SR chapter 4

Emilio Horner

2022-09-29

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──
## ✓ ggplot2 3.3.6      ✓ purrr   0.3.4
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.0      ✓ stringr 1.4.1
## ✓ readr   2.1.2      ✓ forcats 0.5.2
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
```

4e1. yi ~Normal(μ,σ)

4e2. Two, μ,σ

4e3. The variable yi has a normal distribution of the mean and standard deviation.

yi ~ Normal(μ, σ) μ ~ Normal(0, 10) σ ~ Exponential(1)

P(μ,σ | yi) ∝ Likelihood × Prior Probability

Pr(μ, σ|yi) = Normal(yi |μ, σ) Normal(μ|0, 10)Uniform (σ|0, 50)/Normal(yi |μ, σ)Normal(μ|0, 10)Uniform(σ|0, 50)dμdσ

4e4. μi = α + βxi is the linear model

4e5.

three parameters α , β, and x

4m1.

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

4m2.

yi ~Normal(μ,σ) μ ~Normal(0,10) σ ~Exponential(1)

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

4m3. flist <- alist(y ~ dnorm(mu , sigma), mu <- a + b*x, a ~ dnorm(0 , 10), b ~ dunif(0 , 1), sigma ~ dexp(1))

y ~Normal(μ,σ) μi = α + β(xi−¯x) α ~Normal(0,50) β ~Normal(0,10) σ ~Uniform(0,50)

Its on page 96 in the book. It doesn't copy super neatly.

4m4. the priors would be something like average height and average growth rate by child that age.

hi ~Normal(μi,σ) μi = α + β(xi-x) α ~Normal(56,10)
β ~Normal(2,1) σ ~Uniform(0,50)

I picked 56 as the average fifth grader height (in inches), and 10 inches as the standard deviation around that. for the second prior I picked 2 inches as the mean that the seconds grow every year and 1 inch as the standard deviation around that mean.

4m5.

Since we know that all the students must get taller I would assume that the students are pretty young. Most students that are in college have stopped growing. I might decrease my average height knowing their age since they're probably younger.

4m6. I would use this to limit the standard deviation numbers since now we know that the variance can't be more than 64cm.

4m7.

```
library(rethinking)
```

```
## Loading required package: rstan
```

```
## Loading required package: StanHeaders
```

```
## rstan (Version 2.21.5, 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)
```

```
##
## Attaching package: 'rstan'
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
## Loading required package: cmdstanr
```

```
## This is cmdstanr version 0.5.3
```

```
## - CmdStanR documentation and vignettes: mc-stan.org/cmdstanr
```

```
## - Use set_cmdstan_path() to set the path to CmdStan
```

```
## - Use install_cmdstan() to install CmdStan
```

```
## Loading required package: parallel
```

```
## rethinking (Version 2.23)
```

```
##
## Attaching package: 'rethinking'
```

```
## The following object is masked from 'package:rstan':
##
##     stan
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```
## The following object is masked from 'package:stats':
##
##     rstudent
```

```
data(Howell1);
d <- Howell1; d2 <- d[ d$age >= 18 , ]
# define the average weight, x-bar
xbar <- mean(d2$weight)
# fit model
```

This is the original model

```
quap(
alist(
height ~ dnorm( mu , sigma ) ,
mu <- a + b*( weight - xbar ) ,
a ~ dnorm( 178 , 20 ) ,
b ~ dlnorm( 0 , 1 ) ,
sigma ~ dunif( 0 , 50 )
) , data=d2 )
```

```
##
## Quadratic approximate posterior distribution
##
## Formula:
## height ~ dnorm(mu, sigma)
## mu <- a + b * (weight - xbar)
## a ~ dnorm(178, 20)
## b ~ dlnorm(0, 1)
## sigma ~ dunif(0, 50)
##
## Posterior means:
##           a           b       sigma
## 154.6013672   0.9032808   5.0718794
##
## Log-likelihood: -1071.01
```

this is the equation without the xbar

