Causal Graphs for Basic Epidemiologic Data

Part 2—The story behind the data

Tomás J. Aragón

2019-11-22

Case studies revisited: stories behind the data

Data are not sufficient to draw causal inferences: we must know how the data was generated. In other words, we must know the "story behind the data."

Case study 1

This was an observational study where 700 patients were given access to a new drug for an ailment. A total of 350 patients chose to take the drug and 350 patients did not. The patients were assessed for clinical recovery. Here are some additional facts:

- Estrogen has a negative effect on recovery
- Women are more likely to take the drug compared to men

To analyze this data we need to

- Design a causal graph based on the "story behind the data."
- Set the conditional probabilities
- Evaluate the primary causal question (next part)

```
library(bnlearn)
jpdrugrx.bn <- jpdrugrx[, c("Recovered", "Drug", "Gender")]
names(jpdrugrx.bn) <- c("R", "D", "G")
dag1 <- empty.graph(nodes = c("D", "R", "G"))
dag1 <- set.arc(dag1, from = "D", to = "R")
dag1 <- set.arc(dag1, from = "G", to = "D")
dag1 <- set.arc(dag1, from = "G", to = "R")
plot(dag1, main = "Directed acyclic graph for Case study 1")
```

Directed acyclic graph for Case study 1
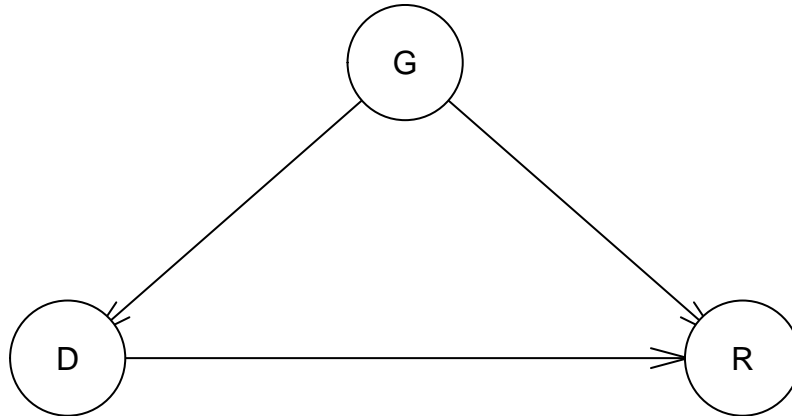

```
dag1
```

```
## 
##   Random/Generated Bayesian network
## 
##   model:
##    [G][D|G][R|D:G]
##   nodes:                              3
##   arcs:                               3
##     undirected arcs:                  0
##     directed arcs:                    3
##   average markov blanket size:        2.00
##   average neighbourhood size:         2.00
##   average branching factor:           1.00
## 
##   generation algorithm:               Empty
```

```
modelstring(dag1)
```

```
## [1] "[G][D|G][R|D:G]"
```

```
nodes(dag1)
```

```
## [1] "D" "R" "G"
```

```
arcs(dag1)
```

```
##      from to
## [1,] "D"  "R"
## [2,] "G"  "D"
## [3,] "G"  "R"
```

Set conditional probability tables (CPTs) from the data

```
(bn1.mle <- bn.fit(dag1, data = jpdrugrx.bn))
```

```
##
##   Bayesian network parameters
##
##   Parameters of node D (multinomial distribution)
##
## Conditional probability table:
##
##      G
## D          Men      Women
##   No  0.7563025 0.2332362
##   Yes 0.2436975 0.7667638
##
##   Parameters of node R (multinomial distribution)
##
## Conditional probability table:
##
## , , G = Men
##
##      D
## R            No        Yes
##   No  0.13333333 0.06896552
##   Yes 0.86666667 0.93103448
##
## , , G = Women
##
##      D
## R            No        Yes
##   No  0.31250000 0.26996198
##   Yes 0.68750000 0.73003802
##
##
##   Parameters of node G (multinomial distribution)
##
## Conditional probability table:
##    Men Women
##   0.51  0.49
```

