DATA 582: Bayesian Inference

Lecture 3: Beta-Binomial Model

Dr. Irene Vrbik

UBCO MDS

Introduction

- In today lecture we will be considering Bayesian inference on a binomial proportion.

- That is, suppose we are studying a Binomial experiment wherein the number of "successes" is recorded in n independent Bernoulli trails.

- Our goal will be to reasonably estimate the probability of success θ, i.e. the proportion of successes in a repeated set of trials.

Recall the Steps to Bayesian Inference

Recall Bayes Rule:

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta)$$
$$\text{posterior} \propto \text{prior} \times \text{likelihoood}$$

The steps of Bayesian inference can be summarized as follows:

1. Specify a prior
2. Identify the sampling density
3. Specify a likelihood
4. Derive a posterior

Bayesian Inference of a Binomial Proportion

- Recall, a Bernoulli RV can take on one of only two values: “success” (represented by 1), or “failure” (represented by 0).
- A binomial experiment conducts n independent Bernoulli trials and counts the number of successes. If we denote the number of successes in n independent Bernoulli trials by X, we say, $X \sim \text{Binomial}(n, \theta)$.
- We describe the goal of estimating θ as *Bayesian inference of a Binomial proportion*.

Bayesian Inference of a Binomial Proportion

- This experiment is one that arises commonly in practice.
- For example consider estimating the proportion of:
  - machines with a defect
  - individuals who would vote for a certain candidate running for president
  - patients who respond to an experimental drug
  - error transactions in audits
  - individuals who click on a website

- Apart from begin an extremely important statistical technique, this example will form the basis for more complicated topics to follow.

- To demonstrate the steps of the Bayesian data analysis process, we will look at a specific example involving tuition increase at the University of Iowa.

- We will make some generalizations so that the steps outlined here can be applicable to countless problems, which include the examples given in the previous slide.

University of Iowa (UI) Example

On March 15, 2002, the *Iowa City Press Citizen* carried an article about the intended 19% tuition increase to go into effect at the University of Iowa (UI) for the next academic year. Suppose you wish estimate the proportion of current UI students who are likely to quit school if tuition is raised that much. Rather than interviewing all 28,000+ students, we pick a simple random sample of $n = 50$ students and ask them if they'd be likely to quit school if tuition were raised by 19%.

Source: MC Ex 3.1

- Each student's response as an outcome of a Bernoulli random variable with success/failure corresponding to answering yes/no, respectively. [1]

- The total number of "yes"es can be modelled by

$$Y \sim \text{Binomial}(n, \theta)$$

where n is the number of UI students being interviewed and θ is the proportion in the entire population of UI students who would be likely to quit school if tuition is raised by 19%

[1] We assume a simple random sample, such that the responses from the individual students are independent.

The Likelihood

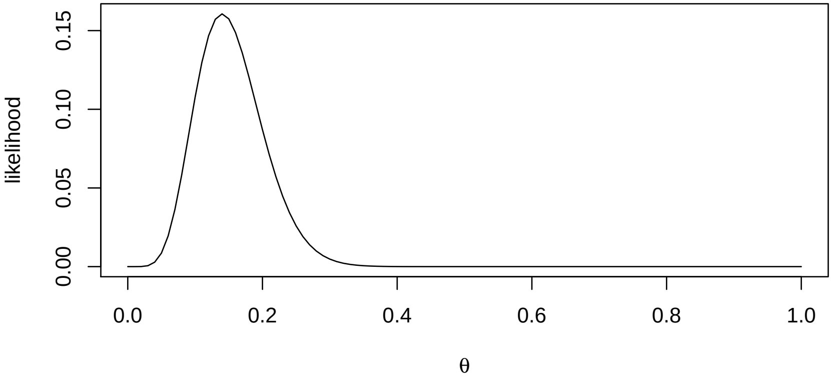
- Recall that the likelihood $p(y \mid \theta)$ is often written $\mathcal{L}(\theta \mid y)$ or $\mathcal{L}(\theta)$ to emphasize that θ is unknown and our data y is what is "given".

- All the relevant information about θ will be contained in our likelihood function which in this case is:

$$\begin{aligned}\mathcal{L}(\theta) &= \binom{50}{y}\theta^{y}(1-\theta)^{50-y}, \quad 0 \le \theta \le 1 \\ &\propto \theta^{y}(1-\theta)^{50-y}\end{aligned}$$

Note: Recall that the likelihood is only defined up to a constant of proportionality, hence, we can rescale the likelihood by dropping the $\binom{50}{y}$ (constant w.r.t θ).

The actual shape of the likelihood will depend on the random sample we obtain. For instance, if 7 of the 50 people sampled answer yes to our question, our likelihood would look as below. To keep in generic, let's say k of the sampled 50 answer yes.



