

Statistical Rethinking: Chapter 4 - Linear Models

Chris Hayduk

January 3, 2019

1 Easy

Problem 4E1.

In the model definition below, which line is the likelihood?

1. $y_i \sim \text{Normal}(\mu, \sigma)$
2. $\mu \sim \text{Normal}(0, 10)$
3. $\sigma \sim \text{Uniform}(0, 10)$

Line 1 represents the likelihood.

Problem 4E2.

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

There are two parameters: μ and σ .

Problem 4E3.

Using the model definition above, write down the appropriate form of Bayes' theorem that includes the proper likelihood and priors.

$$\Pr(\mu, \sigma | y) = \frac{\prod_i \text{Normal}(y_i | \mu, \sigma) \text{Normal}(\mu | 0, 10) \text{Uniform}(\sigma | 0, 10)}{\int \int \prod_i \text{Normal}(y_i | \mu, \sigma) \text{Normal}(\mu | 0, 10) \text{Uniform}(\sigma | 0, 10) d\mu d\sigma}$$

Problem 4E4.

In the model definition below, which line is the linear model?

1. $y_i \sim \text{Normal}(\mu, \sigma)$
2. $\mu_i = \alpha + \beta x_i$
3. $\alpha \sim \text{Normal}(0, 10)$
4. $\beta \sim \text{Normal}(0, 10)$
5. $\sigma \sim \text{Uniform}(0, 10)$

Line 2 represents the linear model.

Problem 4E5.

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

There are three parameters: α , β , and σ .

