# PSY 607 Bayesian Analysis

Sara Weston

Spring 2025

# Contents

# Chapter 1

# Course description

## 1.1 Structure

## 1.2 Materials

## 1.3 Schedule

## 1.4 Grades

## 1.5 Policies

### 1.5.1 Abscences

### 1.5.2 Communication

If you have questions about course policies, have trouble submitting an assignment, or want to schedule a meeting, please email. I will make an effort to respond to emails within one business day. Note that I neither plan nor commit to checking email outside of normal business hours (9am-5pm, Mon-Fri).

If you are having trouble understanding a concept covered in class, please come to office hours, schedule a meeting with me, or ask for clarification during class periods. I will not explain course concepts over email.

Occasionally, I will send out announcements to the entire class via Canvas announcements. These will typically appear when you open Canvas, but you can update your Canvas settings to receive these announcements as emails. It is strongly recommended that you do so.

### 1.5.3 Classroom

All members of the class (students and instructor) can expect to:

*Participate and Contribute:* All students are expected to participate by sharing ideas and contributing to the learning environment. This entails preparing, following instructions, and engaging respectfully and thoughtfully with others.

While all students should participate, participation is not just talking, and a range of participation activities support learning. Participation might look like speaking aloud in the full class and in small groups and collaborating on homework assignments.

*Expect and Respect Diversity:* All classes at the University of Oregon welcome and respect diverse experiences, perspectives, and approaches. What is not welcome are behaviors or contributions that undermine, demean, or marginalize others based on race, ethnicity, gender, sex, age, sexual orientation, religion, ability, or socioeconomic status. We will value differences and communicate disagreements with respect.

*Help Everyone Learn:* Part of how we learn together is by learning from one another. To do this effectively, we need to be patient with each other, identify ways we can assist others, and be open-minded to receiving help and feedback from others. Don't hesitate to contact me to ask for assistance or offer suggestions that might help us learn better.

### 1.5.4 Workload

This is a 3-credit hour course, so you should expect to complete 120 hours of work for the course—an average of about 12 hours each week (this includes time in-class).

### 1.5.5 Generative AI

### 1.5.6 Plagiarism

### 1.5.7 Accessibility

### 1.5.8 Basic needs

### 1.5.9 Reporting obligations

### 1.5.10 Campus emergencies

# Chapter 2

# Week 1: Introduction to Bayesian Analysis

## 2.1 Class 1: Probability

### 2.1.1 Slides

This lecture is based on the following article: Etz and Vandekerckhove (2018). download here.

## 2.2 Class 2: Bayes as counting

Before class, be sure to watch both of the following lectures by McElreath: * Science before statistics * Garden of forking data

This lecture is based on Chapters 2 and 3 in *Statistical Rethinking* by Richard McElreath.

# Chapter 3

# Week 2: Linear models and causal inference

## 3.1 Class 1: Geocentric models

Before class, be sure to watch the lecture by McElreath.

### 3.1.1 Slides

This lecture is based on Chapters 4 in *Statistical Rethinking* by Richard McElreath.

## 3.2 Class 2: Categories (and curves)

Before class, be sure to watch the lecture by McElreath.

### 3.2.1 Slides

This lecture is based on Chapters 4 and 5 in *Statistical Rethinking* by Richard McElreath.

### 3.2.2 Curves (splines)

We won't have time to address curves in class, and McElreath doesn't give you a lot of code to work with in the lecture. Here's how to reproduce his spline model from the lecture.

#### 3.2.2.1 Data preparation

```r
library(rethinking)
library(psych)
library(tidyverse)
library(splines)
```

```r
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
```

```r
psych::describe(d)
```

```
##            vars    n    mean     sd  median trimmed    mad    min     max
## year          1 1215 1408.00 350.88 1408.00 1408.00 450.71 801.00 2015.00
## doy           2  827  104.54   6.41  105.00  104.54   5.93  86.00  124.00
## temp          3 1124    6.14   0.66    6.10    6.11   0.61   4.67    8.30
## temp_upper    4 1124    7.19   0.99    7.04    7.10   0.92   5.45   12.10
## temp_lower    5 1124    5.10   0.85    5.14    5.10   0.72   0.75    7.74
##              range  skew kurtosis    se
## year       1214.00  0.00    -1.20 10.07
## doy          38.00  0.00    -0.15  0.22
## temp          3.63  0.40     0.11  0.02
## temp_upper    6.65  1.05     1.71  0.03
## temp_lower    6.99 -0.17     1.88  0.03
```

Note that some of the values for `doy` (day of year) are missing. The `quap()` function can't handle missingness, so we'll remove those rows before proceeding.

```r
d2 <- d[ complete.cases(d$doy) , ] # complete cases on doy
```

