Chapter 4 Questions

Melissa Van Bussel

June 15, 2018

Chapter 4 points: 2 + 10 + 4 = 16 points total.

Easy (2 points total)

4E1

Line 1

4E2

2 parameters (μ and σ)

4E3

$Pr(\mu, \sigma|h) = \frac{\Pi_i \text{Normal}(h_i|\mu,\sigma)\text{Normal}(\mu|0,10)\text{Normal}(\sigma|0,10)}{\int\int \text{Normal}(h_i|\mu,\sigma)\text{Normal}(\mu|0,10)\text{Normal}(\sigma|0,10)d\mu d\sigma}$

4E4

Line 2

4E5

Three parameters (α, β, and σ).

Medium (10 points total)

4M1 (2 points)

```
library(rethinking)
```

```
## Loading required package: rstan

## Warning: package 'rstan' was built under R version 3.3.3

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.3.3

## Loading required package: StanHeaders

## Warning: package 'StanHeaders' was built under R version 3.3.3
```

```
## rstan (Version 2.17.3, GitRev: 2e1f913d3ca3)

## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)

## Loading required package: parallel

## rethinking (Version 1.59)
```

```
mu_samples <- rnorm(10000, 0, 10)
sigma_samples <- runif(10000, 0, 10)
simulated_heights <- rnorm(10000, mean = mu_samples, sd = sigma_samples)
```

4M2 (2 points)

```
flist <- alist(
  height ~ dnorm(mu, sigma),
  mu ~ dnorm(0, 10),
  sigma ~ dunif(0, 10)
)
# map4M2 <- map(flist, some_data_goes_here)
```

4M3 (2 points)

$y_i \sim$ Normal(μ, σ) $\mu = \alpha + \beta x_i$ $\alpha \sim$ Normal(0, 50) $\beta \sim$ Uniform(0, 10) $\sigma \sim$ Uniform(0, 50)

4M4 (2 points)

We don't know the age of the students, so it's hard to guess their average height, so we expect a high value of sigma at first, and our guess at the average height is pretty much a shot in the dark (since a kindergartener and a 12th grader are both "students", yet their average heights are vastly different). For the intercept, I'll choose 3 feet (90cm), and I'll guess that students grow around 6cm per year (as per a Google search). Sigma will be pretty large here; let's say

$y_i \sim$ Normal(μ, σ) $\mu = \alpha + \beta x_i$ $\alpha \sim$ Normal(90, 20) $\beta \sim$ Normal(6, 2) $\sigma \sim$ Uniform(0, 50)

4M5 (2 points)

I would change my intercept to be closer to 120. The intercept is at year "0", so I'd change it to 120-6 (since we're saying students grow approx 6cm/year). I would also add more precision (by making the variance on the alpha smaller). Change beta from being normal to being uniform since students can't get any smaller each year so maybe Uniform(0, 6)

4M6 (2 points)

Change sigma to be Uniform(0, 8)

HArd (4 points total)

4H1 (1 point)

```r
data(Howell1)
d <- Howell1
d2 <- d[d$age >= 18,]
weight_centred <- (d2$weight - mean(d2$weight) ) / sd(d2$weight)
flist <- alist(
  height ~ dnorm(mu, sigma),
  mu <- alpha + beta * weight_centred,
  alpha ~ dnorm(50, 50), # I'll guess mean weight is 50kg, plus or minus..
  beta ~ dnorm(0, 10), # 10 kg increase per cm seems extreme but..
  sigma ~ dunif(0, 8) # Previous question told us 64 is a good value for maximum variance
  )
map4H1 <- map(flist, d2)

# Figure out associated heights from given weights
given_weights <- c(46.95, 43.72, 64.78, 32.59, 54.63)
given_weights_centred <- data.frame((given_weights - mean(given_weights)) / sd(given_weights))
#simulated_heights <- sim(map4H1, data = given_weights_centred)
#shmean <- apply(simulated_heights, 2, mean)
#sh_HPDI <- apply(simulated_heights, 2, HPDI, prob = 0.89)
#sh_HPDI
```