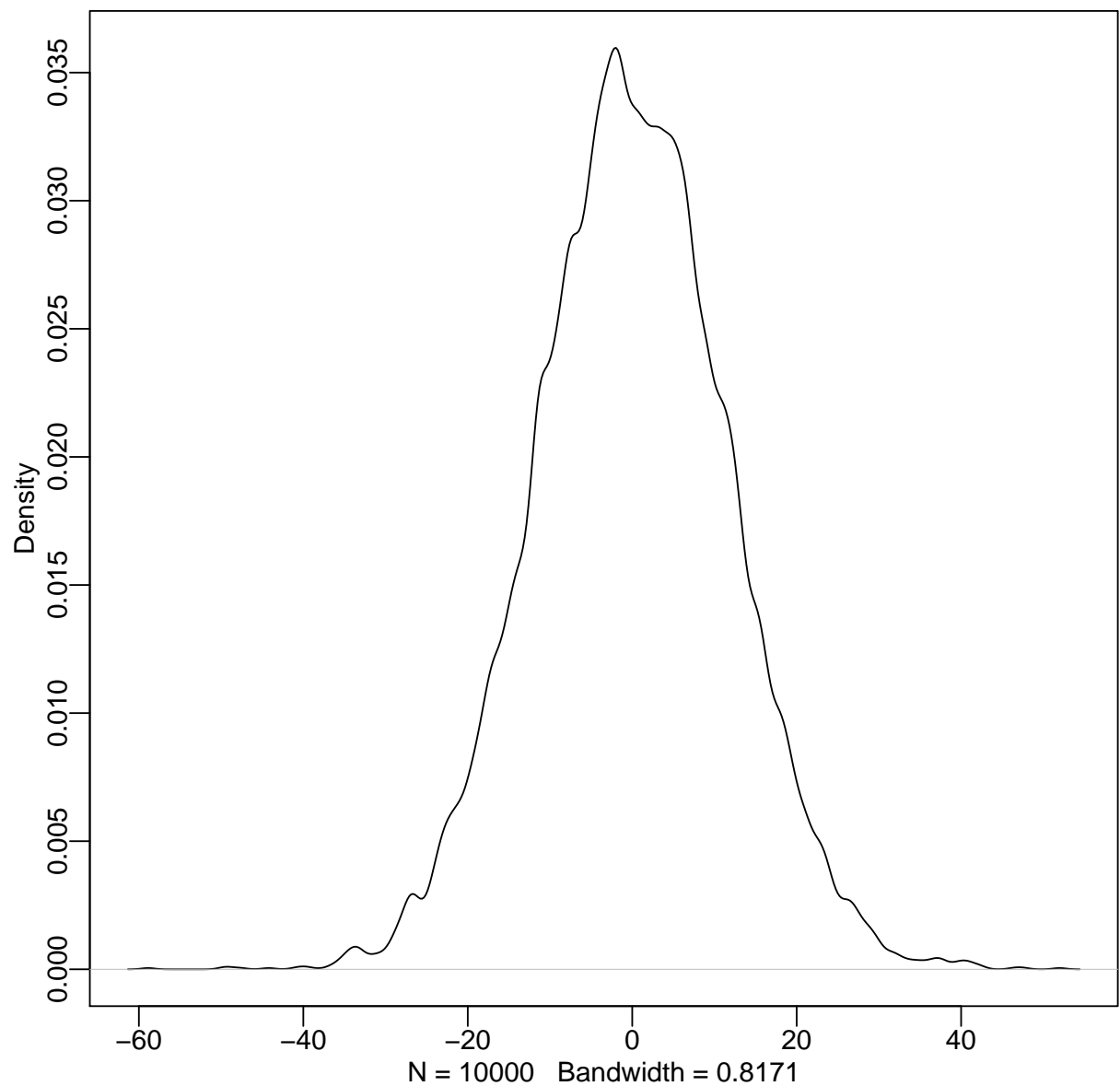
2 Medium

Problem 4M1.

For the model definition below, simulate the observed heights from the prior (not the posterior).

$$\begin{aligned}y_i &\sim \text{Normal}(\mu, \sigma) \\ \mu &\sim \text{Normal}(0, 10) \\ \sigma &\sim \text{Uniform}(0, 10)\end{aligned}$$

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



Problem 4M2.

Translate the model just above into a *map* formula.

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

Problem 4M3.

Translate the *map* model formula below into a mathematical model definition.

```
flist <- alist(
  y ~ dnorm(mu, sigma),
  mu <- a + b*x,
  a ~ dnorm(0, 50),
  b ~ dunif(0, 10),
  sigma ~ dunif(0, 50)
)
```

Model:

$$\begin{aligned}
 y_i &\sim \text{Normal}(\mu_i, \sigma) \\
 \mu_i &= \alpha + \beta x_i \\
 \alpha &\sim \text{Normal}(0, 50) \\
 \beta &\sim \text{Uniform}(0, 10) \\
 \sigma &\sim \text{Uniform}(0, 50)
 \end{aligned}$$

Problem 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 for this regression, using any variable names and priors you choose. Be prepared to defend your choice of priors.

Model:

$$\begin{aligned}
 y_i &\sim \text{Normal}(\mu_i, \sigma) \\
 \mu_i &= \alpha + \beta x_i \\
 \alpha &\sim \text{Normal}(152, 25) \\
 \beta &\sim \text{Normal}(6, 3) \\
 \sigma &\sim \text{Uniform}(0, 50)
 \end{aligned}$$

The priors for α represent an average of about 5 feet (152 cm) and a standard deviation of about 10 inches (25 cm). The priors for β represent an average increase of about 2.4 inches per year (6 cm - chosen using average growth rate for children) and a standard deviation of about 1.19 inches (3 cm). The prior for σ is a uniform prior between 0 and 50.

Problem 4M5.

Now suppose I tell you that the average height in the first year was 120 cm and that every student got taller each year and every student got taller each year. Does this information lead you to change your choice of prior?

With this no information, we should adjust α . We now have:

$$\begin{aligned}
 y_i &\sim \text{Normal}(\mu_i, \sigma) \\
 \mu_i &= \alpha + \beta x_i \\
 \alpha &\sim \text{Normal}(120, 25) \\
 \beta &\sim \text{Normal}(6, 3) \\
 \sigma &\sim \text{Uniform}(0, 50)
 \end{aligned}$$

The priors for α have been adjusted to account for the new information we have about the average height in the first year. We already chose β as a positive number with a relatively small standard deviation, so the information about students growing taller each year does not effect our choice of prior.

