STAT 237
Hierarchical Models

April 11-15, 2022

Colin Reimer Dawson

Outline

Outline

Review: Beta-Bernoulli Model

We have worked extensively with the following model (prior and likelihood) for conditionally independent binary observations, $\mathbf{y} = (y_1, \ldots, y_N)$

$$\mu \sim \text{Beta}(a_0, b_0)$$

$$y_n \mid \mu \overset{i.i.d.}{\sim} \text{Bernoulli}(\mu), n = 1, \ldots, N$$

Review: Beta-Bernoulli Model

We have worked extensively with the following model (prior and likelihood) for conditionally independent binary observations, $\mathbf{y} = (y_1, \ldots, y_N)$

$$\mu \sim \text{Beta}(a_0, b_0)$$
$$y_n \mid \mu \overset{i.i.d.}{\sim} \text{Bernoulli}(\mu), n = 1, \ldots, N$$

Defining $N_1 := \sum_{n=1}^{N} y_n$ to be the number of "successes" in the data, after updating, the posterior is then

$$\mu \mid \mathbf{y} \sim \text{Beta}(a_0 + N_1, b_0 + N - N_1)$$

In other words, the number of "successes" (1s) and "failures" (0s) in the data add to the a and b parameters of the Beta respectively

Review: Beta-Bernoulli Model

The mean of a Beta(a, b) distribution is $\frac{a}{a+b}$, so we have

$$\mathbb{E}[\theta] = \frac{a_0}{a_0 + b_0}$$

$$\begin{aligned}\mathbb{E}[\theta \mid \mathbf{y}] &= \frac{a_0 + N_1}{a_0 + N_1 + b_0 + N - N_1} \\ &= \frac{a_0 + N_1}{a_0 + b_0 + N}\end{aligned}$$

Review: Beta-Bernoulli Model

The mean of a Beta(a, b) distribution is $\frac{a}{a+b}$, so we have

$$\mathbb{E}[\theta] = \frac{a_0}{a_0 + b_0}$$

$$\begin{aligned}\mathbb{E}[\theta \mid \mathbf{y}] &= \frac{a_0 + N_1}{a_0 + N_1 + b_0 + N - N_1} \\ &= \frac{a_0 + N_1}{a_0 + b_0 + N}\end{aligned}$$

Note that we can rewrite the posterior mean as follows:

$$\begin{aligned}\mathbb{E}[\theta \mid \mathbf{y}] &= \frac{a_0}{a_0 + b_0 + N} + \frac{N_1}{a_0 + b_0 + N} \\ &= \frac{a_0}{a_0 + b_0}\frac{a_0 + b_0}{a_0 + b_0 + N} + \frac{N_1}{N}\frac{N}{a_0 + b_0 + N} \\ &= \mathbb{E}[\theta]\left(\frac{a_0 + b_0}{a_0 + b_0 + N}\right) + \bar{y}\left(\frac{N}{a_0 + b_0 + N}\right)\end{aligned}$$

Review: Beta-Bernoulli Model

Defining

$$\alpha = \frac{a_0 + b_0}{a_0 + b_0 + N}$$

this is

$$\begin{aligned} \mathbb{E}\left[\theta \mid \mathbf{y}\right] &= \mathbb{E}\left[\theta\right]\left(\frac{a_0 + b_0}{a_0 + b_0 + N}\right) + \bar{y}\left(\frac{N}{a_0 + b_0 + N}\right) \\ &= \mathbb{E}\left[\theta\right]\alpha + \bar{y}(1 - \alpha) \end{aligned}$$

Review: Beta-Bernoulli Model

Defining

$$\alpha = \frac{a_0 + b_0}{a_0 + b_0 + N}$$

this is

$$\begin{aligned} \mathbb{E}\left[\theta \mid \mathbf{y}\right] &= \mathbb{E}\left[\theta\right]\left(\frac{a_0 + b_0}{a_0 + b_0 + N}\right) + \bar{y}\left(\frac{N}{a_0 + b_0 + N}\right) \\ &= \mathbb{E}\left[\theta\right]\alpha + \bar{y}(1 - \alpha) \end{aligned}$$

