) = c("mean", "sd")
print(round(su, 3))
```

```
         x1     x2     x3    x4    x5    x6
mean -0.006 -0.036 -0.049 0.041 0.029 0.030
sd    1.576  0.699  2.010 1.004 1.046 2.137
```

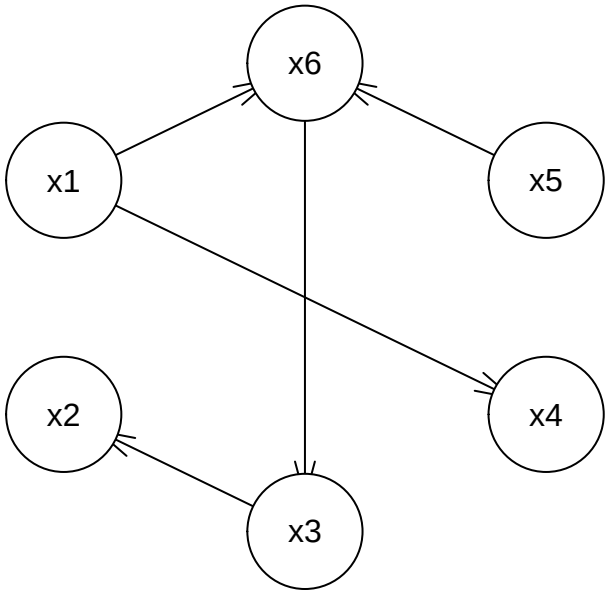
```
cor(simdat)
```

```
             x1          x2         x3          x4           x5         x6
x1  1.000000000  0.26462675 -0.5144856  0.44016843  0.001734173 -0.7165991
x2  0.264626747  1.00000000 -0.4226666  0.07493489 -0.031748666 -0.2255491
x3 -0.514485612 -0.42266665  1.0000000 -0.15835467  0.082101198  0.6547706
x4  0.440168435  0.07493489 -0.1583547  1.00000000 -0.044839369 -0.3011004
x5  0.001734173 -0.03174867  0.0821012 -0.04483937  1.000000000  0.1869784
x6 -0.716599059 -0.22554914  0.6547706 -0.30110036  0.186978417  1.0000000
```

Structure learning

Hill climbing

```
out <- hc(simdat, score = "bic-g", debug = FALSE)
f <- modelstring(out)
plot(model2network(f))
```



```
G <- DAG(x4 ~ x1, x2~x3, x1 ~ x6+x5, x5 ~ x6, x3 ~ x6 )
co <- structure(c(9, 56, 15, 74, 95, 82, 47, 48,
                  80, 80, 49, 20), dim = c(6L, 2L))
drawGraph(G, coor = co)
```