Single Parameter Models

ISTA 410 / INFO 510 - Bayesian Modeling and Inference

University of Arizona School of Information

February 1, 2021

Outline

Last time:

- Joint distribution of several variables
- Conditional probability and independence
- Marginal distributions and marginalization

Today:

- One parameter models:
  - Beta-binomial models and conjugate priors
  - Normal model with known variance — Wednesday
- Summarizing posterior inferences

Recap: that one homework problem

The dice problem

Recall:

(fair)

- Have a box containing four (fair) dice: 1 d4, 2 d6s, 1 d12
- Experiment: select a die at random, roll it 1000 times
- Outcome variable: proportion of 1s

Why this problem? Think about data-generating processes with an underlying, unknown parameter.

The dice problem

The marginal distribution of the count of 1s is a mixture of binomials:

$$y \sim 0.25 \times \mathrm{Binomial}(1000, 1/4) + 0.5 \times \underline{\mathrm{Binomial}(1000, 1/6)}$$
$$+ 0.25 \times \mathrm{Binomial}(1000, 1/12)$$

distribution of "successes" in 1000 independent trials, success prob 1/6.

Let's sketch this...

The dice problem

Data-generating process:

observed outcomes.

"stage 1" "stage 2"

$p^{-1} \longrightarrow y$

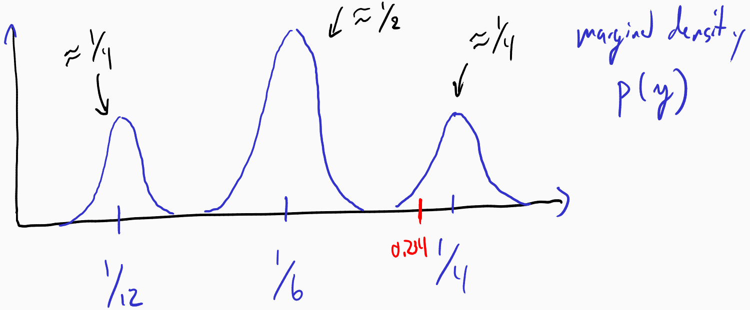
distribution of y is conditioned on p^{-1}

hidden parameters

p^{-1} – $\Pr(p^{-1}=4)=1/4$, $\Pr(p^{-1}=6)=1/2$,
$\Pr(p^{-1}=12)=1/4$

$y \sim \text{Binomial}(100, p^{-1})$

The dice problem



quantiles: $\underbrace{5\%}_{}$, 25%, 50%, 75%, 95%.

20% of 25% ~ Find 20% quantile of 1/12 peak

$$1/12 \sim \text{Normal}\left(1/12, \sqrt{\frac{1/12 \cdot 11/12}{1000}}\right)$$

qnorm in R, sp.stats.norm.ppf

Beta-binomial model

Using a Bayesian model

Steps:

- Set up a probabilistic model for the observed data, dependent on un-observed parameters
- Apply a *prior distribution* to the parameters, representing our knowledge before observing data
- Apply Bayes' theorem to update the distribution of the parameters, resulting in a *posterior distribution*
- Summarize relevant results

Bayes' theorem

model of the data-generating process; prob. of observing certain data given values of parameters.

Recall Bayes' theorem:

$$\mathrm{Pr}(H|E) = \frac{\mathrm{Pr}(E|H)\mathrm{Pr}(H)}{\mathrm{Pr}(E)}$$

Terminology:

- $\mathrm{Pr}(H|E)$ – *posterior probability*
- $\mathrm{Pr}(H)$ – *prior probability*
- $\mathrm{Pr}(E|H)$ – *likelihood*
- $\mathrm{Pr}(E)$ – *normalizing constant*

H - hypothesis

E - evidence.

expand into

$$\sum_i Pr(H_i)\, Pr(E|H_i)$$

Bayes' theorem with densities

Most commonly we have probability density functions that depend on unknown parameters:

- y – data
- θ – parameters

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$$p(\theta|y) \propto p(y|\theta)p(\theta) \leftarrow \text{proportionality.}$$

The normalizing constant $p(y)$ is gotten by marginalizing over θ by computing $\int p(y|\theta)p(\theta)d\theta$; this integral may be intractable, so we work with the proportionality statement.

Binomial model

If we are observing binary categorical outcomes, a binomial likelihood makes sense. $\mathrm{Binomial}(n, \theta)$ is the distribution of the count of "successes" in n independent trials with a fixed probability θ of success.

$$p(y \text{ successes} | \theta) = \overbrace{\binom{n}{y}}^{{}_nC_y} \underbrace{\theta^y}_{(1/4)^{234}} \underbrace{(1-\theta)^{n-y}}_{(3/4)^{766}}$$

$_nC_y$ – # of size-y subsets of a set of size n.

