Homework-2A

Sam Bashevkin

2022-06-10

```
library(rethinking)
```

4E1. In the model definition below, which line is the likelihood?

The first line

4E2. In the model definition just above, how many parameters are in the posterior distribution?

2: μ and σ

4E4. In the model definition below, which line is the linear model?

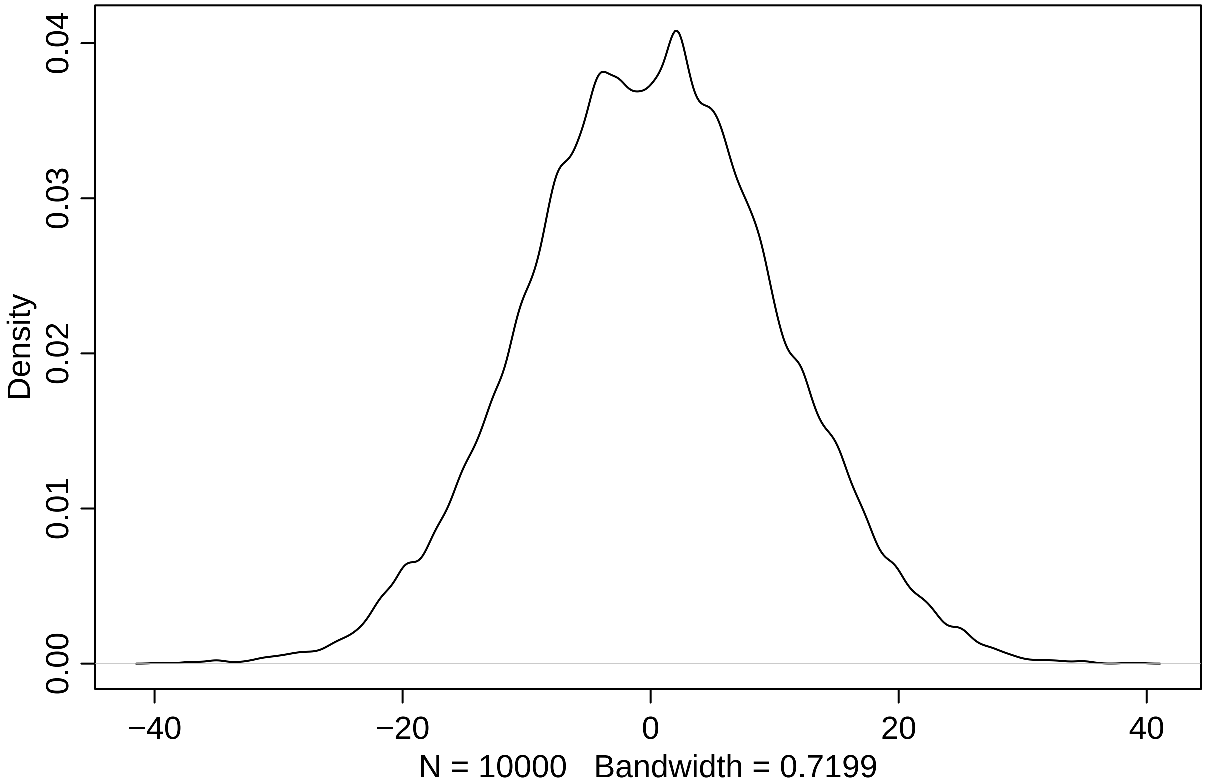
The second line

4E5. In the model definition just above, how many parameters are in the posterior distribution?

3: σ, α, and β

4M1. For the model definition below, simulate observed y values from the prior (not the posterior).

```
sample_mu <- rnorm( 1e4 , 0 , 10 )
sample_sigma <- rexp( 1e4 , 1)
prior_h <- rnorm( 1e4 , sample_mu , sample_sigma )
dens( prior_h )
```



4M2. Translate the model just above into a quap formula.

```
flist<-alist(y ~ dnorm(mu, sigma),
           mu ~ dnorm(0, 10),
           sigma ~ dexp(1))
```

4M3. Translate the quap model formula below into a mathematical model definition.

$$
\begin{aligned}
y_i &\sim Normal(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta x_i \\
\alpha &\sim Normal(0, 10) \\
\beta &\sim Normal(0, 1) \\
\sigma &\sim Exponential(1)
\end{aligned}
$$

4M4. A sample of students is measured for height each year for 3 years. After the third year, you want

