CS&SS/STAT 566 Class Lab

d-separation (discrete data)

Our DAG:

$B \rightarrow X \leftarrow C$;

$B \rightarrow Y \leftarrow C$.

$B \sim Binomial(2, 0.5)$; $Pr(B = 0) = Pr(B = 2) = 0.25, Pr(B = 1) = 0.5$.

$C \sim Binomial(2, 0.5)$; $Pr(C = 0) = Pr(C = 2) = 0.25, Pr(C = 1) = 0.5$.

```
# Number of simulations
nsim <- 5000
set.seed(5000)

b <- rbinom(nsim,2,0.5)
bf <- factor(b)

c <- rbinom(nsim,2,0.5)
cf <- factor(c)
```

Values of X for each value of B and C (similarly for Y).

B	C	X
0	0	0
0	1	$Bernoulli(0.5)$
1	0	$Bernoulli(0.5)$
1	1	$Bernoulli(0.5) + Bernoulli(0.5)$
0	2	1
2	0	1
1	2	$1 + Bernoulli(0.5)$
2	1	$1 + Bernoulli(0.5)$
2	2	2

```
x <- (0*(bf==0) + rbinom(nsim,1,0.5)*(bf==1) + 1*(bf==2)
      + 0*(cf==0) + rbinom(nsim,1,0.5)*(cf==1) + 1*(cf==2))
xf <- factor(x)

y <- (0*(bf==0) + rbinom(nsim,1,0.5)*(bf==1) + 1*(bf==2)
      + 0*(cf==0) + rbinom(nsim,1,0.5)*(cf==1) + 1*(cf==2))
yf <- factor(y)

name.list <- list(c(0,1,2),c(0,1,2))
```

Pairwise comparisons of the generated variables.

Marginal: B vs. C.

```
table(bf,cf,dnn=c("b","c"))
```

```
##    c
## b      0    1    2
##   0  308  629  331
##   1  634 1247  625
##   2  316  625  285
```

```
# Divide by row totals
round(table(bf,cf,dnn=c("b","c"))/matrix(rep(table(bf),3),nrow=3),2)
```

```
##    c
## b      0    1    2
##   0 0.24 0.50 0.26
##   1 0.25 0.50 0.25
##   2 0.26 0.51 0.23
```

```
# Divide by col totals
round(table(bf,cf,dnn=c("b","c"))/matrix(rep(table(cf),3),nrow=3,byrow=TRUE),2)
```

```
##    c
## b      0    1    2
##   0 0.24 0.25 0.27
##   1 0.50 0.50 0.50
##   2 0.25 0.25 0.23
```

```
# Correlation
round(cor(b,c),3)
```

```
## [1] -0.022
```

```
# Chi-squared test of independence
chisq.test(table(bf,cf,dnn=c("b","c")))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(bf, cf, dnn = c("b", "c"))
## X-squared = 2.9676, df = 4, p-value = 0.5633
```

Marginal: X vs. Y.

```
table(xf,yf,dnn=c("x","y"))
```

```
##    y
## x      0    1    2
##   0  693  480   82
##   1  500 1568  474
##   2   69  454  680
```

```
round(table(xf,yf,dnn=c("x","y"))/matrix(rep(table(xf),3),nrow=3),2)
```

```
##    y
## x      0    1    2
##   0 0.55 0.38 0.07
##   1 0.20 0.62 0.19
##   2 0.06 0.38 0.57
```

```
round(table(xf,yf,dnn=c("x","y"))/matrix(rep(table(yf),3),nrow=3,byrow=TRUE),2)
```

```
##    y
## x      0    1    2
##   0 0.55 0.19 0.07
##   1 0.40 0.63 0.38
##   2 0.05 0.18 0.55
```

```
# Correlation
round(cor(x,y),3)
```

```
## [1] 0.493
```

```
# Chi-squared test of independence
chisq.test(table(xf,yf,dnn=c("x","y")))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(xf, yf, dnn = c("x", "y"))
## X-squared = 1497.704, df = 4, p-value < 2.2e-16
```