In the dice problem, θ was an unknown parameter with a distribution

$$\left.\begin{aligned} \Pr(\theta = 1/4) &= 1/4 \\ \Pr(\theta = 1/6) &= 1/2 \\ \Pr(\theta = 1/12) &= 1/4 \end{aligned}\right\} \text{prior distribution for } \theta$$

↑ probabilities of rolling 1

Binomial model

Suppose we ran our experiment and got 234 1s. Then, we can plug into Bayes' theorem:

$$\Pr(\theta = 1/4 | y = 234) = \frac{\Pr(y = 234 | \theta = 1/4)\Pr(\theta = 1/4)}{\Pr(y = 234)}$$

H: $\theta = 1/4$

E: $y = 234$

$$\frac{\overbrace{(0.25)}^{\text{prior}}\overbrace{(1/4)^{234}(3/4)^{766}}^{\text{likelihood}}}{[\underbrace{(0.25)}_{\text{prior}}(1/4)^{234}(3/4)^{766} + \underbrace{(0.5)}_{\text{prior}}(1/6)^{234}(5/6)^{766} + \underbrace{(0.25)}_{\text{prior}}(1/12)^{234}(11/12)^{766}]}$$

which comes out to about 1. (So we can be very certain that we are rolling the d4.)

Binomial model

Suppose we ran our experiment and got 234 1s. Then, we can plug into Bayes' theorem:

$$\Pr(\theta = 1/4 | y = 234) = \frac{\Pr(y = 234 | \theta = 1/4) \Pr(\theta = 1/4)}{\Pr(y = 234)}$$

$$\frac{(0.25)(1/4)^{234}(3/4)^{766}}{[(0.25)(1/4)^{234}(3/4)^{766} + (0.5)(1/6)^{234}(5/6)^{766} + (0.25)(1/12)^{234}(11/12)^{766}]}$$

which comes out to about 1. (So we can be very certain that we are rolling the d4.)

So, this is the discrete version. What if we don't have four discrete physical dice, but an unknown continuous θ?

A continuous prior

A common choice of prior for a binomial likelihood is a beta distribution:

$$\theta \sim \mathrm{Beta}(\alpha, \beta)$$

where $\alpha, \beta > 0$ are chosen ahead of time.

Beta distribution: defined on $[0, 1]$ by the PDF

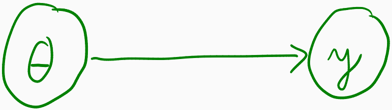
$$p(\theta) = \boxed{\frac{1}{B(\alpha, \beta)}} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$B(\alpha, \beta)$ is the normalizing constant, called a *Beta function*. There are formulas for it but not important for us right now.

What is the data-generating process?

The generative procedure now:

1. Draw a value of θ from $\mathrm{Beta}(\alpha, \beta)$
2. Draw a value of y from $\mathrm{Binomial}(n, \theta)$



What is the data-generating process?

The generative procedure now:

1. Draw a value of θ from $\mathrm{Beta}(\alpha, \beta)$
2. Draw a value of y from $\mathrm{Binomial}(n, \theta)$

- The dice problem: y's distribution was a finite mixture of binomials, weighted by the PMF of θ
- Now: y's distribution is an infinite mixture of binomials, weighted by the PDF of θ

Conjugate prior

One reason for the choice of beta prior: *conjugacy*

A distribution $p(\theta)$ is conjugate to a likelihood $p(y|\theta)$ if the posterior distribution $p(\theta|y)$ is a member of the same family as $p(\theta)$:

In the beta-binomial model,

$$p(\theta|y) = \underbrace{\frac{1}{p(y)}}_{\text{From Bayes.}} \overbrace{\frac{1}{B(\alpha, \beta)}}^{\text{prior}} \overbrace{\binom{n}{y} \theta^y (1-\theta)^{n-y}}^{\text{binomial likelihood}} \overbrace{\theta^{\alpha-1}(1-\theta)^{\beta-1}}^{\text{comes from prior}}$$

Bayes' Thm.

The leading three factors don't depend on θ, so we absorb them into a single constant.

$\frac{1}{Z}$

Conjugate prior

Now:

$$
\begin{aligned}
p(\theta|y) &= \frac{1}{Z}\theta^{y}(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= \frac{1}{Z}\theta^{\alpha+y-1}(1-\theta)^{\beta+(n-y)-1}
\end{aligned}
$$

Since the dependence of the density on θ is that of a beta distribution with parameters $(\alpha + y, \beta + (n - y))$, the constant Z must be the corresponding beta function, and

$$\theta|y \sim \mathrm{Beta}(\alpha + y, \beta + (n - y))$$

Computationally very convenient! Convenience less important these days than it used to be, though.

