Week 2 Problems

Juan Pablo Fonseca-Zamora

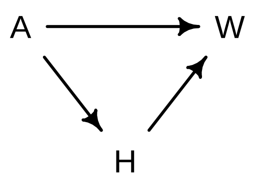
1. From the `Howell1` dataset, consider only the people younger than 13 years old. Estimate the casual association between age and weight. Assume that age influences weight through two paths. First, age influences height, and height influences weight. Second, age directly influences weight through age-related changes in muscle growth and body proportions. Draw the DAG that represents these casual relationships. And then write a generative simulation that takes age as an input and simulates height and weight, obeying the relationships in the DAG.

Solution

```
library(rethinking)
library(dagitty)
```

The DAG showing the two paths A -> H -> W and A -> W.

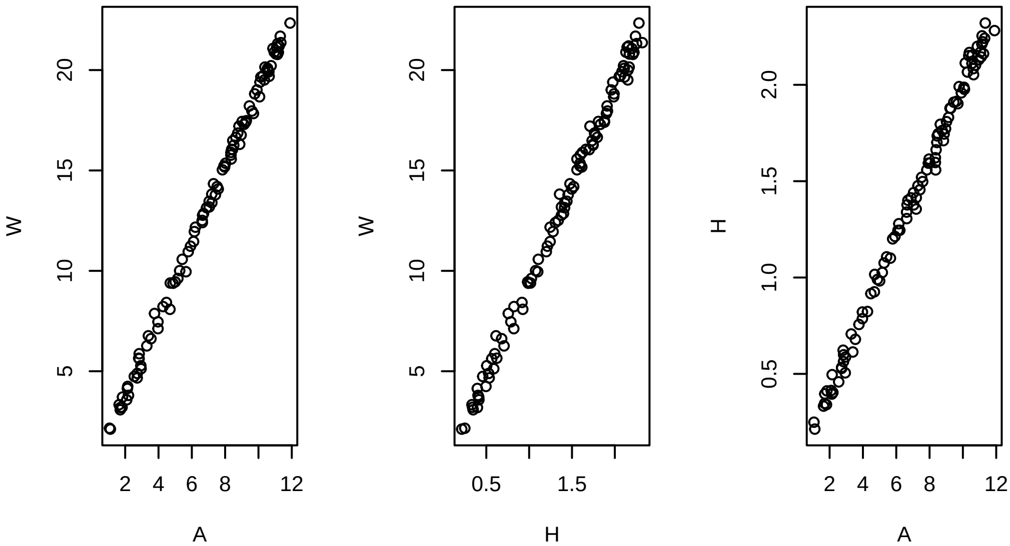
```
dag_w2e1 <- dagitty("dag{A->H;H->W;A->W }")

coordinates(dag_w2e1) <- list(x=c(A=0,W=2,H=1),y=c(A=0,W=0,H=1))

drawdag(dag_w2e1)
```



The generative model

```
sim_HW <- function(age, # The age of the individual(s)
                   bAH, # Beta for the influence of age over height
                   bHW, # Beta for the influence of height over weight
                   bAW  # Beta for the influence of age over weight
                   ){
  N <- length(age)
  H <- rnorm(N,bAH*age,0.05)
  W <- rnorm(N,bHW*H+bAW*age,0.3)
  data.frame(A = age,H,W)
}
```

```r
samples_HW = sim_HW(runif(100,1,12),
                    bAH = 0.2,
                    bHW = 2,
                    bAW = 1.5)
par(mfrow = c(1, 3))
plot(W~A,data = samples_HW)
plot(W~H,data = samples_HW)
plot(H~A,data = samples_HW)
```



2. Estimate the **total** causal effect of each year of growth on weight.

Solution

The total effect of each year of age can be estimated with a linear model

```r
# Subset of data
data("Howell1")
d = Howell1[Howell1$age<13,]

mex2 = quap(
  alist(
    weight ~ dnorm(mu,sigma),
    mu <-  a + b*age,
    a ~ dnorm(3,5),
    b ~ dunif(0,10),
    sigma ~ dunif(0,10)
  ),
  data = d
)
```

The distribution of `b` corresponds to the effect of each year of growth on weight

```
precis(mex2)[2,]
```

```
      mean         sd     5.5%    94.5%
b 1.344614 0.05471316 1.257172 1.432056
```

3 - OPTIONAL CHALLENGE. The data in `data(Oxboys)` (`rethinking` package) are growth records for 26 boys measured over 9 periods. I want you to model their growth. Specifically, model the increments in growth from one period (`Occasion` in the data table) to the next. Each increment is simply the difference between height in one occasion and the height in the previous occasion. Since none of these boys shrunk during the study, all of the growth increments are grater than zero. Estimate the posterior distribution of these increments. Constrain the distribution so it is always positive -it should not be possible for the model to think that boys can shrink from year to year. Finally compute the posterior distribution of the total growth over all 9 occasions.