4H2 (1 point)

```r
# part a
data(Howell1)
d <- Howell1
d2 <- d[d$age < 18,]
flist <- alist(
  height ~ dnorm(mu, sigma),
  mu <- alpha + beta *weight,
  alpha ~ dnorm(0, 100),
  beta ~ dnorm(0, 10),
  sigma ~ dunif(0, 10)
  )
map4H2 <- map(flist, d2)

precis(map4H2)
```

```
##        Mean StdDev  5.5% 94.5%
## alpha 58.22   1.40 55.99 60.45
## beta   2.72   0.07  2.61  2.83
## sigma  8.44   0.43  7.75  9.13
```

```r
coef(map4H2)['b']
```

```
## <NA>
##   NA
```

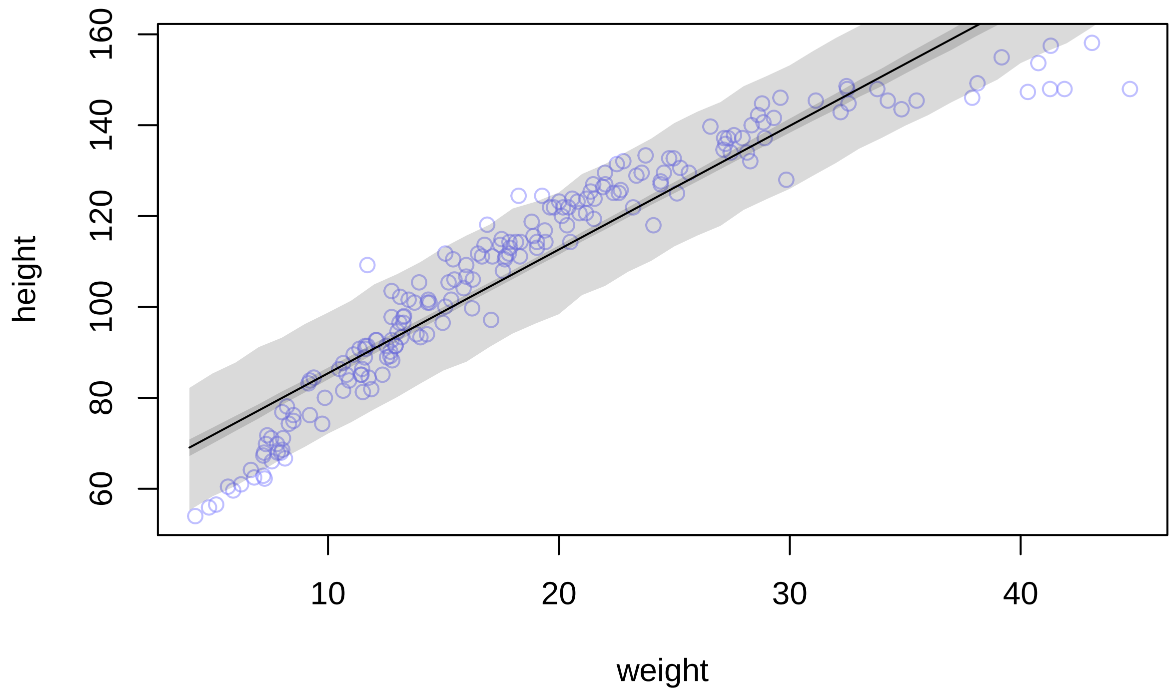
```r
post <- extract.samples(map4H2)

# part b
weight2 <- seq(4, 45, 1)
d2 <- d2[order(d2$weight),]
mu <- link(map4H2, data = data.frame(weight = weight2))
```

