Homework Chapter 11

Emily Maloney

March 24, 2019

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------------------------------- tidyverse 1.2.

## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  2.0.1     v dplyr   0.7.8
## v tidyr   0.8.2     v stringr 1.3.1
## v readr   1.3.1     v forcats 0.3.0

## -- Conflicts ------------------------------------------------------------------ tidyverse_conflicts(
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(rethinking)
```

```
## Loading required package: rstan

## Loading required package: StanHeaders

## rstan (Version 2.18.2, 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)

## For improved execution time, we recommend calling
## Sys.setenv(LOCAL_CPPFLAGS = '-march=native')
## although this causes Stan to throw an error on a few processors.

##
## Attaching package: 'rstan'

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

## Loading required package: parallel

## rethinking (Version 1.59)

##
## Attaching package: 'rethinking'

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

```r
library(brms)
```

```
## Loading required package: Rcpp

## Loading 'brms' package (version 2.7.0). Useful instructions
## can be found by typing help('brms'). A more detailed introduction
## to the package is available through vignette('brms_overview').
```

```
## Run theme_set(theme_default()) to use the default bayesplot theme.
##
## Attaching package: 'brms'
## The following objects are masked from 'package:rethinking':
##
##     LOO, stancode, WAIC
## The following object is masked from 'package:rstan':
##
##     loo
```

```r
library(knitr)
library(bayesplot)
```

```
## This is bayesplot version 1.6.0
## - Online documentation and vignettes at mc-stan.org/bayesplot
## - bayesplot theme set to bayesplot::theme_default()
##    * Does _not_ affect other ggplot2 plots
##    * See ?bayesplot_theme_set for details on theme setting
```

Chapter 11

Easy Problems

11E1

An ordered categorical variable is one in which the points on the scale correspond to levels while an unordered categorical variable cannot be ranked against each other. An example of an ordered categorical variable is grades (A, B, C, D, F), and an example of an unordered categorical variable is sex (Male, Female, Intersex).

11E2

An ordered logistic regression employs a cumulative logit link, which means that it is the odds for that value on the scale or anything below it.

11E3

When count data are zero-inflated, using a model that ignores zero-inflation will tend to induce Type II error.

11E4

An example of a process that might produce over-dispersed counts would be the number of times someone talks in class, because everyone has a different threshold for speaking. An example of a process that might produce under-dispersed counts would be how many times someone is stopped by the police by race and gender, because I think the variance would be lower than the expected value.

Medium Problems

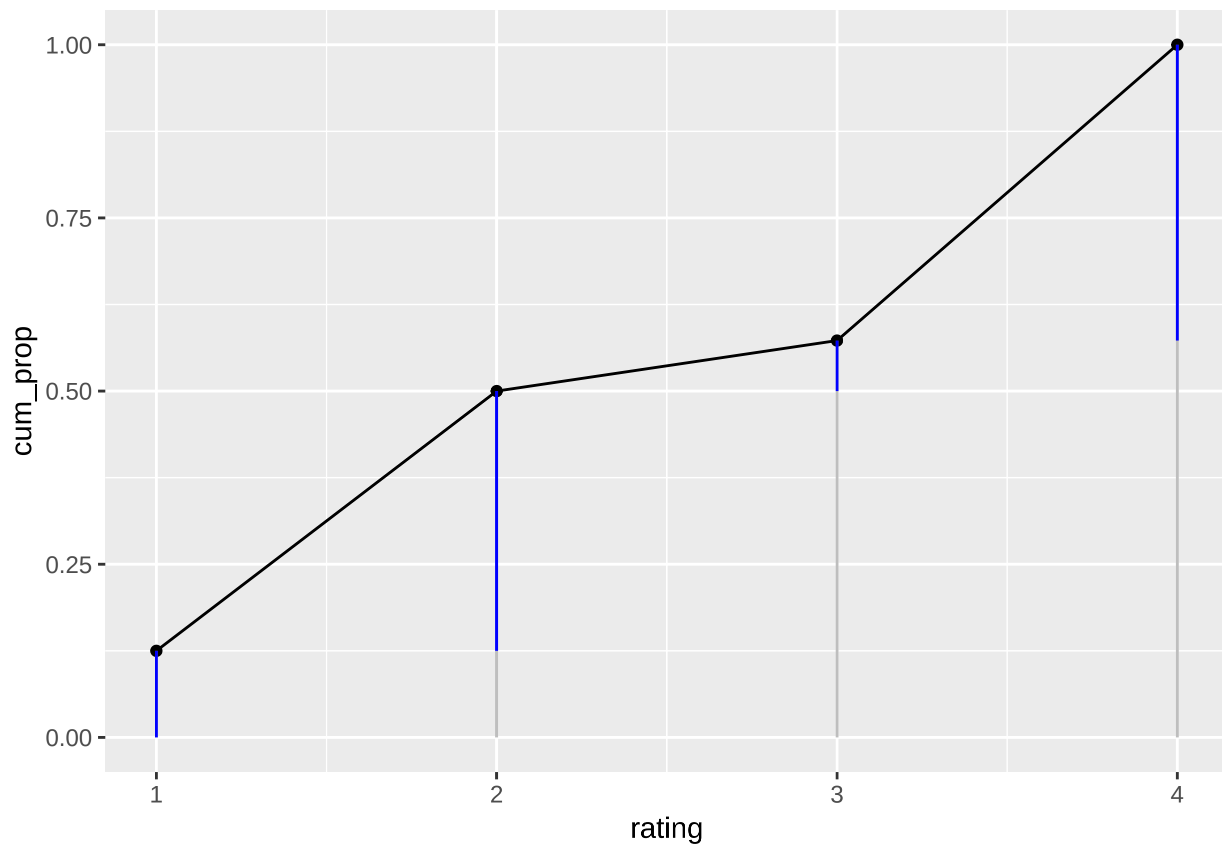
11M1

```r
d <- tibble(rating = seq(1, 4, by = 1),
               freq = c(12, 36, 7, 41))

d <- d %>% mutate(total = sum(freq),
                  cum_sum = cumsum(freq),
                  cum_prop = cum_sum/total,
                  cum_odds = cum_prop/(1-cum_prop),
                  log_cum_odds = log(cum_odds),
                  prev_prop = lag(cum_prop, default = 0))

table <- d %>% select(rating, log_cum_odds)
kable(table)
```

rating	log_cum_odds
1	-1.9459101
2	0.0000000
3	0.2937611
4	Inf

11M2

```r
ggplot(data = d, mapping = aes(x = rating, y = cum_prop)) +
  geom_line() +
  geom_point() +
  geom_linerange(mapping = aes(ymin = 0, ymax = cum_prop), alpha = 3/4, color = "dark gray") +
  geom_linerange(mapping = aes(ymin = prev_prop, ymax = cum_prop), color = "blue", alpha = 1)
```



11M3

With probability d that secondary process produces 0, n equal to the number of trials, and p the probability of success, the likelihood of a zero result is:
d + (1- d)(1-p)^n
and the likelihood of a non-zero result is:
(1 - d)(p^y)(n choose y)(1-p)^{(n-y)}