The Prior

- To carry out Bayesian inference on θ, we need to assess our belief about θ *before* we observe the data from the survey.
- Last class we assumed a *discrete* prior distribution for θ, i.e. θ could only be 0.5, 0.6, and 0.9.
- Now we allow θ be any number between 0 and 1 and specify a *continuous* prior distribution.
- A convenient family of densities for modeling proportions is the beta distribution which has support $[0, 1]$ and two positive real-valued shape parameters: α, β.

Beta Distribution

The pdf of the beta distribution, for $0 \leq \theta \leq 1$, and shape *hyperparameters* $\alpha, \beta > 0$ is given by:

$$p(\theta|\alpha, \beta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}/\mathrm{B}(\alpha, \beta) \tag{1}$$

where $\mathrm{B}(\alpha, \beta)$ is the Beta function $= \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\Gamma(z)$ is the Gamma function$= \int_0^\infty x^{z-1}e^{-x}\, dx$.

Define: A *hyperparameter* is a parameter used in a prior model.

Note: For any positive integer n, $\Gamma(n) = (n-1)!$ *When* $\alpha = \beta = 1$ *we get the standard uniform* $\sim \mathcal{U}(0, 1)$*.*

- Using the same logic that we made for the likelihood, since we only need to calculate the posterior up to a constant of proportionality, we can remove any constants (w.r.t. θ) in the prior so that

$$\begin{aligned} p(\theta) &= \theta^{\alpha-1}(1-\theta)^{\beta-1}/\mathrm{B}(\alpha,\beta) \\ &\propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \end{aligned}$$

- Prior distributions that do not integrate to 1 are called *improper priors*.

- As we found in the previous lecture, this will be of no consequence since we will normalize the posterior as a final step.

- With little-to-no knowledge about university students, we might consider all values equally likely, i.e. a uniform prior.
- The uniform prior is an example of an *non-informative* prior.
- Non-informative priors generally represent the “state of ignorance” wherein very little is known *a priori*.
- Non-informative (or “vague”) priors have little impact on the posterior and allow the data to “speak for themselves”.
- Let’s use *a* and *b* as place holders for these hyperparameters to generalize the fundamental Beta-Binomial Bayesian model ...

Let's find our posterior $\propto$ prior × likelihoood ...

$$
\begin{aligned}
p(\theta \mid y) &\propto p(\theta)p(y \mid \theta) \\
&\propto \theta^{a-1}(1-\theta)^{b-1}/\mathrm{B}(a,b) \times \quad \text{(\textit{Beta} with hyperparameters } a \text{ and } b\text{)} \\
&\quad \binom{50}{k}\theta^{k}(1-\theta)^{50-k} \quad \text{(the \textit{Binomial} likelihood with } k \text{ yes's)} \\
&\propto \theta^{a-1}(1-\theta)^{b-1} \times \theta^{k}(1-\theta)^{50-k} \\
&\propto \theta^{k+a-1}(1-\theta)^{50-k+b-1}
\end{aligned}
$$

Hence the *Beta-Binomial* Bayesian model.

Recall that the "$\propto$" allows us to remove any constants with respect to θ from the above expression.

- The last line of the previous slide is the unnormalized posterior, i.e., it is *not* a proper probability distribution function.

- We will use the same trick from last lecture, namely we can multiply by the normalizing constant c that makes the posterior distribution with are under the curve equal to 1.

- But rather than solving for this, we need only recognize that this has the *functional form*, of a Beta distribution . . .

- To see this, compare the PDF of the Beta:

$$p(\theta|\alpha,\beta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}/\mathrm{B}(\alpha,\beta)$$

  with the our unnormalized posterior...

$$p(\theta \mid y) \propto \theta^{k+a-1}(1-\theta)^{50-k+b-1}$$

- Hence $\theta \mid y \sim \text{Beta}(\alpha = \qquad , \beta = \qquad)$ so that

$$p(\theta \mid y) = \theta^{\alpha-1}(1-\theta)^{\beta-1}/\mathrm{B}(\alpha,\beta)$$

- Beta distribution with shape parameters equal to

$$\alpha = \qquad\qquad \beta =$$

- To see this, compare the PDF of the Beta:

$$p(\theta|\alpha,\beta) = \theta^{\alpha-1}(1-\theta)^{\beta-1}/\mathrm{B}(\alpha,\beta)$$

with the our unnormalized posterior...

$$p(\theta \mid y) \propto \theta^{k+a-1}(1-\theta)^{50-k+b-1}$$

- Hence $\theta \mid y \sim$ Beta($\alpha = a + k, \beta = b + n$ -k) so that

$$p(\theta \mid y) = \theta^{\alpha-1}(1-\theta)^{\beta-1}/\mathrm{B}(\alpha,\beta)$$

- Beta distribution with shape parameters equal to

$$\alpha = a + k \qquad\qquad \beta = b + n \text{ -} k$$

To see why, note that the pdf of Beta dist must integrate to 1:

$$\begin{aligned} \int_0^1 \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\mathrm{B}(\alpha,\beta)} d\theta &= 1 \\ \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta &= 1 \\ \implies \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta &= \mathrm{B}(\alpha,\beta) \end{aligned}$$