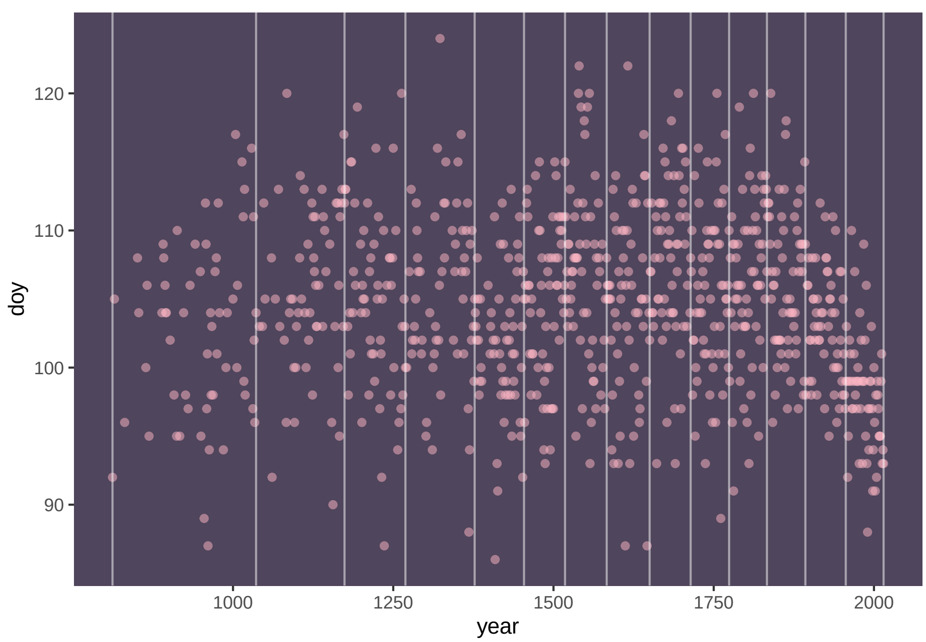
Next we set up an arbitrary number of knots. Knots divide our data into $N_{knots}+1$ equal bins, such that each bin has the same number of data points. McElreath chooses 15 knots.

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

```
##        0% 7.142857% 14.28571% 21.42857% 28.57143% 35.71429% 42.85714%       50%
##       812      1036      1174      1269      1377      1454      1518      1583
## 57.14286% 64.28571% 71.42857% 78.57143% 85.71429% 92.85714%      100%
##      1650      1714      1774      1833      1893      1956      2015
```

To visualize how our data have been partitioned:

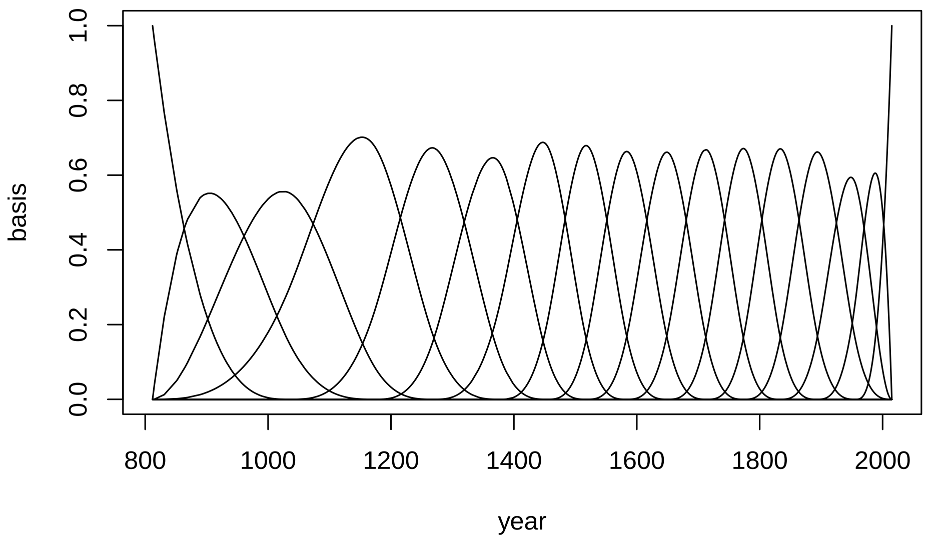
```
d2 %>%
  ggplot(aes(x = year, y = doy)) +
  geom_point(color = "#ffb7c5", alpha = 1/2) +
  geom_vline(xintercept = knot_list, color = "white", alpha = .5) +
  theme(panel.grid = element_blank(),
        panel.background = element_rect(fill = "#4f455c"))
```




Next, we'll use functions in the `spline` package to create basis spline functions based on these knots