```
quap(
alist(
height ~ dnorm( mu , sigma ) ,
mu <- a + b*( weight ) ,
a ~ dnorm( 178 , 20 ) ,
b ~ dlnorm( 0 , 1 ) ,
sigma ~ dunif( 0 , 50 )
) , data=d2 )
```

```
##
## Quadratic approximate posterior distribution
##
## Formula:
## height ~ dnorm(mu, sigma)
## mu <- a + b * (weight)
## a ~ dnorm(178, 20)
## b ~ dlnorm(0, 1)
## sigma ~ dunif(0, 50)
##
## Posterior means:
##           a           b       sigma
## 114.5350229   0.8907156   5.0726946
##
## Log-likelihood: -1071.07
```

The posterior means decrease

Centering is subtracting a constant so the slope shouldn’t change but the intercept would.

4e8.

```
library(rethinking)
data(cherry_blossoms)
d <- cherry_blossoms
precis(d)
```

```
##                    mean          sd      5.5%      94.5%       histogram
## year       1408.000000 350.8845964 867.77000 1948.23000  ▇▇▇▇▇▇▇▇▇▇▇▇▁
## doy         104.540508   6.4070362  94.43000  115.00000        ▁▂▅▇▇▃▁▁
## temp          6.141886   0.6636479   5.15000    7.29470        ▁▃▅▇▃▂▁▁
## temp_upper    7.185151   0.9929206   5.89765    8.90235  ▁▂▅▇▇▅▂▂▁▁▁▁▁▁▁
## temp_lower    5.098941   0.8503496   3.78765    6.37000  ▁▁▁▁▁▁▁▃▅▇▃▂▁▁▁
```

```
d |>
  count(is.na(doy)) |>
  mutate(percent = 100 * n / sum(n))
```

```
##   is.na(doy)   n  percent
## 1      FALSE 827 68.06584
## 2       TRUE 388 31.93416
```

```
d2 <-
  d |>
  drop_na(doy)
```

```
num_knots <- 15
knot_list <- quantile(d2$year, probs = seq(from = 0, to = 1, length.out = num_knots))
```

same thing but with higher number of knots