Multiplying the unnormalized posterior by $1/\mathrm{B}(a+y, b+n-y)$ ...

$$\int_0^1 \frac{\theta^{a+k-1}(1-\theta)^{b+n-k-1}}{\mathrm{B}(a+y, b+n-k)} d\theta = \frac{\mathrm{B}(a+k, b+n-k)}{\mathrm{B}(a+k, b+n-k)} = 1$$

You can see a more formal proof to this result here

Conjugate prior

- Notice that the prior and posterior distribution have the same functional form as a Beta distribution.
- To put another way, both distributions can be written (up to a constant of proportionality) as $\theta^{xx}(1-\theta)^{yy}$
- The general form can be referred to as the *Beta kernel*[2] of the beta distribution.
- When the posterior and prior are of the same functional form this is known as *conjugacy*. More formally...

[2] the kernel of a probability density function (pdf) or probability mass function (pmf) is the form of the pdf or pmf in which any factors that are not functions of any of the variables in the domain are omitted

Conjugate prior

Conjugacy

If the posterior distributions $p(\theta \mid y)$ are in the same probability distribution family as the prior probability distribution $p(\theta)$, the prior and posterior are then called *conjugate distributions*, and the *prior* is called a *conjugate prior* for the likelihood function.

Note: By "same family", we mean that both the prior and the posterior have the same probability distribution but with differing parameters.

Conjugate prior

- In this example we have just shown that the *beta prior* is conjugate for the *binomial likelihood*, since the posterior distribution is a member of the same parametric family as the prior distribution.
- We will encounter other pairings in which a particular family of densities is conjugate for a particular likelihood family. In each case, the resulting posterior density will be in the same family as the prior.
- Conjugate priors are extremely convenient, however, for most complex real-world models, **no conjugate family exists**.

Beta-Binomial Distribution

The Beta distribution is a conjugate distribution of the binomial likelihood. That is, with $Y \mid \theta \sim \text{Binomial}(n, \theta)$ we have

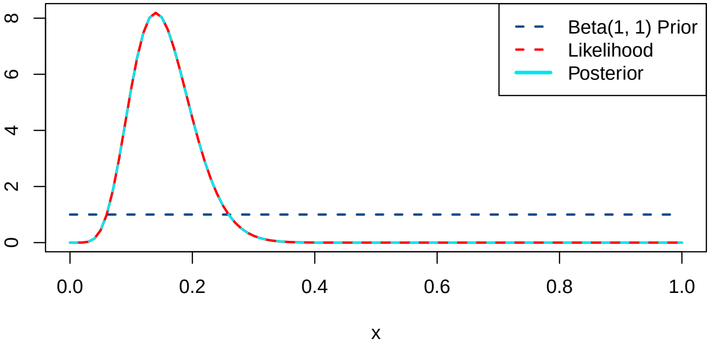
$$\theta \sim \text{Beta}(\alpha, \beta) \qquad \text{(Prior)}$$
$$\theta \mid (Y = y) \sim \text{Beta}(y + \alpha, n - y + \beta) \qquad \text{(Posterior)}$$

where k is the observed number of yes's in n independent Bernoulli trials.

- To demonstrate the generality of this, let's consider the uniform, i.e. Beta(1,1) prior, for θ and consider more "informative" priors shortly.
- We will suppose that 7 of the 50 people we sampled answered *yes* to the question if they are likely to quit school if tuition is raised by 19%.
- To specify this Beta-Binomial model we might write:

$$\begin{aligned} \theta &\sim \text{Beta}(1,1) \\ \theta \mid (Y = k) &\sim \text{Beta}(8,44) \end{aligned}$$

Summary of UI example



The following image was produced using “in-house” code but we will move to packages for doing these things.

Choosing the prior parameters

- Notice that the *non-informative* uniform prior did nothing in terms of our tug-of-war fight with the likelihood.

$$
\begin{aligned}
p(\theta \mid y) &\propto p(\theta)p(y \mid \theta) \\
&\propto p(\theta)\mathcal{L}(\theta) \\
&\propto \theta^{1-1}(1-\theta)^{1-1} \times \theta^{7}(1-\theta)^{50-7} \\
&\propto 1 \times \theta^{7}(1-\theta)^{50-7} \\
&\propto \theta^{7}(1-\theta)^{43}, \qquad \text{for } 0 \leq \theta \leq 1
\end{aligned}
$$