Conditional: X vs. Y given B.

```
table(xf[(bf==0)],yf[(bf==0)],dnn=c("x","y"))
```

```
##    y
## x     0   1   2
##   0 468 145   0
##   1 176 479   0
##   2   0   0   0
```

```
table(xf[(bf==1)],yf[(bf==1)],dnn=c("x","y"))
```

```
##    y
## x     0   1   2
##   0 225 335  82
##   1 324 620 316
##   2  69 307 228
```

```
table(xf[(bf==2)],yf[(bf==2)],dnn=c("x","y"))
```

```
##    y
## x     0   1   2
##   0   0   0   0
##   1   0 469 158
##   2   0 147 452
```

```
names(name.list) <- c("x","y")

## Dividing by row totals
round(matrix(table(xf[(bf==0)],yf[(bf==0)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(xf[bf==0]),3),nrow=3),2)
```

```
##    y
## x      0    1   2
##   0 0.76 0.24   0
##   1 0.27 0.73   0
##   2  NaN  NaN NaN
```

```
round(matrix(table(xf[(bf==1)],yf[(bf==1)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(xf[bf==1]),3),nrow=3),2)
```

```
##    y
## x      0    1    2
##   0 0.35 0.52 0.13
##   1 0.26 0.49 0.25
##   2 0.11 0.51 0.38
```

```
round(matrix(table(xf[(bf==2)],yf[(bf==2)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(xf[bf==2]),3),nrow=3),2)
```

```
##    y
## x     0    1    2
##   0 NaN  NaN  NaN
##   1   0 0.75 0.25
##   2   0 0.25 0.75
```

```
## Dividing by col totals
round(matrix(table(xf[(bf==0)],yf[(bf==0)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(yf[bf==0]),3),nrow=3,byrow=TRUE),2)
```

```
##    y
## x     0    1   2
##   0 0.73 0.23 NaN
##   1 0.27 0.77 NaN
##   2 0.00 0.00 NaN
```

```
round(matrix(table(xf[(bf==1)],yf[(bf==1)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(yf[bf==1]),3),nrow=3,byrow=TRUE),2)
```

```
##    y
## x     0    1    2
##   0 0.36 0.27 0.13
##   1 0.52 0.49 0.50
##   2 0.11 0.24 0.36
```

```
round(matrix(table(xf[(bf==2)],yf[(bf==2)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(yf[bf==2]),3),nrow=3,byrow=TRUE),2)
```

```
##    y
## x     0    1    2
##   0 NaN 0.00 0.00
##   1 NaN 0.76 0.26
##   2 NaN 0.24 0.74
```

```
# Correlations
round(cor(x[(bf==0)],y[(bf==0)]),3)
```

```
## [1] 0.495
```

```
round(cor(x[(bf==1)],y[(bf==1)]),3)
```

```
## [1] 0.243
```

```
round(cor(x[(bf==2)],y[(bf==2)]),3)
```

```
## [1] 0.502
```

```r
# Chi-squared test of independence
chisq.test(table(xf[(bf==1)],yf[(bf==1)],dnn=c("x","y")))
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(xf[(bf == 1)], yf[(bf == 1)], dnn = c("x", "y"))
## X-squared = 150.0687, df = 4, p-value < 2.2e-16
```