Problem 4M6.

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

σ is just the square root of the variance. Thus, we would like β 's standard deviation to be less than $\sqrt{64} = 8$. Since we already chose a standard deviation of 3 cm, we do not need to revise this choice.

3 Hard

Problem 4H1.

The weights listed below were recorded in the !Kung census, but heights were not recorded for these individuals. Provide predicted heights and 89% intervals (either HPDI or PI) for each of these individuals. That is, fill in the table below, using model-based predictions.

```
data(Howell1)

d1 <- Howell1

m1 <- map(
  alist(
    height ~ dnorm(a + b*weight, sigma),
    a ~ dnorm(178, 100),
    b ~ dnorm(0, 10),
    sigma ~ dunif(0, 50)
  ), data = d1)

weights <- c(46.95, 43.72, 64.78, 32.59, 54.63)

sim.height <- sim(m1, data=list(weight=weights), n = 1e4, silent = TRUE)

## [ 1000 / 10000 ]
[ 2000 / 10000 ]
[ 3000 / 10000 ]
[ 4000 / 10000 ]
[ 5000 / 10000 ]
[ 6000 / 10000 ]
[ 7000 / 10000 ]
[ 8000 / 10000 ]
```

```

[ 9000 / 10000 ]
[ 10000 / 10000 ]

mean.height <- apply(sim.height, 2, mean)

print(mean.height)

## [1] 158.2247 152.4695 189.6582 132.7213 171.8540

height.PI <- apply(sim.height, 2, PI, prob=0.89)

print(height.PI)

##           [,1]      [,2]      [,3]      [,4]      [,5]
## 5%  143.4416 137.6086 174.4575 117.6613 156.9577
## 94% 173.2767 167.3117 205.1101 147.7628 186.9470

```

Individual	weight	expected height	89% interval
1	46.95	158.43	(143.67, 173.42)
2	43.72	152.56	(137.83, 167.69)
3	64.78	189.54	(174.28, 204.93)
4	32.59	133.09	(118.25, 148.41)
5	54.63	171.67	(156.48, 186.94)

Problem 4H2.

Select out all the rows in the Howell1 data with ages below 18 years of age. If you do it right, you should end up with a new data frame with 192 rows in it.

- Fit a linear regression to these data, using *map*. Present and interpret the estimates. For every 10 units of increase in weight, how much taller does the model predict a child gets?
- Plot the raw data, with height on the vertical axis and weight on the horizontal axis. Superimpose the MAP regression line and 89% HPDI for the mean. Also superimpose the 89% HPDI for predicted heights.
- What aspects of the model fit concern you? Describe the kinds of assumptions you would change, if any, to improve the model. You don't have to write any new code. Just explain what the model appears to be doing a bad job of, and what you hypothesize would be a better model.

```

(a) d2 <- Howell1
d2 <- d2[d2$age < 18, ]

str(d2)

## 'data.frame': 192 obs. of 4 variables:
## $ height: num 121.9 105.4 86.4 129.5 109.2 ...
## $ weight: num 19.6 13.9 10.5 23.6 16 ...
## $ age : num 12 8 6.5 13 7 17 16 11 17 8 ...
## $ male : int 1 0 0 1 0 1 0 1 0 1 ...

m2 <- map(
  alist(
    height ~ dnorm(a + b*weight, sigma),
    a ~ dnorm(178, 100),
    b ~ dnorm(0, 10),
    sigma ~ dunif(0, 50)
  ), data = d2)

print(m2)

##
## Maximum a posteriori (MAP) model fit
##
## Formula:
## height ~ dnorm(a + b * weight, sigma)
## a ~ dnorm(178, 100)
## b ~ dnorm(0, 10)
## sigma ~ dunif(0, 50)
##
## MAP values:
##      a      b      sigma
## 58.256579 2.718923 8.437103
##
## Log-likelihood: -681.9