That is, the posterior mean is a **weighted average** of the **prior mean** and the **data mean**, with weights coming from the **virtual sample size** of the prior and **actual sample size** of the data

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- There are 10000 flips overall, each with a binary result

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- There are 10000 flips overall, each with a binary result
- We could model $Y_1, \ldots Y_{10000} \mid \mu \sim \text{Bernoulli}(\mu)$ but... that throws away structure

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- There are 10000 flips overall, each with a binary result
- We could model $Y_1, \ldots Y_{10000} \mid \mu \sim \text{Bernoulli}(\mu)$ but... that throws away structure
- If **some** people are psychic and others aren't, then the psychic individuals' have a different μ

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- There are 10000 flips overall, each with a binary result
- We could model $Y_1, \ldots Y_{10000} \mid \mu \sim \text{Bernoulli}(\mu)$ but... that throws away structure
- If **some** people are psychic and others aren't, then the psychic individuals' have a different μ
- We'd like a model that allows each person to have a different μ, while also being able to learn something about the population as a whole

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- Let $s = 1, \ldots, S$ index subjects (people; here $S = 1000$) and $n = 1, \ldots N$ index trials (here, $N = 10$)

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- Let $s = 1, \ldots, S$ index subjects (people; here $S = 1000$) and $n = 1, \ldots N$ index trials (here, $N = 10$)
- We can define $\theta_1, \ldots, \theta_S$ to be the success chances for each individual

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- Let $s = 1, \ldots, S$ index subjects (people; here $S = 1000$) and $n = 1, \ldots N$ index trials (here, $N = 10$)
- We can define $\theta_1, \ldots, \theta_S$ to be the success chances for each individual
- How to estimate these?

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- Let $s = 1, \ldots, S$ index subjects (people; here $S = 1000$) and $n = 1, \ldots N$ index trials (here, $N = 10$)
- We can define $\theta_1, \ldots, \theta_S$ to be the success chances for each individual
- How to estimate these?
- We need a **prior** the θ_ss. One possibility:

$$\theta_s \overset{i.i.d.}{\sim} \text{Beta}(a_0, b_0)$$

where we specify a_0 and b_0 based on background information

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- Let $s = 1, \ldots, S$ index subjects (people; here $S = 1000$) and $n = 1, \ldots N$ index trials (here, $N = 10$)
- We can define $\theta_1, \ldots, \theta_S$ to be the success chances for each individual
- How to estimate these?
- We need a **prior** the θ_ss. One possibility:

$$\theta_s \overset{i.i.d.}{\sim} \text{Beta}(a_0, b_0)$$

  where we specify a_0 and b_0 based on background information
- This would split the problem into 1000 separate Beta-Bernoulli models. But this doesn't help us learn

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- One possibility:

$$\theta_s \overset{i.i.d.}{\sim} \mathsf{Beta}(a_0, b_0)$$

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- One possibility:

$$\theta_s \overset{i.i.d.}{\sim} \text{Beta}(a_0, b_0)$$

- This would split the problem into 1000 separate Beta-Bernoulli models. But this doesn't help us learn much about the **population as a whole**

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- One possibility:

$$\theta_s \overset{i.i.d.}{\sim} \text{Beta}(a_0, b_0)$$

- This would split the problem into 1000 separate Beta-Bernoulli models. But this doesn't help us learn much about the **population as a whole**
- Instead, we might let the parameters of the Beta distribution be **random variables themselves**

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- One possibility:

$$\theta_s \overset{i.i.d.}{\sim} \text{Beta}(a_0, b_0)$$

- This would split the problem into 1000 separate Beta-Bernoulli models. But this doesn't help us learn much about the **population as a whole**
- Instead, we might let the parameters of the Beta distribution be **random variables themselves**
- For instance, let ω represent the **mean success rate** across the population, and κ tell us something about the **homogeneity of the population**

