CS&SS/STAT 566 Class Lab 5

February 5, 2016

D-separation in the context of linear models

(1) Simulate a dataset from a DAG

Model 1

$$B \to X \leftarrow C$$
$$B \to Y \leftarrow C$$
$$C \to Z \leftarrow D$$
$$D \to F$$
$$Z \to R$$

```
n <- 1000

# Independent, uncorrelated error terms
set.seed(16566)
eb <- rnorm(n,0,1)
ec <- rnorm(n,0,1)
ed <- rnorm(n,0,1)
ex <- rnorm(n,0,1)
ey <- rnorm(n,0,1)
ez <- rnorm(n,0,1)
er <- rnorm(n,0,1)
ef <- rnorm(n,0,1)

b <-  2 + eb
c <- -1 + ec
d <-  2 + ed
x <-  1 + 2*b - 1*c + ex  # b --> x <-- c
y <- -1 - 1*b + 2*c + ey  # b --> y <-- c
z <- -2 + 3*c - 2*d + ez  # c --> z <-- d
f <-  1 + 1*d + ef  # d --> f
r <-  3 + 2*z + er  # z --> r
```

Suppose we didn't know the true underlying DAG from which the observed dataset was obtained. We will use d-separation to work out which coefficients should be non-zero!

(2) Marginal Independence

(i) Is $X \perp\!\!\!\perp Y$?

Hint: check coefficient of X in regression of Y on X.

```
tmp <- lm(y ~ x)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   -1.109     0.1536   -7.22  1.02e-12
## x             -0.649     0.0238  -27.20 2.54e-122
```

(ii) Is $X \perp\!\!\!\perp D$?

```
tmp <- lm(d ~ x)
summary(tmp)$coef
```

```
##              Estimate Std. Error   t value Pr(>|t|)
## (Intercept)  1.97e+00     0.0862 22.899011 1.32e-93
## x           -1.01e-05     0.0134 -0.000756 9.99e-01
```

```
# Will our conclusion change if we do the following regression instead?
tmp <- lm(x ~ d)
# summary(tmp)$coef
```

(iii) Is $B \perp\!\!\!\perp C$?

```
tmp <- lm(b ~ c)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  2.01001     0.0444 45.2714 3.68e-244
## c           -0.00269     0.0320 -0.0841  9.33e-01
```

(iv) Is $R \perp\!\!\!\perp D$?

```
tmp <- lm(r ~ d)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -6.63      0.436   -15.2 3.65e-47
## d              -4.14      0.197   -21.1 8.22e-82
```

(3) Conditional Independence

(i) Is $Y \perp\!\!\!\perp X|\{B, C\}$?

Hint: check coefficient of X in regression of Y on B, C, X.

```
tmp <- lm(y ~ b + c + x)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  -1.0682     0.0841 -12.705  2.20e-34
## b            -1.0292     0.0688 -14.952  9.41e-46
## c             2.0106     0.0441  45.547 9.79e-246
## x             0.0245     0.0309   0.791  4.29e-01
```

(ii) Is $B \perp\!\!\!\perp C | \{X, Y\}$?

```
tmp <- lm(b ~ c + x + y)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   -0.178     0.0373   -4.78  2.04e-06
## c              0.677     0.0241   28.12 1.55e-128
## x              0.330     0.0075   43.98 1.23e-235
## y             -0.178     0.0119  -14.95  9.41e-46
```

(iii) Is $C \perp\!\!\!\perp D | \{R\}$?

```
tmp <- lm(c ~ d + r)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    0.016    0.02629    0.61  5.42e-01
## d              0.577    0.01285   44.89 1.26e-241
## r              0.145    0.00172   84.29  0.00e+00
```

(iv) Is $X \perp\!\!\!\perp D | \{R\}$?

```
tmp <- lm(x ~ d + r)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    5.024     0.1703    29.5 5.88e-138
## d             -0.599     0.0833    -7.2  1.20e-12
## r             -0.145     0.0111   -13.0  1.14e-35
```

(v) Is $X \perp\!\!\!\perp D | \{C, R\}$?

```
tmp <- lm(x ~ d + c+ r)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   5.0385     0.1688  29.854 2.17e-140
## d            -0.0749     0.1434  -0.523  6.01e-01
## c            -0.9091     0.2033  -4.472  8.65e-06
## r            -0.0128     0.0315  -0.406  6.85e-01
```

(vi) Is $D \perp\!\!\!\perp B | \{R\}$?

```
tmp <- lm(d ~ b + r)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.9102    0.08038  11.324 4.72e-28
## b            -0.0180    0.02707  -0.666 5.06e-01
## r            -0.0743    0.00353 -21.055 1.04e-81
```

(vii) Is $D \perp\!\!\!\perp B | \{R, Y\}$?

```
tmp <- lm(d ~ b + r + y)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    0.891    0.06227    14.3  2.09e-42
## b              0.332    0.02497    13.3  3.37e-37
## r             -0.146    0.00389   -37.5 2.51e-192
## y              0.349    0.01352    25.8 8.21e-113
```

(viii) Is $Z \perp\!\!\!\perp F | \{D\}$?

```
tmp <- lm(z ~ f + d)
summary(tmp)$coef
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.7346     0.2362 -20.049 2.17e-75
## f            -0.0715     0.0955  -0.748 4.55e-01
## d            -2.0079     0.1356 -14.804 5.70e-45
```

Recall that both:

$$Z \leftarrow D \rightarrow F$$
$$Z \leftarrow D \leftarrow F$$
$$\Rightarrow Z \perp\!\!\!\perp F | \{D\}.$$

In our CPDAG, we will represent the edge between D and F as an undirected edge:

$$Z \leftarrow D - F$$

How many DAGS are there in the Markov equivalence class?