Chapter 4 Exercises

Hard

4H1

```
library(rethinking)
library(greta)

data(Howell1)
d <- Howell1

# data
height <- d$height
weight <- as_data(d$weight - mean(d$weight))

# variables & priors
alpha <- normal(178, 20)
beta <- normal(0, 10)
sigma <- uniform(0, 50)

# likelihood
mu <- alpha + beta * weight

# observation model
distribution(height) <- normal(mu, sigma)

m4h1 <- model(alpha, beta, sigma)

# find the MAP estimate using
opt_m4h1 <- opt(m4h1, optimiser = bfgs(), hessian = TRUE)

x_new <- c(46.95, 43.72, 64.78, 32.59, 54.63)
mu_hat <- opt_m4h1$par$alpha + opt_m4h1$par$beta * x_new
y_hat_sim <- sapply(mu_hat, function(mean) rnorm(1000, mean, opt_m4h1$par$sigma))
(y_hat_mean <- apply(y_hat_sim, 2, mean))
```

```
## [1] 221.5117 214.9113 252.5506 196.0385 234.7882
```

```
(y_hat_89ci <- apply(y_hat_sim, 2, quantile, probs = c(0.055, 0.945)))
```

```
##           [,1]     [,2]     [,3]     [,4]     [,5]
## 5.5%  206.4799 200.5049 236.7198 181.0055 219.2905
## 94.5% 236.7069 229.3812 267.3859 212.0404 250.0745
```

4H2

```r
d2 <- subset(d, age < 18)
nrow(d2)
```

```
## [1] 192
```

(a)

```r
height <- d2$height
weight <- as_data(d2$weight)

alpha <- normal(178, 20)
beta <- normal(0, 10)
sigma <- uniform(0, 50)

mu <- alpha + beta * weight

distribution(height) <- normal(mu, sigma)

m4h2a <- model(alpha, beta, sigma)

opt_m4h2a <- opt(m4h2a, optimiser = bfgs(), hessian = TRUE)

opt_m4h2a$par
```

```
## $alpha
## [1] 58.81755
##
## $beta
## [1] 2.694244
##
## $sigma
## [1] 8.458539
```

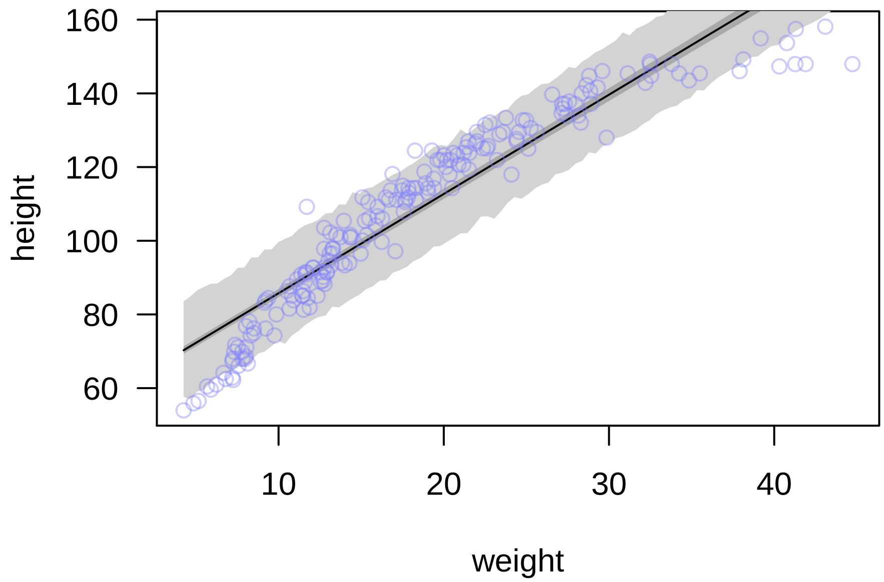
(b)

```r
# calculate SE of params
# following https://github.com/greta-dev/greta/issues/226
se <-
  lapply(opt_m4h2a$hessian, function(h) {
    sqrt(diag(solve(h)))
  })

# interval of mean
x_new <- seq(min(d2$weight), max(d2$weight), length.out = 100)
mu_sim <-
  rnorm(1000, opt_m4h2a$par$alpha, se$alpha) +
  outer(rnorm(1000, opt_m4h2a$par$beta, se$beta), (x_new))
mu_ci <- apply(mu_sim, 2, quantile, probs = c(0.055, 0.945))

# prediction interval
mu_hat <- opt_m4h2a$par$alpha + opt_m4h2a$par$beta * (x_new)
y_hat_sim <- sapply(mu_hat, function(mean) rnorm(1000, mean, opt_m4h2a$par$sigma))
y_hat_mean <- apply(y_hat_sim, 2, mean)
y_hat_89ci <- apply(y_hat_sim, 2, quantile, probs = c(0.055, 0.945))
```

```r
with(d2, plot(weight, height, type = "n", las = 1))
polygon(c(x_new, rev(x_new)), c(y_hat_89ci[1,], rev(y_hat_89ci[2,])),
        col = "lightgrey", border = NA)
polygon(c(x_new, rev(x_new)), c(mu_ci[1,], rev(mu_ci[2,])),
        col = "darkgrey", border = NA)
curve(opt_m4h2a$par$alpha + opt_m4h2a$par$beta * x,
      from = min(d2$weight),
      to = max(d2$weight),
      add = TRUE)
with(d2, points(weight, height, col = col.alpha(rangi2,0.4)))
```



4H3

###(a)

```r
# data
height <- d$height
weight <- as_data(d$weight)

# variables & priors
alpha <- normal(178, 20)
beta <- lognormal(0, 1)
sigma <- uniform(0, 50)

# likelihood
mu <- alpha + beta * log(weight)

# observation model
distribution(height) <- normal(mu, sigma)

m4h3 <- model(alpha, beta, sigma)

# find the MAP estimate using
opt_m4h3 <- opt(m4h3, optimiser = bfgs(), hessian = TRUE)
opt_m4h3$par
```

```
## $alpha
## [1] -22.88353
##
## $beta
## [1] 46.82048
##
## $sigma
## [1] 5.141227
```

(b)

```r
# interval of mean
x_new <- seq(min(log(d$weight)), max(log(d$weight)), length.out = 100)
mu_sim <-
  rnorm(1000, opt_m4h3$par$alpha, se$alpha) +
  outer(rnorm(1000, opt_m4h3$par$beta, se$beta), x_new)
mu_ci <- apply(mu_sim, 2, quantile, probs = c(0.015, 0.985))

# prediction interval
mu_hat <- opt_m4h3$par$alpha + opt_m4h3$par$beta * x_new
y_hat_sim <- sapply(mu_hat, function(mean) rnorm(1000, mean, opt_m4h3$par$sigma))
y_hat_mean <- apply(y_hat_sim, 2, mean)
y_hat_97ci <- apply(y_hat_sim, 2, quantile, probs = c(0.015, 0.985))

plot( height ~ weight , data=Howell1, type = "n", las = 1)
polygon(c(exp(x_new), rev(exp(x_new))), c(y_hat_97ci[1,], rev(y_hat_97ci[2,])),
        col = "lightgrey", border = NA)
polygon(c(exp(x_new), rev(exp(x_new))), c(mu_ci[1,], rev(mu_ci[2,])),
        col = "darkgrey", border = NA)
with(d, points(weight, height, col = col.alpha(rangi2,0.4)))
curve(opt_m4h3$par$alpha + opt_m4h3$par$beta * log(x),
      from = min(d$weight),
      to = max(d$weight),
      add = TRUE)
```