Set conditional probability tables (CPTs) from external sources

```
G.lv <- c("Men", "Women")
D.lv <- c("No", "Yes")
```

```
R.lv <- c("No", "Yes")
G.prob <- array(c(0.51, 0.49), dim = 2, dimnames = list(G = G.lv))
D.prob <- array(c(0.76, 0.24, 0.23, 0.77), dim = c(2, 2),
    dimnames = list(D = D.lv, G = G.lv))
R.prob <- array(c(0.13, 0.87, 0.07, 0.93, 0.31, 0.69, 0.27, 0.73),
    dim = c(2, 2, 2), dimnames = list(R = R.lv, D = D.lv, G = G.lv))
(cpt1 <- list(G = G.prob, D = D.prob, R = R.prob))
```

```
## $G
## G
##   Men Women
##  0.51  0.49
##
## $D
##      G
## D     Men Women
##   No  0.76  0.23
##   Yes 0.24  0.77
##
## $R
## , , G = Men
##
##      D
## R      No  Yes
##   No  0.13 0.07
##   Yes 0.87 0.93
##
## , , G = Women
##
##      D
## R      No  Yes
##   No  0.31 0.27
##   Yes 0.69 0.73
```

```
(bn1 <- custom.fit(dag1, dist = cpt1))
```

```
##
##   Bayesian network parameters
##
##   Parameters of node D (multinomial distribution)
##
## Conditional probability table:
##
##      G
## D     Men Women
##   No  0.76  0.23
```

```
##   Yes 0.24  0.77
##
##   Parameters of node R (multinomial distribution)
##
## Conditional probability table:
##
## , , G = Men
##
##      D
## R       No  Yes
##   No  0.13 0.07
##   Yes 0.87 0.93
##
## , , G = Women
##
##      D
## R       No  Yes
##   No  0.31 0.27
##   Yes 0.69 0.73
##
##
##   Parameters of node G (multinomial distribution)
##
## Conditional probability table:
##  G
##   Men Women
##  0.51  0.49
```

Case study 2

Again data are not sufficient to draw causal inferences: we must know how the data was generated. In other words, we must know the "story behind the data."

This was a treatment study with 700 patients, half of whom were assigned a new drug for their ailment. At the end of the study the patients were assessed for clinical recovery, and their blood pressure was measured. Here are some additional facts:

- Blood pressure was measured at the end of the study
- Drug treatment affects recovery by lowering blood pressure
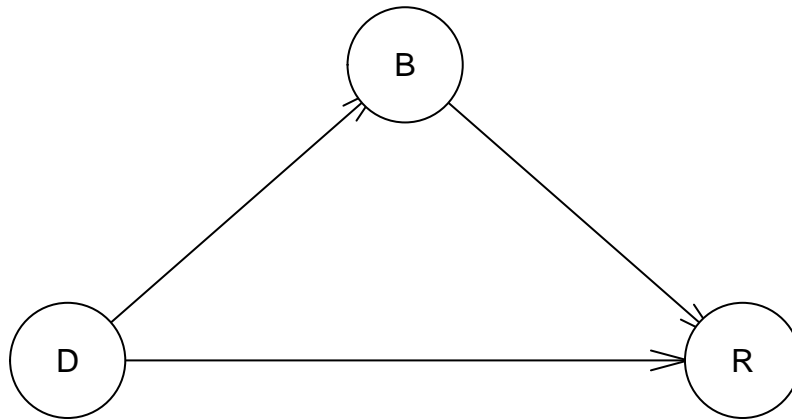- Lowering blood pressure also has toxic side effects

```
library(bnlearn)
jpdrugrx2.bn <- jpdrugrx2[, c("Recovered", "Drug", "BP")]
names(jpdrugrx2.bn) <- c("R", "D", "B")
dag2 <- empty.graph(nodes = c("D", "R", "B"))
dag2 <- set.arc(dag2, from = "D", to = "R")
```

```
dag2 <- set.arc(dag2, from = "D", to = "B")
dag2 <- set.arc(dag2, from = "B", to = "R")
plot(dag2, main = "Directed acyclic graph for Case study 2")
```