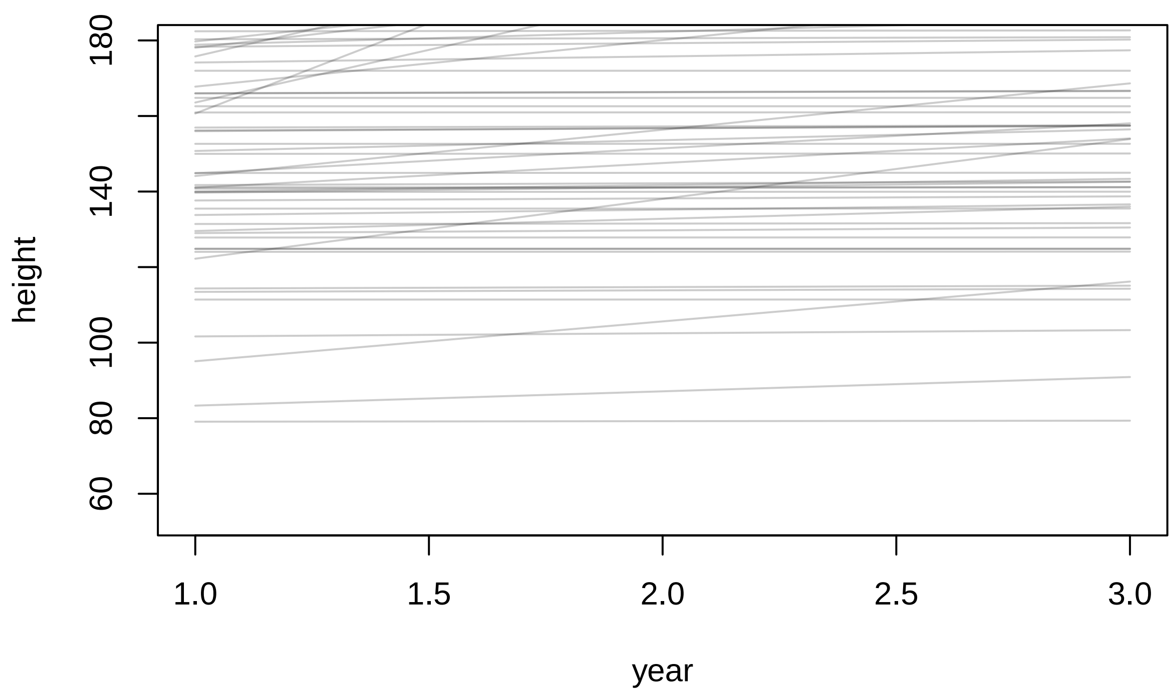
to fit a linear regression predicting height using year as a predictor. Write down the mathematical model definition for this regression, using any variable names and priors you choose. Be prepared to defend your choice of priors.

$$\begin{aligned}
height_i &\sim Normal(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta year_i \\
\alpha &\sim Normal(167, 40) \\
\beta &\sim Log - Normal(0, 5) \\
\sigma &\sim Exponential(1)
\end{aligned}$$

I chose the prior for α as a wide interval around my height. The prior for β assumes that these are minors <18 years old. If the population is college students, which could include older non-traditional students, the β prior should encompass positive and negative values.