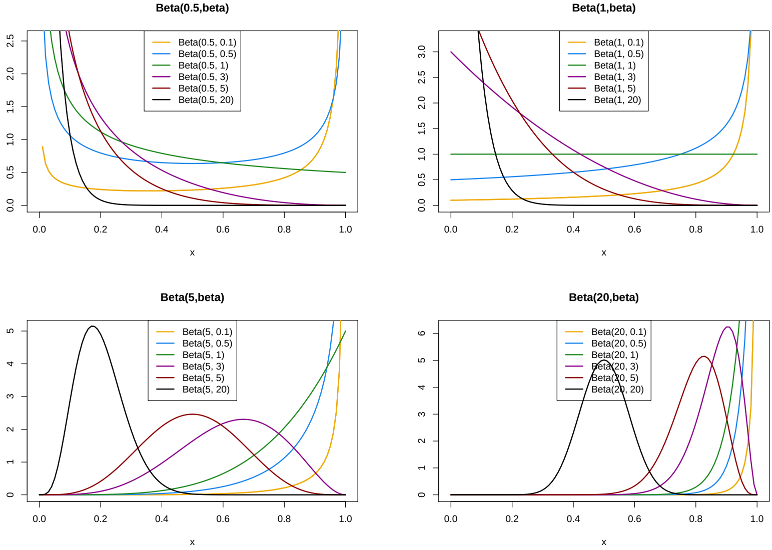
Note: If we want to plot the likelihood on the same scale as the prior and the posterior, we simply recognize that the functional from of the likelihood is also Beta with shape parameters $\alpha = 8$ and $\beta = 44$

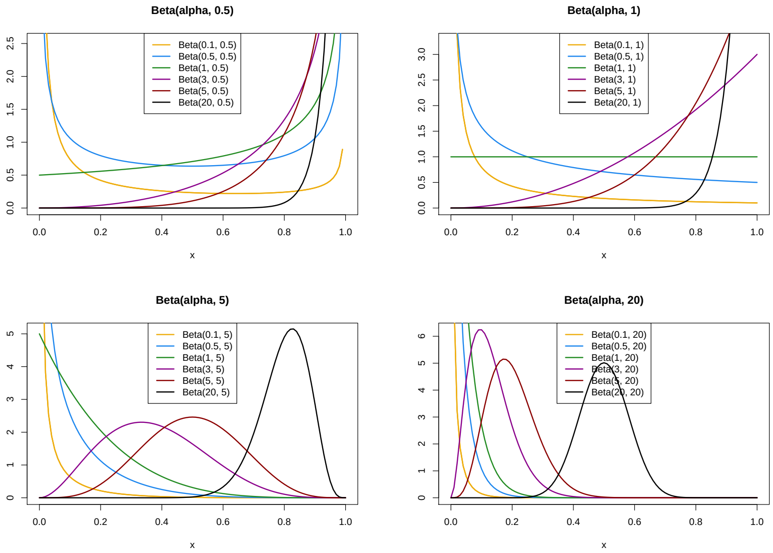
Choosing the prior parameters

- Lets consider an example of some *informative* priors to represent the instances where a person has knowledge or belief regarding the value of θ.

- To put another way, lets assume $\theta \sim \text{Beta}(\alpha, \beta)$ for values of the distribution parameters α and β other than 1.

- Before we set them, lets try to understand how the values of α and β affect our beta distribution.


Beta(0.5,beta)
Beta(0.5, 0.1)
Beta(0.5, 0.5)
Beta(0.5, 1)
Beta(0.5, 3)
Beta(0.5, 5)
Beta(0.5, 20)
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0
x
Beta(1,beta)
Beta(1, 0.1)
Beta(1, 0.5)
Beta(1, 1)
Beta(1, 3)
Beta(1, 5)
Beta(1, 20)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
x
Beta(5,beta)
Beta(5, 0.1)
Beta(5, 0.5)
Beta(5, 1)
Beta(5, 3)
Beta(5, 5)
Beta(5, 20)
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
x
Beta(20,beta)
Beta(20, 0.1)
Beta(20, 0.5)
Beta(20, 1)
Beta(20, 3)
Beta(20, 5)
Beta(20, 20)
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
x


Beta(alpha, 0.5)
Beta(0.1, 0.5)
Beta(0.5, 0.5)
Beta(1, 0.5)
Beta(3, 0.5)
Beta(5, 0.5)
Beta(20, 0.5)
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0
x
Beta(alpha, 1)
Beta(0.1, 1)
Beta(0.5, 1)
Beta(1, 1)
Beta(3, 1)
Beta(5, 1)
Beta(20, 1)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
x
Beta(alpha, 5)
Beta(0.1, 5)
Beta(0.5, 5)
Beta(1, 5)
Beta(3, 5)
Beta(5, 5)
Beta(20, 5)
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
x
Beta(alpha, 20)
Beta(0.1, 20)
Beta(0.5, 20)
Beta(1, 20)
Beta(3, 20)
Beta(5, 20)
Beta(20, 20)
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
x

Strategies for choosing α and β

MC[3] outlines several strategies for choosing the parameters of a beta distribution to express prior beliefs about θ:

1. Graph densities until you find one that matches your beliefs.
2. Note that a Beta(α, β) prior contains the information of a dataset with $\alpha - 1$ (resp. $\beta - 1$) successes (resp. failures).
3. Solve for the values of α and β that yield a desired mean/variance/"sample size".
4. Choose values of α and β that produce a prior probability interval that reflects your belief about θ.