Conditional: B vs. C given X.

```
table(bf[(xf==0)],cf[(xf==0)],dnn=c("b","c"))
```

```
##    c
## b     0   1   2
##   0 308 305   0
##   1 321 321   0
##   2   0   0   0
```

```
table(bf[(xf==1)],cf[(xf==1)],dnn=c("b","c"))
```

```
##    c
## b     0   1   2
##   0   0 324 331
##   1 313 623 324
##   2 316 311   0
```

```
table(bf[(xf==2)],cf[(xf==2)],dnn=c("b","c"))
```

```
##    c
## b     0   1   2
##   0   0   0   0
##   1   0 303 301
##   2   0 314 285
```

```
names(name.list) <- c("b","c")

## Dividing by row totals
round(matrix(table(bf[(xf==0)],cf[(xf==0)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(bf[xf==0]),3),nrow=3),2)
```

```
##    c
## b     0   1   2
##   0 0.5 0.5   0
##   1 0.5 0.5   0
##   2 NaN NaN NaN
```

```
round(matrix(table(bf[(xf==1)],cf[(xf==1)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(bf[xf==1]),3),nrow=3),2)
```

```
##    c
## b      0    1    2
##   0 0.00 0.49 0.51
##   1 0.25 0.49 0.26
##   2 0.50 0.50 0.00
```

```
round(matrix(table(bf[(xf==2)],cf[(xf==2)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(bf[xf==2]),3),nrow=3),2)
```

```
##    c
## b     0    1    2
##   0 NaN  NaN  NaN
##   1   0 0.50 0.50
##   2   0 0.52 0.48
```

```
## Dividing by col totals
round(matrix(table(bf[(xf==0)],cf[(xf==0)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(cf[xf==0]),3),nrow=3,byrow=TRUE),2)
```

```
##    c
## b      0    1   2
##   0 0.49 0.49 NaN
##   1 0.51 0.51 NaN
##   2 0.00 0.00 NaN
```

```
round(matrix(table(bf[(xf==1)],cf[(xf==1)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(cf[xf==1]),3),nrow=3,byrow=TRUE),2)
```

```
##    c
## b     0    1    2
##   0 0.0 0.26 0.51
##   1 0.5 0.50 0.49
##   2 0.5 0.25 0.00
```

```
round(matrix(table(bf[(xf==2)],cf[(xf==2)]),nrow=3,dimnames=name.list)/
        matrix(rep(table(cf[xf==2]),3),nrow=3,byrow=TRUE),2)
```

```
##    c
## b     0    1    2
##   0 NaN 0.00 0.00
##   1 NaN 0.49 0.51
##   2 NaN 0.51 0.49
```

```
# Correlations
round(cor(b[(xf==0)],c[(xf==0)]),3)
```

```
## [1] 0.002
```

```
round(cor(b[(xf==1)],c[(xf==1)]),3)
```

```
## [1] -0.504
```

```
round(cor(b[(xf==2)],c[(xf==2)]),3)
```

```
## [1] -0.023
```