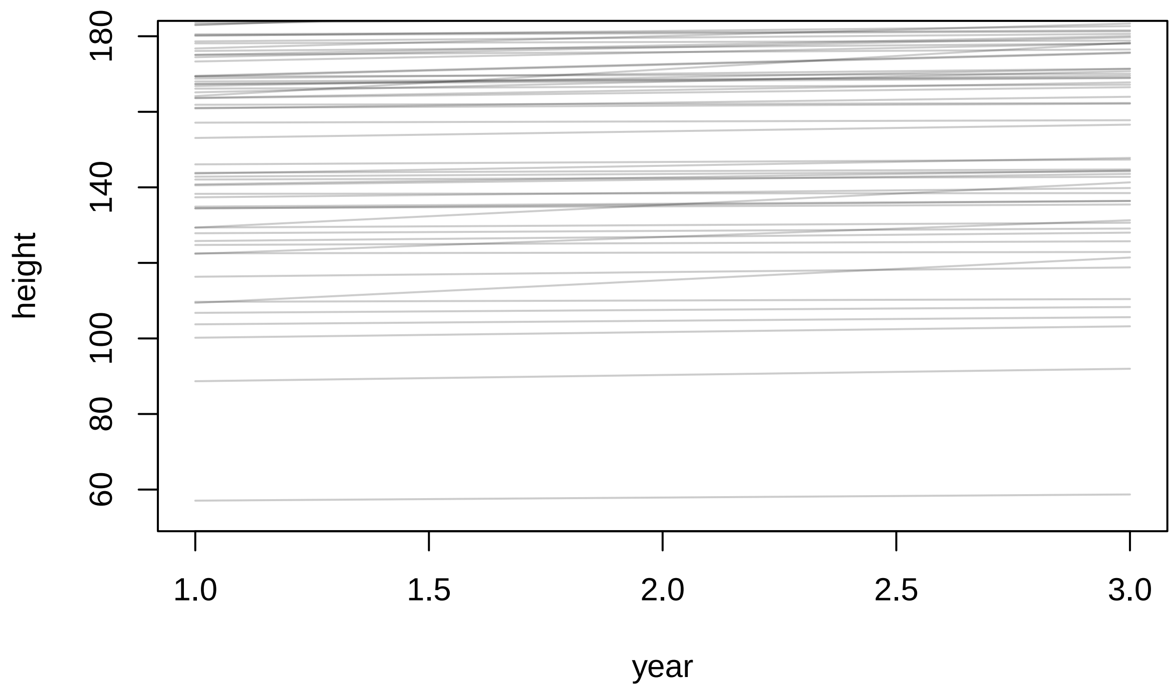
Now let's simulate from the prior to test it:

```
data("Howell1")
N<-100
a <- rnorm( N , 167 , 40 )
b <- rlnorm( N , 0 , 5)

plot( NULL , ylim=range(Howell1$height) , xlim=c(1,3) ,
xlab="year" , ylab="height" )
for ( i in 1:N ) curve( a[i] + b[i]*x ,
from=1 , to=3 , add=TRUE ,
col=col.alpha("black",0.2) )
```



Clearly some of these lines are way too steep. Let's try with $\beta \sim Log - Normal(0, 1)$

```r
data("Howell1")
N<-100
a <- rnorm( N , 167 , 40 )
b <- rlnorm( N , 0 , 1)

plot( NULL , ylim=range(Howell1$height) , xlim=c(1,3) ,
xlab="year" , ylab="height" )
for ( i in 1:N ) curve( a[i] + b[i]*x ,
from=1 , to=3 , add=TRUE ,
col=col.alpha("black",0.2) )
```



4M5. Now suppose I remind you that every student got taller each year. Does this information lead

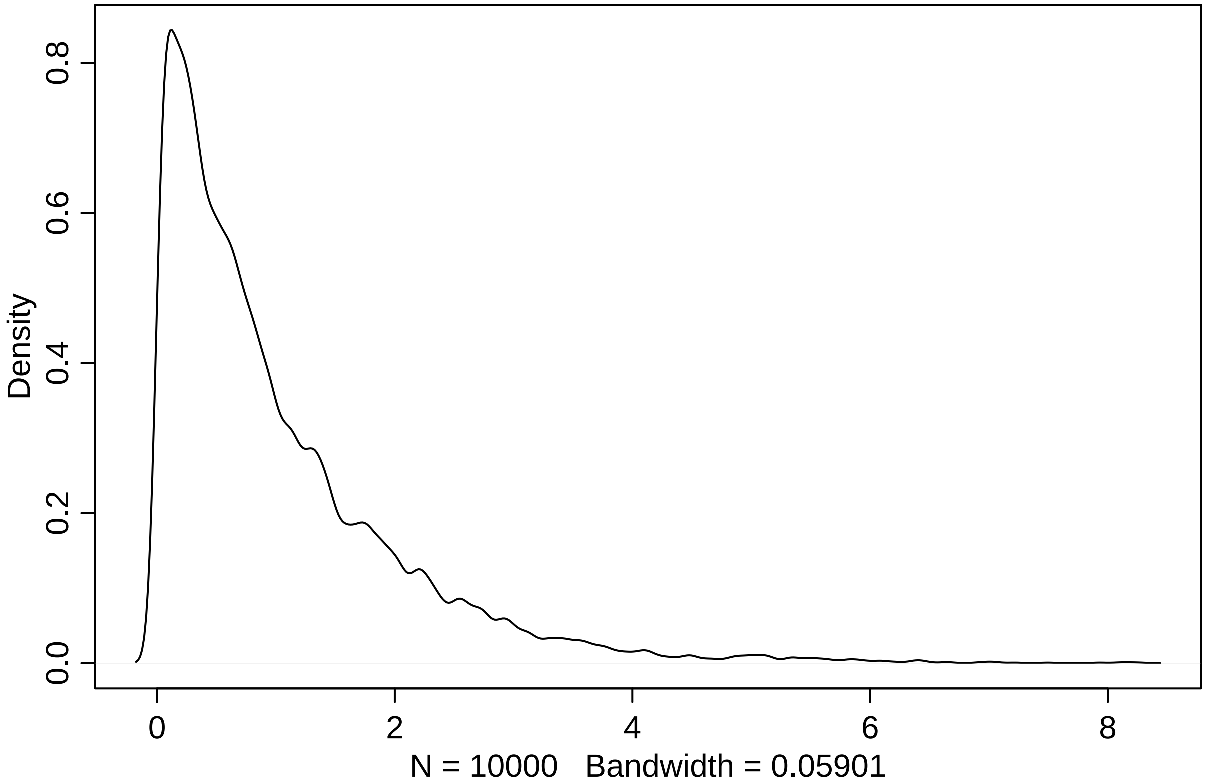
you to change your choice of priors? How?