[3] Cowles, M. K., 'Applied Bayesian statistics: with R and OpenBUGS examples', Vol. 98. Springer Science & Business Media, 2013.

Strategy 1: graphing

- There are a few patterns we can pick up from slide 27/28:
  - As α (resp. β) increases the distribution tends to move to the right (resp. left).
  - As α and β increase then the distribution begins to get narrower, i.e. the variance gets smaller.
  - If $\alpha = \beta$ then the distribution will peak over 0.5.
  - See this interactive website
- However, the task of plot varying values of α and β until we see one that matches our beliefs is a somewhat tedious task.

Strategy 2: prior sample size

- Let's compare the likelihood with the prior to see where this come from. Recall the binomial likelihood:

$$\mathcal{L}(\theta) \propto \theta^k (1-\theta)^{n-k}$$

- and contrast that with the beta prior:

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

- If the likelihood contains all the information about our data, namely that there were k success and $n-k$ failures, then the beta prior contains the same information as a dataset with $\alpha - 1$ successes and $\beta - 1$ failures.

Strategy 2

- Returning to our UI example, our data has 50 observations (7 "yes", 43 "no"s).

- To match this strength in terms of our prior, we would need an equivalent prior sample size of $50 = \alpha - 1 + \beta - 1 = \alpha + \beta - 2$

- More generally, we might say:
  - $\alpha + \beta - 2$ is roughly our *prior sample size*
  - $\alpha - 1$ is roughly the prior number of successes, and
  - $\beta - 1$ is roughly the prior number of failures

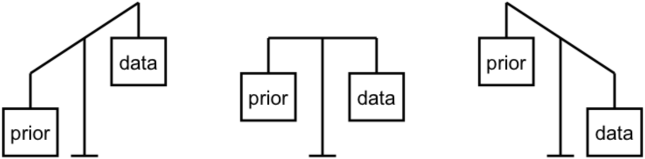
Strategy 2

- Keeping with the tug of war analogy, the strength of the respective likelihood and prior would determine how much “pull” one would have in a game of tug of war.

- Opponents with equal strength would arrive at a posterior somewhere in the middle, whereas a weaker opponent would get dominated by a stronger one.

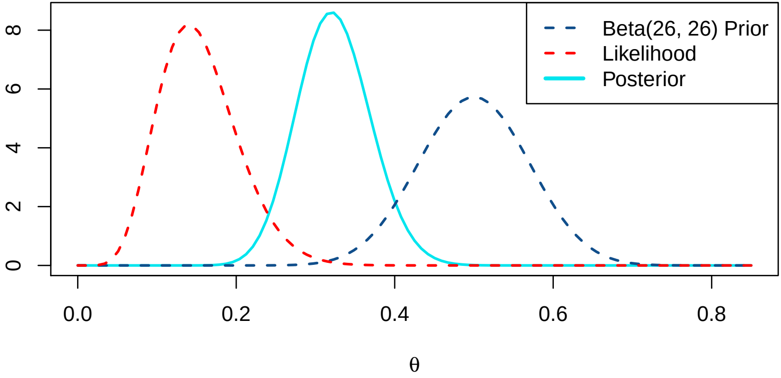
If $n \ll \alpha + \beta - 2$ the posterior will be influenced heavily by our prior belief.

If $n \gg \alpha + \beta - 2$ the posterior will be influenced heavily by the data.



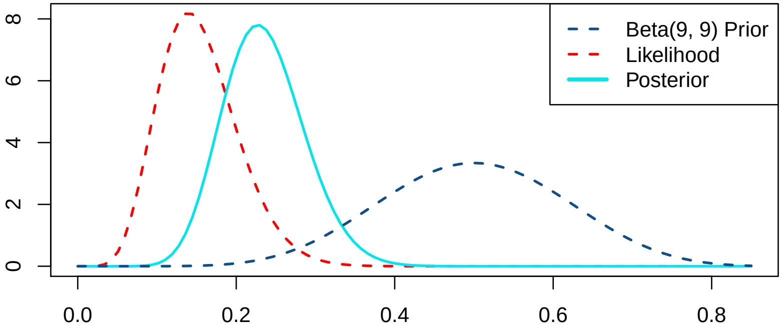
UI example – tug of war 1

Equal strength opponents:
50 samples vs 50 “prior samples”



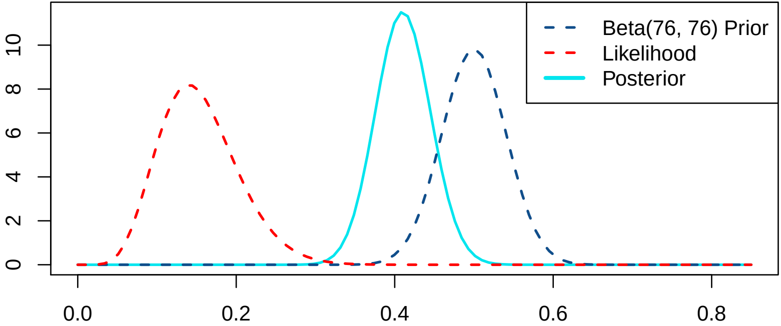
UI example – tug of war 2

Unequal strength opponents:
likelihood is “stronger” than the prior
50 samples vs. 16 “prior samples”