Conditional: X vs. Y given B and C.

```
table(xf[(bf==0)&(cf==0)],yf[(bf==0)&(cf==0)],dnn=c("x","y"))
```

```
##    y
## x     0   1   2
##   0 308   0   0
##   1   0   0   0
##   2   0   0   0
```

```
for (i in 0:2){
    for (j in 0:2){
        cat(paste("  B =",i,";  C =",j,"\n"))
        tbl <- table(xf[(bf==i)&(cf==j)],yf[(bf==i)&(cf==j)],dnn=c("x","y"))
        print(tbl)
        cat("============\n")
        }
    }
```

```
##   B = 0 ;  C = 0
##    y
## x     0   1   2
##   0 308   0   0
##   1   0   0   0
##   2   0   0   0
## ============
##   B = 0 ;  C = 1
##    y
## x     0   1   2
##   0 160 145   0
##   1 176 148   0
##   2   0   0   0
## ============
##   B = 0 ;  C = 2
##    y
## x     0   1   2
##   0   0   0   0
##   1   0 331   0
##   2   0   0   0
## ============
##   B = 1 ;  C = 0
##    y
## x     0   1   2
##   0 155 166   0
##   1 159 154   0
##   2   0   0   0
## ============
##   B = 1 ;  C = 1
##    y
## x     0   1   2
##   0  70 169  82
##   1 165 302 156
```

```
##   2  69 146  88
## ============
##   B = 1 ;  C = 2
##    y
## x     0   1   2
##   0   0   0   0
##   1   0 164 160
##   2   0 161 140
## ============
##   B = 2 ;  C = 0
##    y
## x     0   1   2
##   0   0   0   0
##   1   0 316   0
##   2   0   0   0
## ============
##   B = 2 ;  C = 1
##    y
## x     0   1   2
##   0   0   0   0
##   1   0 153 158
##   2   0 147 167
## ============
##   B = 2 ;  C = 2
##    y
## x     0   1   2
##   0   0   0   0
##   1   0   0   0
##   2   0   0 285
## ============
```

```
for (i in 0:2){
    for (j in 0:2){
        cat(paste("  B =",i,";  C =",j," :", round(cor(x[(bf==i)&(cf==j)],y[(bf==i)&(cf==j)]),3),"\n"))
}
}
```

```
## Warning in cor(x[(bf == i) & (cf == j)], y[(bf == i) & (cf == j)]): the
## standard deviation is zero
```

```
##   B = 0 ;  C = 0  : NA
##   B = 0 ;  C = 1  : -0.019
```

```
## Warning in cor(x[(bf == i) & (cf == j)], y[(bf == i) & (cf == j)]): the
## standard deviation is zero
```

```
##   B = 0 ;  C = 2  : NA
##   B = 1 ;  C = 0  : -0.025
##   B = 1 ;  C = 1  : 0.012
##   B = 1 ;  C = 2  : -0.029
```

```
## Warning in cor(x[(bf == i) & (cf == j)], y[(bf == i) & (cf == j)]): the
## standard deviation is zero
```

```
##   B = 2 ;  C = 0  : NA
##   B = 2 ;  C = 1  : 0.024
```

```
## Warning in cor(x[(bf == i) & (cf == j)], y[(bf == i) & (cf == j)]): the
## standard deviation is zero
```

```
##   B = 2 ;  C = 2  : NA
```

```
### Conditional
## B vs. C given X, Y

table(bf[(xf==0)&(yf==0)],cf[(xf==0)&(yf==0)],dnn=c("b","c"))
```

```
##    c
## b     0   1   2
##   0 308 160   0
##   1 155  70   0
##   2   0   0   0
```

```
for (i in 0:2){
    for (j in 0:2){
        cat(paste("  X =",i,";  Y =",j))
        tbl <- table(bf[(xf==i)&(yf==j)],cf[(xf==i)&(yf==j)],dnn=c("b","c"))
        print(tbl)
        cat("============\n")
        }
    }
```

```
##   X = 0 ;  Y = 0   c
## b     0   1   2
##   0 308 160   0
##   1 155  70   0
##   2   0   0   0
## ============
##   X = 0 ;  Y = 1   c
## b     0   1   2
##   0   0 145   0
##   1 166 169   0
##   2   0   0   0
## ============
##   X = 0 ;  Y = 2   c
## b    0  1  2
##   0  0  0  0
##   1  0 82  0
##   2  0  0  0
## ============
##   X = 1 ;  Y = 0   c
## b     0   1   2
##   0   0 176   0
##   1 159 165   0
##   2   0   0   0
## ============
```

```
##   X = 1 ;  Y = 1   c
## b     0   1   2
##   0   0 148 331
##   1 154 302 164
##   2 316 153   0
## ============
##   X = 1 ;  Y = 2   c
## b     0   1   2
##   0   0   0   0
##   1   0 156 160
##   2   0 158   0
## ============
##   X = 2 ;  Y = 0   c
## b    0  1  2
##   0  0  0  0
##   1  0 69  0
##   2  0  0  0
## ============
##   X = 2 ;  Y = 1   c
## b     0   1   2
##   0   0   0   0
##   1   0 146 161
##   2   0 147   0
## ============
##   X = 2 ;  Y = 2   c
## b     0   1   2
##   0   0   0   0
##   1   0  88 140
##   2   0 167 285
## ============
```

```
for (i in 0:2){
    for (j in 0:2){
        cat(paste("  X =",i,";  Y =",j," :", round(cor(b[(xf==i)&(yf==j)],c[(xf==i)&(yf==j)]),3),"\n"))
}
}
```

```
##   X = 0 ;  Y = 0  : -0.031
##   X = 0 ;  Y = 1  : -0.478

## Warning in cor(b[(xf == i) & (yf == j)], c[(xf == i) & (yf == j)]): the
## standard deviation is zero

##   X = 0 ;  Y = 2  : NA
##   X = 1 ;  Y = 0  : -0.503
##   X = 1 ;  Y = 1  : -0.676
##   X = 1 ;  Y = 2  : -0.505

## Warning in cor(b[(xf == i) & (yf == j)], c[(xf == i) & (yf == j)]): the
## standard deviation is zero

##   X = 2 ;  Y = 0  : NA
##   X = 2 ;  Y = 1  : -0.513
##   X = 2 ;  Y = 2  : 0.016
```