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- One possibility:

$$\theta_s \overset{i.i.d.}{\sim} \text{Beta}(a_0, b_0)$$

- This would split the problem into 1000 separate Beta-Bernoulli models. But this doesn't help us learn much about the **population as a whole**
- Instead, we might let the parameters of the Beta distribution be **random variables themselves**
- For instance, let ω represent the **mean success rate** across the population, and κ tell us something about the **homogeneity of the population**
- Then we might define

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- Let ω represent the **mean success rate** across the population, and κ tell us something about the **homogeneity of the population**

A Hierarchical Extension

ESP

Recall Peter Venkman's experiment to test for extra-sensory perception: Each of 1000 people attempted to call the result of 10 consecutive coin flips. How can we model this data?

- Let ω represent the **mean success rate** across the population, and κ tell us something about the **homogeneity of the population**
- Then we might define

$$\theta_s \mid \omega, \kappa \overset{i.i.d.}{\sim} \text{Beta}(\kappa\omega, \kappa(1-\omega))$$

and put **hyperpriors** on κ and ω. Such as:

$$\omega \sim \text{Beta}(a_0, b_0)$$
$$\kappa \sim \text{Gamma}(k, r)$$

A Hierarchical Extension

- We might define

$$\theta_s \mid \omega, \kappa \overset{i.i.d.}{\sim} \text{Beta}(\kappa\omega, \kappa(1-\omega))$$

and put **hyperpriors** on κ and ω. Such as:

$$\omega \sim \text{Beta}(a_0, b_0)$$
$$\kappa \sim \text{Gamma}(k, r)$$

A Hierarchical Extension

- We might define

$$\theta_s \mid \omega, \kappa \overset{i.i.d.}{\sim} \text{Beta}(\kappa\omega, \kappa(1-\omega))$$

  and put **hyperpriors** on κ and ω. Such as:

$$\omega \sim \text{Beta}(a_0, b_0)$$
$$\kappa \sim \text{Gamma}(k, r)$$

- Then, we use the data $y_{1|1}, \ldots y_{N|1}, y_{1|2}, \ldots, y_{N|2}, \ldots, y_{1|S}, \ldots, y_{N|S}$ to obtain a **joint posterior density**

$$p(\boldsymbol{\theta}, \omega, \kappa \mid \mathbf{Y}) = \frac{p(\omega, \kappa, \boldsymbol{\theta}) p(\mathbf{Y} \mid \omega, \kappa, \boldsymbol{\theta})}{p(\mathbf{Y})}$$

A Hierarchical Extension

- We might define

$$\theta_s \mid \omega, \kappa \overset{i.i.d.}{\sim} \text{Beta}(\kappa\omega, \kappa(1-\omega))$$

A Hierarchical Extension

- We might define

$$\theta_s \mid \omega, \kappa \overset{i.i.d.}{\sim} \text{Beta}(\kappa\omega, \kappa(1-\omega))$$

- Then, we use the data $y_{1|1}, \ldots y_{N|1}, y_{1|2}, \ldots, y_{N|2}, \ldots, y_{1|S}, \ldots, y_{N|S}$ to obtain a **joint posterior density**

$$\begin{aligned} p(\boldsymbol{\theta}, \omega, \kappa \mid \mathbf{Y}) &= \frac{p(\omega, \kappa, \boldsymbol{\theta}) p(\mathbf{Y} \mid \omega, \kappa, \boldsymbol{\theta})}{p(\mathbf{Y})} \\ &= \frac{p(\omega) p(\kappa) \prod_{s=1}^{S} p(\theta_s \mid \omega, \kappa) \prod_{s=1}^{S} \prod_{n=1}^{N} p(y_{n|s} \mid \theta_s)}{p(\mathbf{Y})} \end{aligned}$$