Oops no I already did that

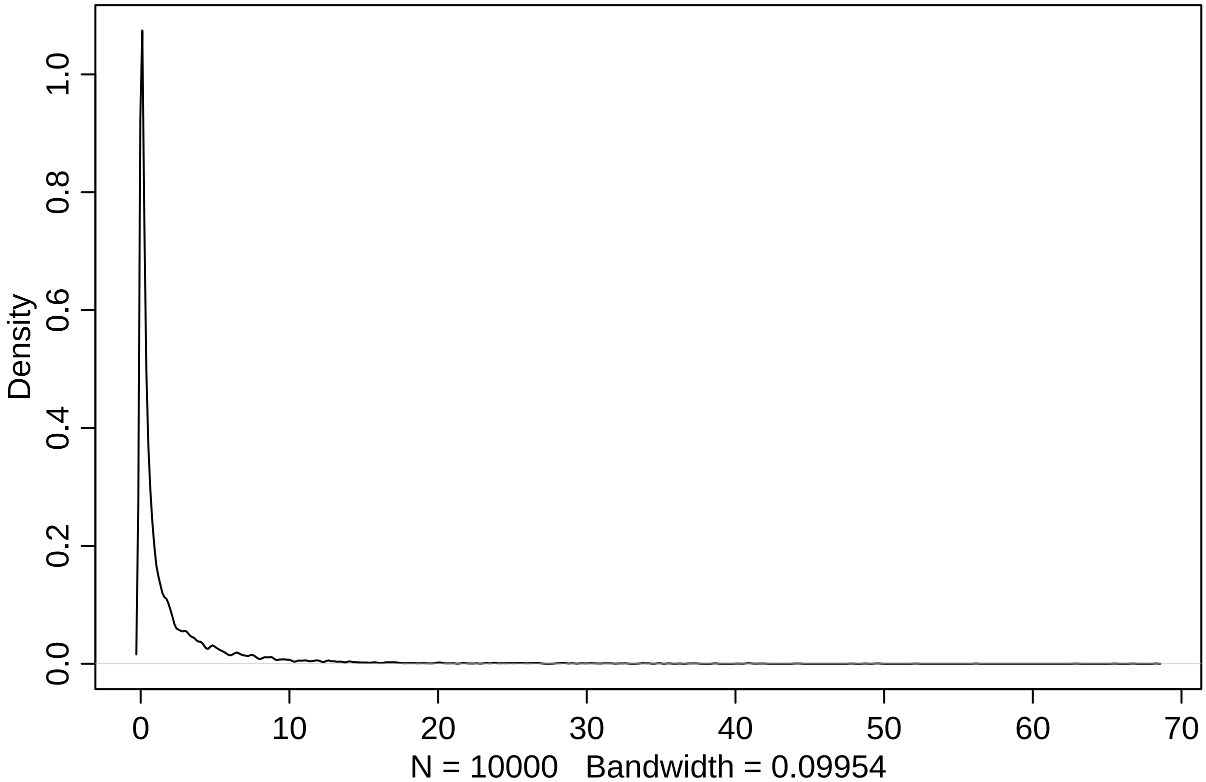
4M6. Now suppose I tell you that the variance among heights for students of the same age is never more than 64cm. How does this lead you to revise your priors?

What does the prior for σ look like now?

```r
sample_sigma <- rexp( 1e4 , 1)
dens(sample_sigma)
```




```
# On the variance scale

dens(sample_sigma^2)
```



I don't think I need to revise the prior because values at 64 and higher are already very improbable.