UI example – tug of war 3

Unequal strength opponents:
likelihood is “weaker” than the prior
50 samples vs. 150 “prior samples”



Strategy 3: solve for desired sample statistics

- To make use of strategy 3, lets take a look at the mean μ and standard deviation σ of the beta distribution:

$$\mu = \frac{\alpha}{\alpha + \beta} \tag{2}$$

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} \tag{3}$$

Rearranging this formulae so that the hyperparameters are in terms of μ and σ will provide a reasonable way for defining our prior

$$\alpha = \left(\frac{1-\mu}{\sigma^2} - \frac{1}{\mu} \right) \mu^2 \tag{4}$$

$$\beta = \alpha \left(\frac{1}{\mu} - 1 \right) \tag{5}$$

For instance, if I believe θ= 0.5, but I'm not particularly certain I may specify a standard deviation of 0.1. Plugging $\mu = 0.5$ and $\sigma = 0.1$ in (4) and (5) gives me

$$\alpha = \beta = 12$$

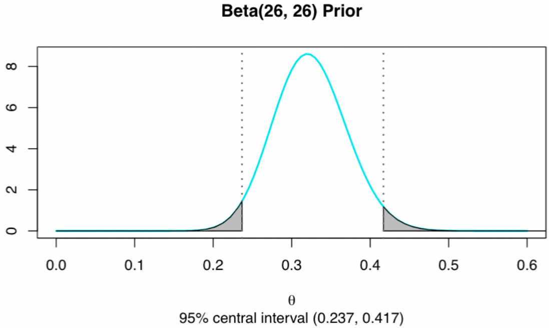
Strategy 4: probability intervals

- One way of summarizing a prior probability distribution is in terms of an interval that traps a specified proportion of the prior probability mass.

- The interval $[l, u]$ is a $100(1-\alpha)\%$ central prior probability interval for θ if the prior mass on values less than l is $\alpha/2$ and on values greater than u is also $\alpha/2$; the remaining $1-\alpha$ prior probability is on the interval $[l, u]$.

Strategy 4: probability intervals

Akin to Frequentist Confidence Intervals, the most commonly considered *prior probability intervals* are at 95% ($\alpha = 0.05$).



UI example– choosing prior parameters

University of Iowa (UI) with ISU data

Suppose we read about a survey taken on 50 *Iowa State University (ISU)* students of whom 10 said they would quit school; 40 said they would not.

Let's apply some of these strategies to the UI example.

UI example– choosing prior parameters

Strategy 2

- Applying Strategy 2, we might suppose a Beta(11, 41) prior.
- However, we need to be very cautious here because the sample on which the prior distribution is to be based was not drawn from the same population (UI students) in which we are interested and from which we will draw our sample.
- So we may want to make use of this information without giving it as much weight as 50 observations from the UI population.

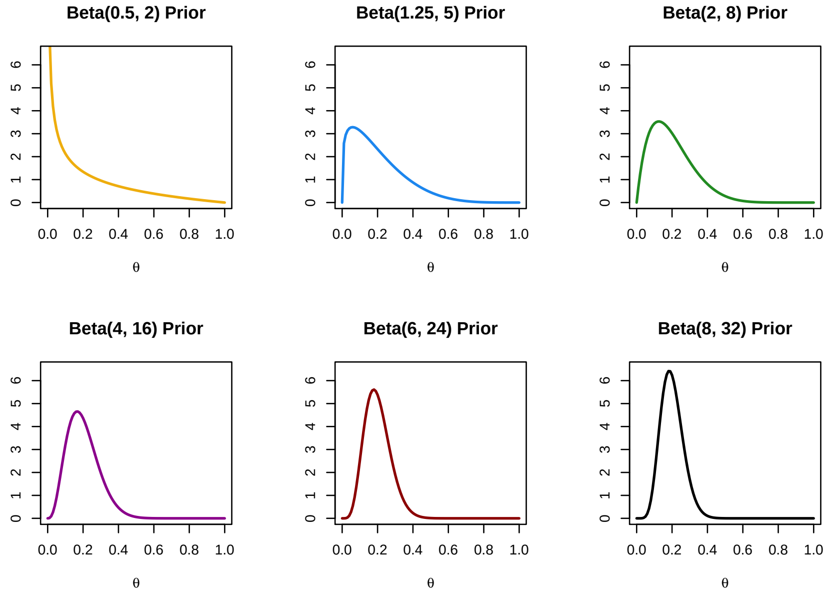
UI example – choosing prior parameters

Strategy 1/3

- **Strategy 3:** set the prior mean equal to the sample proportion from ISU, i.e. $10/50 = 0.2$ with the condition that the equivalent "prior sample size" $= \alpha - 1 + \beta - 1$ should be less than 50.