Solution

```
data("Oxboys")
d <- Oxboys
```

Arranging the dataset to calculate the differences

```
d <- d[order(d$Subject,d$Occasion),]
```

Making the growth for the first occasion `NA`

```
d$growth <- c(NA_real_,diff(d$height,lag = 1))
d$growth[d$Occasion==1] <- NA
```

`NA` values are discarded.

```
d <- d[!is.na(d$growth),]
```

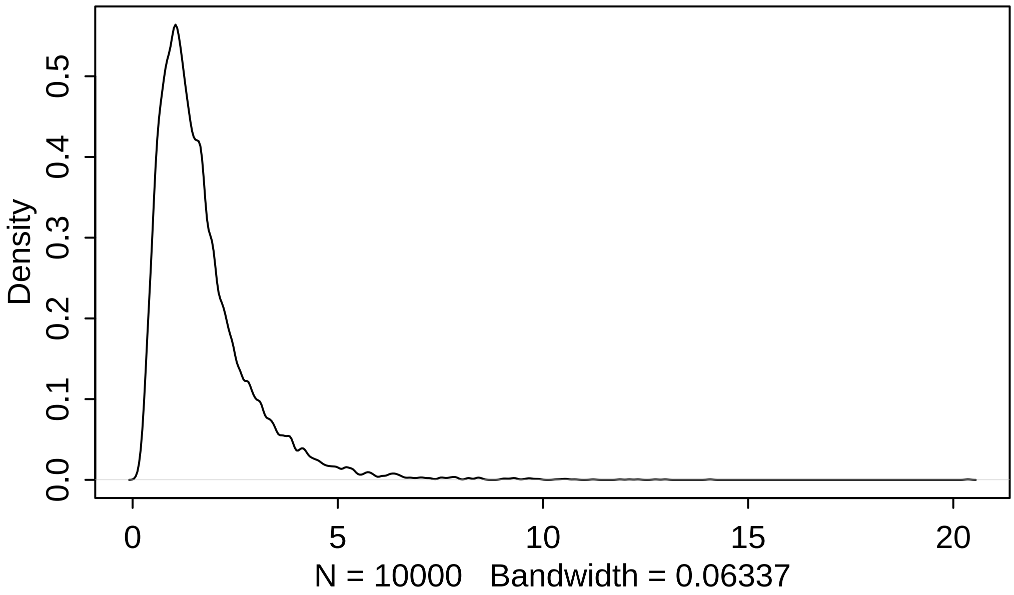
To compute the posterior

```
mex4 <- quap(alist(
  growth ~ dlnorm(mu,sigma),
  mu ~ dnorm(1,1),
  sigma ~ dexp(1)),
  data = d
)

precis(mex4)
```

```
           mean          sd      5.5%     94.5%
mu    0.3436076 0.04284802 0.2751282 0.4120870
sigma 0.6185287 0.03025981 0.5701676 0.6668897
```

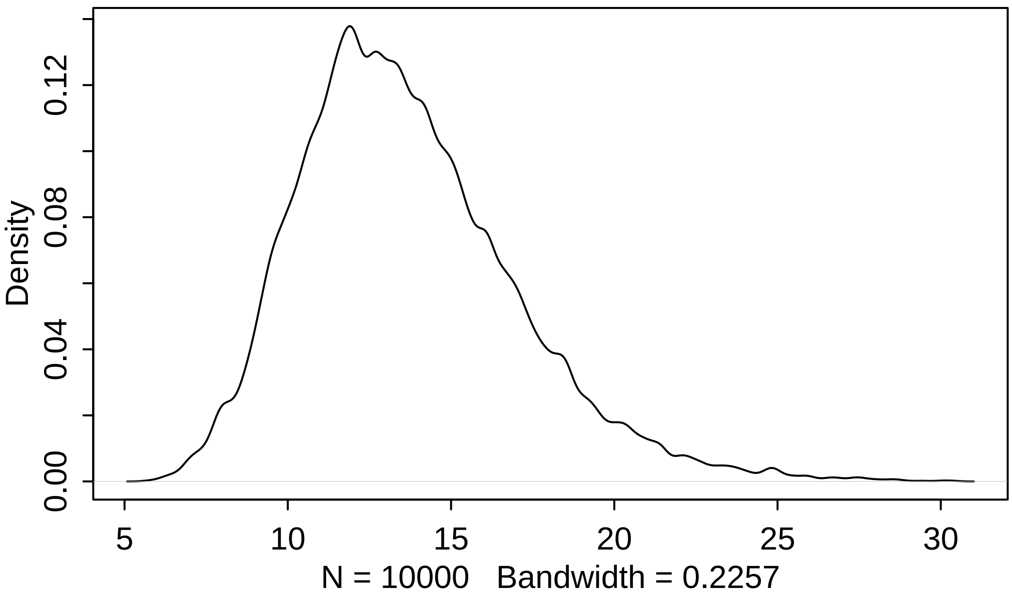
The distribution of the estimated growth by occasion can be extracted as follows

```r
posterior <- extract.samples(mex4,n = 1e4)

sim_growth <- rlnorm(1e4,posterior$mu,posterior$sigma)

dens(sim_growth)
```



The total growth would be the result of the accumulated growth over 8 occasions. So, I produce a matrix with growths sampled from the posterior and then calculate the total growth (sum of all elements in a column)

```r
simmatrix <- matrix(rlnorm(8*1e4,posterior$mu,posterior$sigma),nrow = 8)

sim_total_growth <- apply(simmatrix,2,sum)

dens(sim_total_growth)
```


Density
0.00
0.04
0.08
0.12
5
10
15
20
25
30
N = 10000   Bandwidth = 0.2257