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

```r
plot( NULL , xlim=range(d2$year) , ylim=c(0,1) , xlab="year" , ylab="basis" )
for ( i in 1:ncol(B) ) lines( d2$year , B[,i] )
```




#### 3.2.2.2 Mathematical model and `quap()`

Here is the mathematical model for the spline model:

$$
\begin{aligned}
D_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha + \sum_{k=1}^{K} w_k B_{k,i} \\
\alpha &\sim \text{Normal}(100, 10) \\
w_j &\sim \text{Normal}(0, 10) \\
\sigma &\sim \text{Exponential}(1)
\end{aligned}
$$

And here's how we can fit this using `quap()`. Note that this is the first use of the `start` function, which gives the estimation algorithm a start value. This can be helpful if you find that models are not converging.

```r
m5<- quap(

  alist(
  D ~ dnorm( mu , sigma ) ,
  mu <- a + B %*% w ,
```

```
  a ~ dnorm(100,10),
  w ~ dnorm(0,10),
  sigma ~ dexp(1)),

  data=list( D=d2$doy , B=B ) ,

  start=list( w=rep( 0 , ncol(B) ) ) )
```

# Chapter 4

# Week 3: Causes, Confounds, and Colliders

## 4.1 Class 1: Elemental confounds

Before class, be sure to watch the lecture by McElreath.

### 4.1.1 Slides

This lecture is based on Chapters 5 and 6 in *Statistical Rethinking* by Richard McElreath.

You'll also want this code to simulate some fake plant data in class.

```
set.seed(71)
# number of plants
N <- 100
# simulate initial heights
h0 <- rnorm(N,10,2)
# assign treatments and simulate fungus and growth
treatment <- rep( 0:1 , each=N/2 )
fungus <- rbinom( N , size=1 , prob=0.5 - treatment*0.4 )
h1 <- h0 + rnorm(N, 5 - 3*fungus)
# compose a clean data frame
d <- data.frame( h0=h0 , h1=h1 , treatment=treatment , fungus=fungus )
```

## 4.2 Class 2: Categories (and curves)

Before class, be sure to watch the lecture by McElreath.

### 4.2.1 Slides

This lecture is based on Chapter 6 in *Statistical Rethinking* by Richard McElreath.

### 4.2.2 Simulation Code

#### 4.2.2.1 Simulation 1: Simple Confounding

In this code, the true causal effect of X on Y is 0, but confounded by U.

```
#number of sims
N = 1000
# Generate data
U <- rnorm(N)  # Unobserved confounder
X <- rnorm(N, mean = 0.5 * U)  # Treatment affected by U
Y <- rnorm(N, mean = 0.8 * U)  # Outcome affected by U
Z <- rnorm(N, mean = 0.6 * U)  # Observed variable that captures U

d <- data.frame(X, Y, Z)

# Fit models
flist1 <- alist(
  Y ~ dnorm(mu, sigma),
  mu <- a + bX*X,
  a ~ dnorm(0, .5),
  bX ~ dnorm(0, .25),
  sigma ~ dexp(1)
)

m32.1 <- quap(flist1, d)
precis(m32.1)

# Fit models
flist2 <- alist(
  Y ~ dnorm(mu, sigma),
  mu <- a + bX*X +bZ*Z,
  a ~ dnorm(0, .5),
  bX ~ dnorm(0, .25),
```

```
  bZ ~ dnorm(0, .25),
  sigma ~ dexp(1)
)

m32.2 <- quap(flist2, d)
precis(m32.2)

post.1 <- extract.samples(m32.1)
post.2 <- extract.samples(m32.2)

results_df = data.frame(naive = post.1$bX,
                        adjusted = post.2$bX)
results_df %>%
  pivot_longer(everything()) %>%
  ggplot(aes(x = value, fill = name)) +
  geom_density(alpha = .5) +
  geom_vline(aes(xintercept = 0), linetype = "dashed")
```

#### 4.2.2.2 Simulation 2: Collider Bias

In this code, the true causal effect of X on Y is 0, but controlling for Z (collider) creates bias.

# Chapter 5

# Week 4: Overfitting/MCMC

## 5.1 Class 1: Overfitting

Before class, be sure to watch the lecture by McElreath.

### 5.1.1 Slides

This lecture is based on Chapter 7 in *Statistical Rethinking* by Richard McElreath.

## 5.2 Class 2: MCMC

Before class, be sure to watch the lecture by McElreath.

### 5.2.1 Slides

This lecture is based on Chapter 9 in *Statistical Rethinking* by Richard McElreath.

# Bibliography

Etz, A. and Vandekerckhove, J. (2018). Introduction to Bayesian Inference for Psychology. *Psychonomic Bulletin & Review*, 25(1):5–34.