```

For every 10 units of increase in weight, height increases by $10*b \approx 27$.

```

(b) #Define sequence of weights to compute predictions for
#These values will be on the horizontal axis
weight.seq <- seq(from = 0, to = 50, by = 1)

#Use link to compute mu for each sample from the posterior

```

```

#and for each weight in weight.seq
post <- extract.samples(m2)
mu.link <- function(weight) post$a + post$b*weight
mu <- sapply(weight.seq, mu.link)

#Summarize the distribution of mu
mu.mean <- apply(mu, 2, mean)
mu.HPDI <- apply(mu, 2, HPDI, prob = 0.89)

#Simulate heights from the posterior
sim.height <- sim(m2, data=list(weight=weight.seq))

## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]

#Summarize the simulated heights
height.PI <- apply(sim.height, 2, PI, prob=0.89)

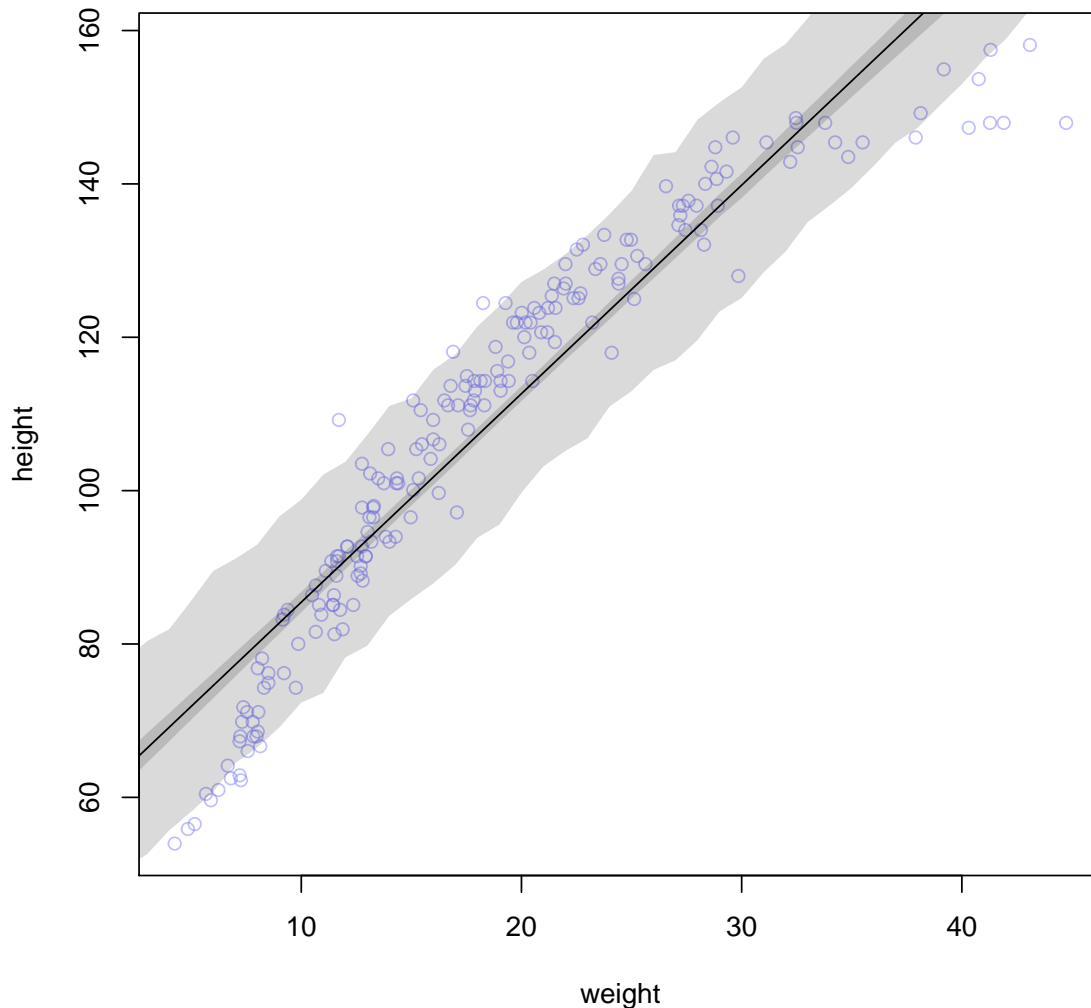
#Plot raw data
plot(height ~ weight, d2, col=col.alpha(rangi2, 0.5))

#Draw MAP line
lines(weight.seq, mu.mean)

#Draw HPDI region for line
shade(mu.HPDI, weight.seq)

#Draw PI region for simulated heights
shade(height.PI, weight.seq)