- Mathematically, we can see that in order for the beta prior to have a mean $\mu = \dfrac{\alpha}{\alpha + \beta} = \dfrac{10}{50} = 0.2 \implies \alpha = \frac{1}{4} \times \beta$.

- **Strategy 1:** we can look at the graphs of several different beta distributions with mean 0.2 and prior sample sizes less than 50 and seek one that matches our beliefs.


Beta(0.5, 2) Prior
Beta(1.25, 5) Prior
Beta(2, 8) Prior
Beta(4, 16) Prior
Beta(6, 24) Prior
Beta(8, 32) Prior
0
1
2
3
4
5
6
0.0
0.2
0.4
0.6
0.8
1.0
θ

- All of the distributions on the previous slide have a mean of 0.2.
- The belief expressed in the Beta(2,8) prior, for example, is as strong as if we had seen a previous sample from the population of interest (UI students) of size 8 with 1 successes and 7 failures.
- The Beta(8, 32) prior, for example, is as strong as a sample of size 38 with 7 successes and 31 failures.

Note: the smaller (resp. larger) the sum $(\alpha + \beta)$ the larger (resp. smaller) the variance.

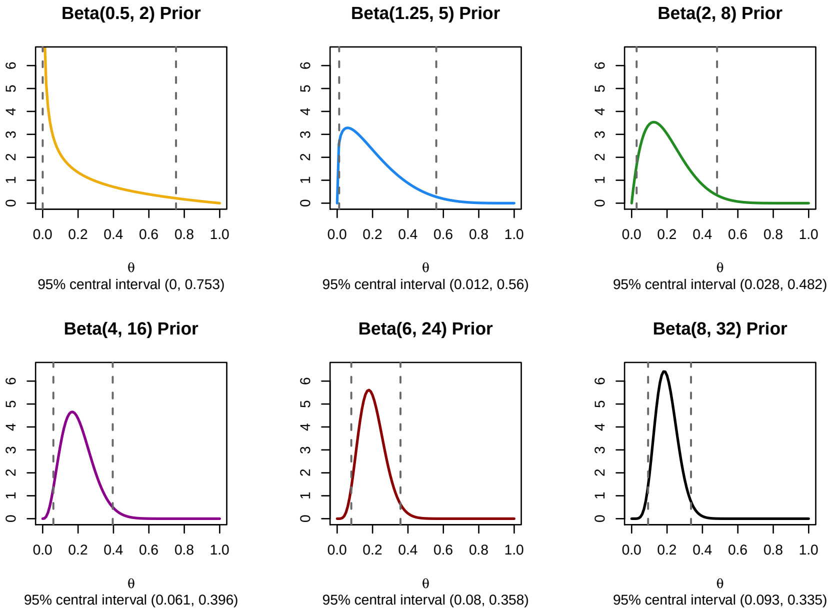
Strategy 4

- We can investigate the corresponding 95% central intervals for these priors to see if they correspond to our prior belief.

- In R, we would find the values of $[l, u]$ (l/u called `lower`/`upper` below) using the following code:

```
> lower = qbeta(0.025, shape1=a, shape2=b)
> upper = qbeta(0.025, a, b, lower.tail = F)
```

- Hence $l =$ the 0.025 *quantile* of the probability distribution, and $u = 0.975$ *quantile*.



Beta(0.5, 2) Prior

95% central interval (0, 0.753)

Beta(1.25, 5) Prior

95% central interval (0.012, 0.56)

Beta(2, 8) Prior

95% central interval (0.028, 0.482)

Beta(4, 16) Prior

95% central interval (0.061, 0.396)

Beta(6, 24) Prior

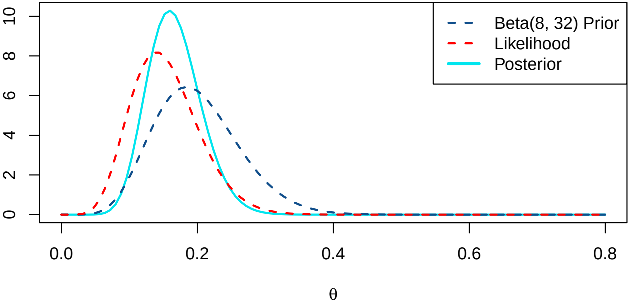
95% central interval (0.08, 0.358)

