Computational Statistics & Probability

Lab 3 - Causal Models

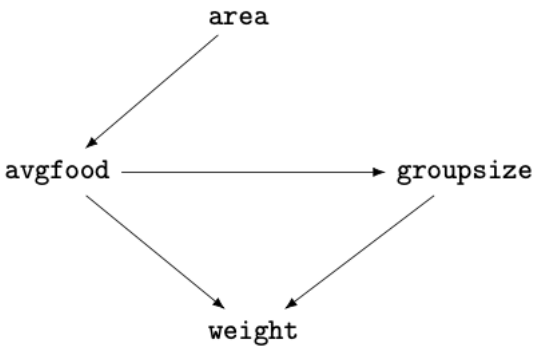
Fall 2022

1. English Foxes

Urban foxes are like street gangs. Groups vary in size between two and eight foxes and each group maintains its own (almost exclusive) urban territory. Some territories are larger than others. the data set `foxes` in the `rethinking` package consists of data for 116 foxes from 30 different urban groups in England.

```
library(rethinking)
data(foxes)
d <- foxes
```

The variable `area` encodes that some territories are larger than others. The variable `avgfood` encodes that some territories have more average food than others. Suppose we wants to model the `weight` of each fox. For our purposes here, assume the following causal DAG:



a) Does territory size causally influence the weight of foxes? Construct a quadratic approximation (*quap*) model to infer the total causal influence of `area` on `weight`. Does increasing the area available to each fox make it healthier (i.e., heavier)? You should standardize your variables and use prior predictive simulations to show that your model's predictions stay within the range of possible outcome values.

```
# From the DAG we can see there are no backdoor paths from `area` to `weight`.
# Unsure? Use `daggity` to check:
library(dagitty)
dag_foxes <- dagitty("dag{
                     A -> F
                     F -> G -> W
                     F -> W}"
                     )

# adjustment set to assess the effect of exposing A to outcome W:
adjustmentSets( dag_foxes, exposure = "A", outcome="W" )
```

```
##  {}
```

```
# which returns the empty set, {}

# So, we only need to include the variable `area`. Here is a model with
# standardized variables and standardized priors (i.e., mean = 0):

d$A <- standardize(d$area)
d$W <- standardize(d$weight)

m1a <- quap(
  alist (
    W ~ dnorm( mu , sigma) ,
    mu <- a + bA*A ,
    a ~ dnorm( 0, 0.2 ) ,
    bA ~ dnorm( 0, 0.5 ) ,
    sigma ~ dexp( 1 )
  ), data = d
)
precis( m1a )
```

```
##               mean         sd       5.5%     94.5%
## a     3.963631e-09 0.08360869 -0.1336228 0.1336228
## bA    1.883364e-02 0.09089584 -0.1264355 0.1641027
## sigma 9.912663e-01 0.06466652  0.8879167 1.0946159
```

```
# The coefficient bA is very close to 0 (~ 0.02). Thus, within this data set, territory
# size (`area`) has no causal influence on the weight of foxes.
```

b) Now infer the causal influence of adding food (`avgfood`) to a territory. Would the introduction of more food make foxes heavier? Which covariates do you need to adjust to estimate the total causal influence of food?

```
# From the DAG we see there are no backdoor paths from `avgfood` to `weight`.
# Again, we can confirm this with dagitty:

# adjustment set to assess the effect of exposing A to outcome W:
adjustmentSets( dag_foxes, exposure = "F", outcome="W" )
```

```
##  {}
```

```
# which returns the empty set, {}. Thus, we only need to include the variable `avgfood`.

#Here is a model with standardized variables and standardized priors.

d$F <- standardize( d$avgfood )
m1b <- quap(
  alist (
    W ~ dnorm( mu, sigma ),
    mu <- a + bF*F,
    a ~ dnorm( 0, 0.2 ),
    bF ~ dnorm( 0, 0.5 ),
    sigma ~ dexp( 1 )
  ), data=d)
precis( m1b )
```

```
##                mean         sd       5.5%     94.5%
## a     -1.984744e-06 0.08360105 -0.1336126 0.1336086
```

```
## bF    -2.421188e-02 0.09088613 -0.1694655 0.1210417
## sigma  9.911566e-01 0.06466063  0.8878164 1.0944967
```

```
# Within this "small world" data set, `avgfood` does not change the weight of foxes.
# This result shouldl be expected if the DAG is correct: `avgfood` comes after
# `area` on the causal paths to `weight`.  Indeed, if we did see an influence, we
# would then have grounds to reject this DAG.
```

c) Now infer the causal influence of group size (`groupsize`). Which covariates do you need to adjust to make this estimate? Inspect the posterior distribution of the resulting model. What do you think explains these data? Specifically, explain the estimates of the effects of `area`, `avgfood`, and `groupsize` on `weight`. How do they make sense together? (Hint: we covered an example in class which exhibited a similar relationship between predictors and outcome variables.)

```
# From the DAG we see there is a backdoor from `groupsize` to `weight` through
# `avgfood`.  We can confirm this with dagitty:

# adjustment set to assess the effect of exposing A to outcome W:
adjustmentSets( dag_foxes, exposure = "G", outcome="W" )
```

```
## { F }
```

```
# which returns the set, {F}.

# Thus, to infer the causal influence of `groupsize` on `weight`, we need to
# include `avgfood` in addition to `groupsize`. Here is a model with standardized
# variables and standardized priors:

d$G <- standardize(d$groupsize)
m1c <- quap(
  alist(
    W ~ dnorm( mu , sigma ), mu <- a + bF*F + bG*G,
    a ~ dnorm( 0,0.2 ),
    bF ~ dnorm( 0,0.5 ),
    bG ~ dnorm( 0,0.5 ),
    sigma ~ dexp( 1 )
    ), data=d
  )
precis( m1c )
```

```
##                mean         sd       5.5%      94.5%
## a     -4.938127e-10 0.08013807 -0.1280761  0.1280761
## bF     4.772531e-01 0.17912322  0.1909796  0.7635266
## bG    -5.735258e-01 0.17914172 -0.8598289 -0.2872228
## sigma  9.420439e-01 0.06175256  0.8433514  1.0407365
```

```
# With model m1c we see that there is a MASKED EFFECT between `avgfood` and
# `groupsize` on the `weight` of foxes.  Specifically, `avgfood` is positively
# associated with `weight`, controlling for group size, while `groupsize` is
# negatively associated with `weight`, controlling for food.

# Put differently, more food increases weight but larger group size decreases
# weight, and the total cause of food and group size cancel each other out.
```