```

- (c) The rate of height increase in relation to weight does not appear to be constant. Thus, the data may be better modeled by a quadratic or cubic function.

Problem 4H3.

Suppose a colleague of yours, who works on allometry, glances at the practice problems just above. Your colleague exclaims, "That's silly. Everyone knows that it's only the *logarithm* of body weight that scales with height!" Let's take your colleague's advice and see what happens.

- (a) Model the relationship between height (cm) and the natural logarithm of weight (log-kg). Use the entire Howell1 data frame, all 544 rows, adults and non-adults. Fit this model using quadratic approximation:

$$h_i \sim \text{Normal}(\mu_i, \sigma)$$

$$\begin{aligned}\mu_i &= \alpha + \beta \log(x_i) \\ \alpha &\sim \text{Normal}(178, 100) \\ \beta &\sim \text{Normal}(0, 100) \\ \sigma &\sim \text{Uniform}(0, 50)\end{aligned}$$

where h_i is the height of individual i and w_i is the weight (in kg) of individual i . The function for computing a natural log in R is just `log`. Can you interpret the resulting estimates?

(b) Begin with this plot:

```
plot(height ~ weight, data = Howell1, col = col.alpha(rangi2, 0.4))
```

Then use samples from the quadratic approximate posterior of the model in (a) to superimpose on the plot: (1) the predicted mean height as a function of weight, (2) the 97% HPDI for the mean, (3) the 97% HPDI for predicted heights.

```
(a) Howell1$weight.log <- log(Howell1$weight)

m3 <- map(
  alist(
    height ~ dnorm(a + b*weight.log, sigma),
    a ~ dnorm(178, 100),
    b ~ dnorm(0, 100),
    sigma ~ dunif(0, 50)
  ), data = Howell1)

print(m3)

##
## Maximum a posteriori (MAP) model fit
##
## Formula:
## height ~ dnorm(a + b * weight.log, sigma)
## a ~ dnorm(178, 100)
## b ~ dnorm(0, 100)
## sigma ~ dunif(0, 50)
##
## MAP values:
##           a           b       sigma
## -23.784397  47.075377   5.134705
##
## Log-likelihood: -1661.9
```

For every one unit increase in log-kg, height increases by about 47 cm. The intercept a is negative and thus is uninterpretable. This arises because $\log(0)$ is undefined and $\log(x) \rightarrow -\infty$ as $x \rightarrow 0$.

(b) `plot(height ~ weight, data = Howell1, col = col.alpha(rangi2,0.4))`

```
#Define sequence of weights to compute predictions for
#These values will be on the horizontal axis
weight.seq <- seq(from = 1, to = 70, by = 1)

#Use link to compute mu for each sample from the posterior
#and for each weight in weight.seq
post <- extract.samples(m3)
mu.link <- function(weight) post$a + post$b*log(weight)
mu <- sapply(weight.seq, mu.link)

#Summarize the distribution of mu
mu.mean <- apply(mu, 2, mean)
mu.HPDI <- apply(mu, 2, HPDI, prob = 0.97)

#Simulate heights from the posterior
sim.height <- sim(m3, data=list(weight.log=log(weight.seq)))

## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]

#Summarize the simulated heights
height.PI <- apply(sim.height, 2, PI, prob=0.97)

#Draw MAP line
lines(weight.seq, mu.mean)

#Draw HPDI region for line
shade(mu.HPDI, weight.seq)
```

```
#Draw PI region for simulated heights  
shade(height.PI, weight.seq)
```