```
num_knots2 <-45
knot_list <- quantile(d2$year, probs = seq(from = 0, to = 1, length.out = num_knots2))
```

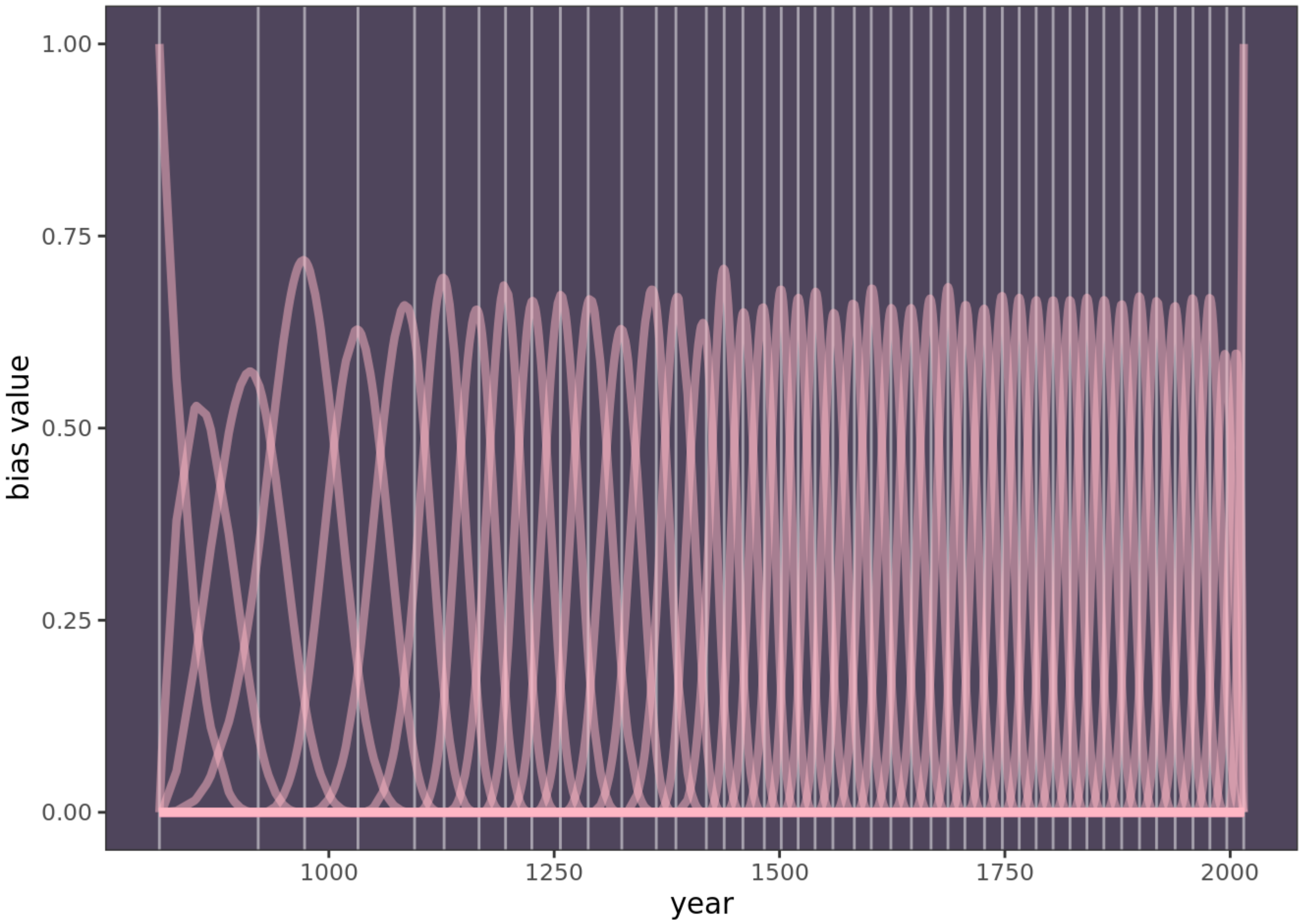
```
library(splines)

B <- bs(d2$year,
        knots = knot_list[-c(1, num_knots2)],
        degree = 3,
        intercept = TRUE)
```

```
B %>% str()
```

```
##  'bs' num [1:827, 1:47] 1 0.92 0.565 0.267 0.245 ...
##  - attr(*, "dimnames")=List of 2
##   ..$ : NULL
##   ..$ : chr [1:47] "1" "2" "3" "4" ...
##  - attr(*, "degree")= int 3
##  - attr(*, "knots")= Named num [1:43] 922 973 1032 1095 1128 ...
##   ..- attr(*, "names")= chr [1:43] "2.272727%" "4.545455%" "6.818182%" "9.090909%" ...
##  - attr(*, "Boundary.knots")= int [1:2] 812 2015
##  - attr(*, "intercept")= logi TRUE
```

```
b <-
  B %>%
  data.frame() %>%
  set_names(str_c(0, 1:9), 10:47) %>%
  bind_cols(select(d2, year)) %>%
  pivot_longer(-year,
               names_to = "bias_function",
               values_to = "bias")
```

```
b %>%
  ggplot(aes(x = year, y = bias, group = bias_function)) +
  geom_vline(xintercept = knot_list, color = "white", alpha = 1/2) +
  geom_line(color = "#ffb7c5", alpha = 1/2, size = 1.5) +
  ylab("bias value") +
  theme_bw() +
  theme(panel.background = element_rect(fill = "#4f455c"),
        panel.grid = element_blank())
```



The increase in the number of knots has made this graph way more wavy and closer together than the one in the book. I guess maybe more knots makes it more wigglier.