Directed acyclic graph for Case study 2


```
dag2
```

```
##
##   Random/Generated Bayesian network
##
##   model:
##    [D][B|D][R|D:B]
##   nodes:                              3
##   arcs:                               3
##     undirected arcs:                  0
##     directed arcs:                    3
##   average markov blanket size:        2.00
##   average neighbourhood size:         2.00
##   average branching factor:           1.00
##
##   generation algorithm:               Empty
```

```
modelstring(dag2)
```

```
## [1] "[D][B|D][R|D:B]"
```

```
nodes(dag2)
```

```
## [1] "D" "R" "B"
```

```
arcs(dag2)
```

```
##      from to
## [1,] "D"  "R"
## [2,] "D"  "B"
## [3,] "B"  "R"
```

Set conditional probabilities from the data

```
(bn2.mle <- bn.fit(dag2, data = jpdrugrx2.bn))
```

```
##
##   Bayesian network parameters
##
##   Parameters of node D (multinomial distribution)
##
## Conditional probability table:
##   No Yes
## 0.5 0.5
##
##   Parameters of node R (multinomial distribution)
##
## Conditional probability table:
##
## , , B = High
##
##       D
## R            No        Yes
##   No  0.26996198 0.31250000
##   Yes 0.73003802 0.68750000
##
## , , B = Low
##
##       D
## R            No        Yes
##   No  0.06896552 0.13333333
##   Yes 0.93103448 0.86666667
##
##
##   Parameters of node B (multinomial distribution)
##
## Conditional probability table:
##
##       D
## B            No        Yes
##   High 0.7514286 0.2285714
##   Low  0.2485714 0.7714286
```

7

Set conditional probabilities from external sources

```
D.lv <- c("No", "Yes")
B.lv <- c("High", "Low")
R.lv <- c("No", "Yes")
D.prob <- array(c(0.5, 0.5), dim = 2, dimnames = list(D = D.lv))
B.prob <- array(c(0.75, 0.25, 0.23, 0.77), dim = c(2, 2),
    dimnames = list(B = B.lv, D = D.lv))
R.prob <- array(c(0.27, 0.73, 0.31, 0.69, 0.07, 0.93, 0.13, 0.87),
    dim = c(2, 2, 2), dimnames = list(R = R.lv, D = D.lv, B = B.lv))
(cpt2 <- list(D = D.prob, B = B.prob, R = R.prob))
```

```
## $D
## D
##  No Yes
## 0.5 0.5
##
## $B
##        D
## B        No  Yes
##   High 0.75 0.23
##   Low  0.25 0.77
##
## $R
## , , B = High
##
##      D
## R       No  Yes
##   No  0.27 0.31
##   Yes 0.73 0.69
##
## , , B = Low
##
##      D
## R       No  Yes
##   No  0.07 0.13
##   Yes 0.93 0.87
```

```
(bn2 <- custom.fit(dag2, dist = cpt2))
```

```
##
##   Bayesian network parameters
##
##   Parameters of node D (multinomial distribution)
##
## Conditional probability table:
```

```
##   D
##   No Yes
## 0.5 0.5
##
##    Parameters of node R (multinomial distribution)
##
## Conditional probability table:
##
## , , B = High
##
##        D
## R        No   Yes
##    No   0.27 0.31
##    Yes  0.73 0.69
##
## , , B = Low
##
##        D
## R        No   Yes
##    No   0.07 0.13
##    Yes  0.93 0.87
##
##
##    Parameters of node B (multinomial distribution)
##
## Conditional probability table:
##
##         D
## B        No   Yes
##    High 0.75 0.23
##    Low  0.25 0.77
```