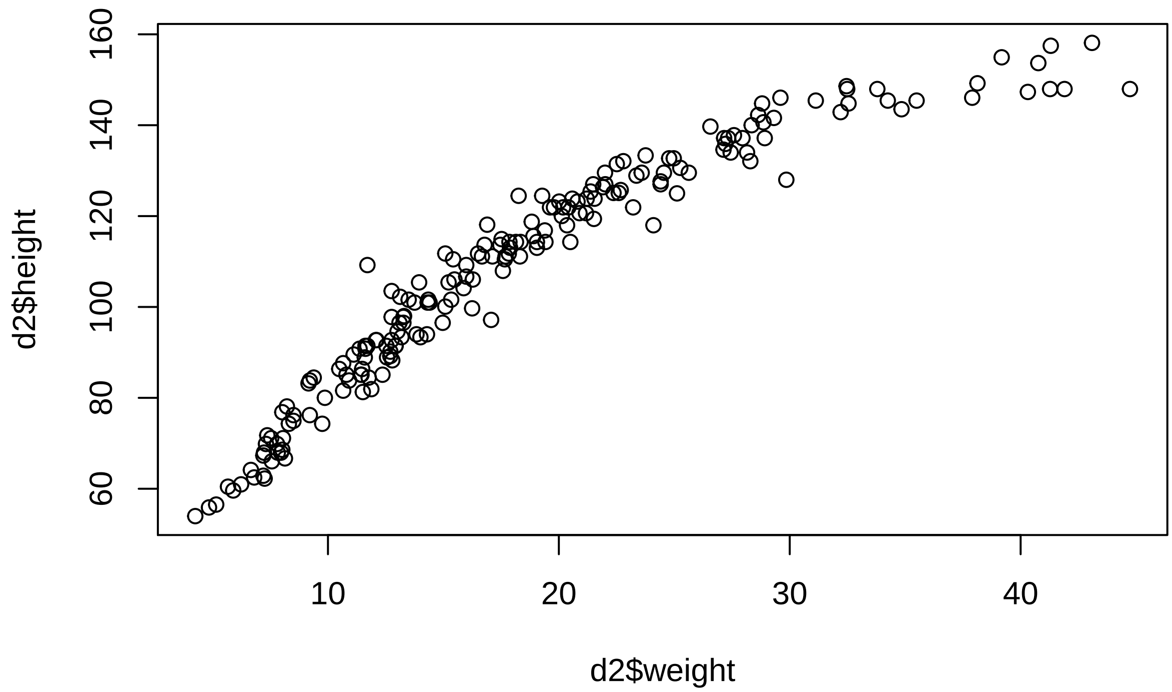
```
## [ 100 / 1000 ]
[ 200 / 1000 ]
[ 300 / 1000 ]
[ 400 / 1000 ]
[ 500 / 1000 ]
[ 600 / 1000 ]
[ 700 / 1000 ]
[ 800 / 1000 ]
[ 900 / 1000 ]
[ 1000 / 1000 ]
```

```r
mu.mean <- apply(mu, 2, mean)
mu.HPDI <- apply(mu, 2, HPDI, prob = 0.89)
sim.height <- sapply(weight2, function(weight) {
  rnorm(
    n = nrow(post),
    mean = post$alpha + post$beta*weight,
    sd = post$sigma
  )
})

height.HPDI <- apply(sim.height, 2, HPDI, prob = 0.89)
plot(height ~ weight, data = d2, col = col.alpha(rangi2, 0.5))
lines(weight2, mu.mean)
shade(mu.HPDI, weight2)
shade(height.HPDI, weight2)
```




```
plot(d2$weight, d2$height)
```




```
#lines(map4H2)
```

4H3 (2 points)

```
df <- Howell1
mod <- map(
  alist(
    height ~ dnorm(mu, sigma),
    mu <- a + b*log(weight),
    a ~ dnorm(178, 100),
    b ~ dnorm(0, 100),
    sigma ~ dunif(0, 50)
  ), data = df
)
precis(mod, corr = TRUE) # they're negative :( interpretation is kind of silly
```

```
##          Mean StdDev   5.5%  94.5%     a     b sigma
## a      -23.78   1.34 -25.92 -21.65  1.00 -0.99     0
## b       47.08   0.38  46.46  47.69 -0.99  1.00     0
## sigma    5.13   0.16   4.89   5.38  0.00  0.00     1
```

```
post <- extract.samples(mod, n = 1e4)
mu.link <- function(weight ) {
  post$a + post$b * log(weight)
}
```

```r
weight.seq <- seq(0, 70, 1)

mu <- sapply(weight.seq, mu.link)
mu.mean <- apply(mu, 2, mean)
mu.HPDI <- apply(mu, 2, HPDI, prob = 0.89)
sim.height <- sim(mod, 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 ]
```

```r
height.PI <- apply(sim.height, 2, PI, prob = 0.97)

plot(height ~ weight, df,  ylim = c(0, 180), col = col.alpha(rangi2, 0.5))
lines(weight.seq, mu.mean)
shade(mu.HPDI, weight.seq)
shade(height.PI, weight.seq)
```