Posterior distribution

So, in a beta-binomial model:

$$y \sim \text{Binomial}(n, \theta)$$
$$\theta \sim \text{Beta}(\alpha, \beta)$$

if we observe y successes and $n - y$ failures, the posterior distribution of θ is

$$\theta | y \sim \text{Beta}(\alpha + y, \beta + (n - y))$$

Posterior distribution

So, in a beta-binomial model:

$$y \sim \mathrm{Binomial}(n, \theta)$$
$$\theta \sim \mathrm{Beta}(\alpha, \beta)$$

if we observe y successes and $n - y$ failures, the posterior distribution of θ is

$$\theta | y \sim \mathrm{Beta}(\alpha + y, \beta + (n - y))$$

Interpretation: we may think of the prior parameters $\alpha - 1, \beta - 1$ as *pseudocounts*

$\mathrm{Beta}(1,1) = \mathrm{Uniform}([0,1])$

Summarizing inferences; example

Inferences from the posterior

The posterior distribution is the primary product of inference; it contains all that we know about the parameter after incorporating prior and data.

In practice, often want to distill out some summary statistics:

- posterior mean – expected value of θ under the posterior distribution
  Statistical Rethinking – 89%
- posterior intervals – 95% common, but arbitrary. Note difference between highest density and central intervals
- maximum a posteriori estimate – often not a good choice, especially if the model has many parameters

Example: Pfizer's vaccine trial

Prominent recent example: beta-binomial model in analysis of Pfizer's COVID-19 vaccine

Trial procedure:

- Study participants divided randomly into two "arms": control/placebo and vaccine
- Control arm given placebo, vaccine arm given vaccine
- Watch both groups and count cases, running the analysis when a predetermined number of cases is observed

$\theta \sim \text{Beta}(\alpha, \beta)$ $\qquad$ $y \sim \text{Binomial}(\underset{\text{predetermined \#}}{\underline{n}}, \theta)$

Beta-binomial model

Defining parameters:

- π_c: probability that a control subject becomes ill
- π_v: probability that a vaccinated subject becomes ill
- Derived quantity: Vaccine efficacy:

$$VE = 1 - \frac{\pi_v}{\pi_c}$$

Parameter for the model:

$$\theta = \frac{1 - VE}{2 - VE} = \frac{\pi_v}{\pi_v + \pi_c}$$

goes into binomial likelihood.

Measures the probability that a case came from the vaccine arm

Pfizer's prior

Let y be the number of cases that come from the vaccinated group.

The model:

$$y \sim \text{Binomial}(\theta, n)$$
$$\theta \sim \text{Beta}(\underline{0.700102}, \underline{1})$$

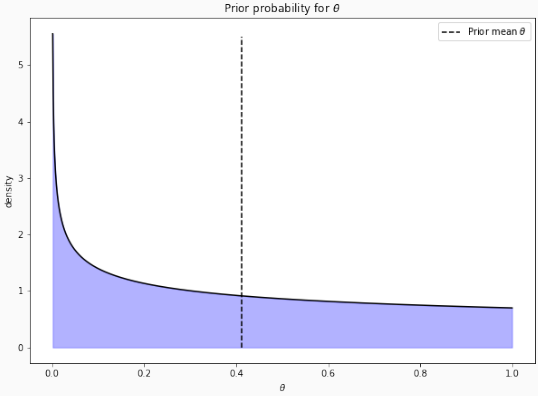
Prior was stated in Pfizer's press release. No specific reason given for these parameters, but:

presumably actual starting point.

- VE at prior mean θ is 30%
- fairly uninformative: 95% interval of VE is about $(-26.2, 0.995)$.

Pfizer's prior

On the θ scale:

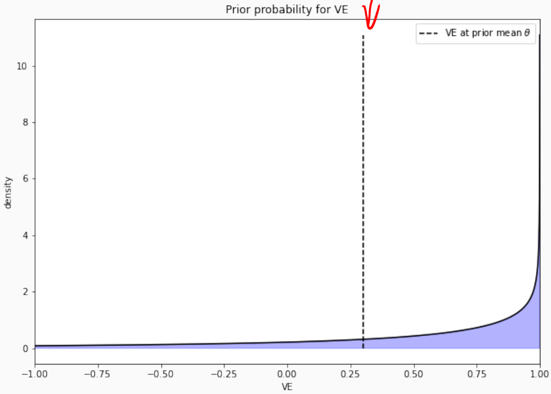


Pfizer's prior

On the VE scale

$VE = 1 - \frac{\pi_v}{\pi_c}$

prior mean 30%



What's the data?

The result of the study submitted to the FDA to obtain an emergency use authorization had a total of 170 observed cases, 8 of which were in the vaccine arm. So:

$$\theta | y \sim \mathrm{Beta}(\underbrace{0.700102 + 8}_{+8 \text{ from vaccine arm}}, \underbrace{1 + 162}_{+162 \text{ from non-vaccine arm}})$$

Let's examine this graphically...

Summary

Today:

- Beta-binomial model
- Posterior summary statistics

Next time:

- Normal model with known variance
- Some considerations for choosing priors

HW1 on D2L tonight! Due Friday, 2/12.