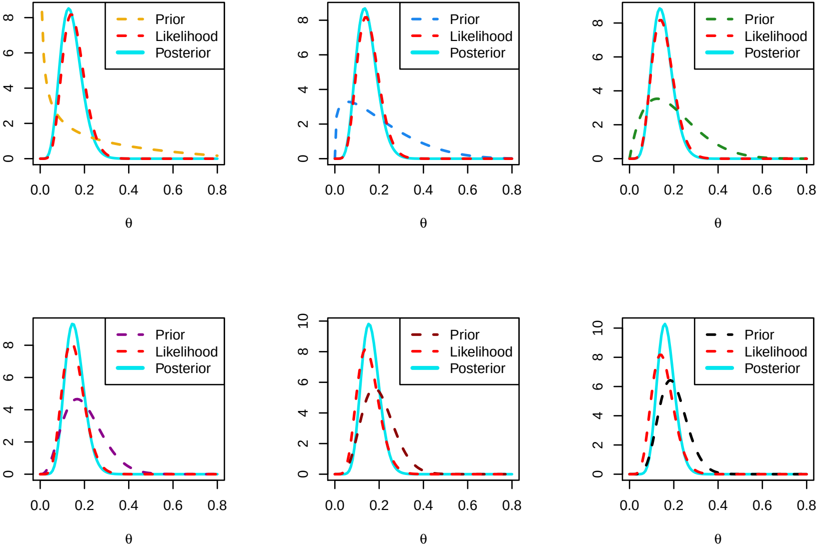
Beta(8, 32) Prior

95% central interval (0.093, 0.335)

UI example using a Beta(8, 32) prior

- A stronger our prior belief, i.e. a Beta distribution with larger values of $\alpha + \beta$, will have greater impact on the posterior distribution. Focussing on the last plot ...




Prior
Likelihood
Posterior
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
θ

Plotting the Prior, Likelihood, and Posterior

- To have the prior, likelihood, and posterior appear on the same scale, all of them needed to be *normalized*.

- The prior and posterior are already proper, the likelihood is not.
- We can apply the same trick as used for the posterior, namely:

$$\mathcal{L}(\theta) \propto \theta^{y}(1-\theta)^{n-y} \tag{6}$$

- This has the functional form of a Beta distribution with

$$\alpha = \qquad\qquad\qquad \beta =$$

Plotting the Prior, Likelihood, and Posterior

- To have the prior, likelihood, and posterior appear on the same scale, all of them needed to be *normalized*.

- The prior and posterior are already proper, the likelihood is not.
- We can apply the same trick as used for the posterior, namely:

$$\mathcal{L}(\theta) \propto \theta^{y}(1-\theta)^{n-y} \tag{6}$$

- This has the functional form of a Beta distribution with

$$\alpha = y + 1 \qquad \beta = n - y + 1$$

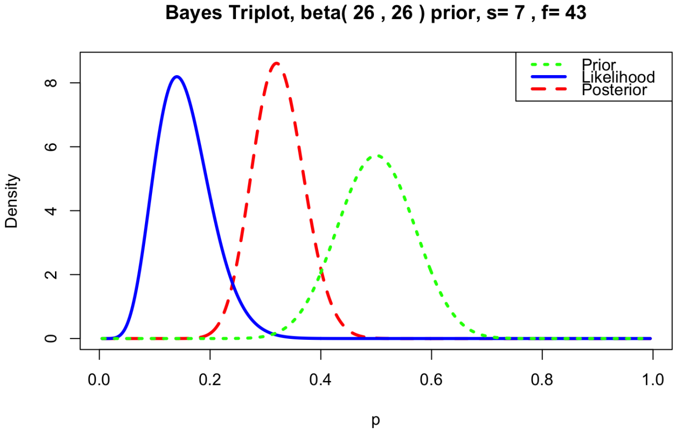
Packages for plotting

- Plotting these three curve can easily be done using base plots in R (you will see this code in lab this week).

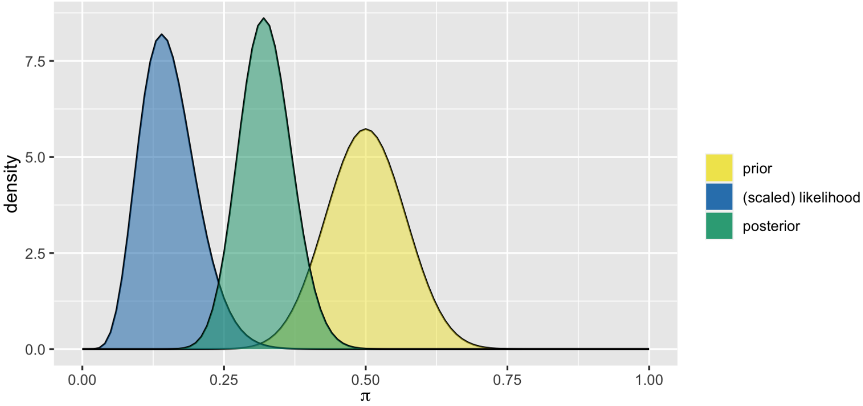
- However, you will can instead use the easy to use functions available in multiple packages for Bayesian analysis.

- For instance: the `triplot` function from the **LearnBayes** package, or the `plot_beta_binomial` function from the **bayesrules** package.

```
library(LearnBayes)
a = 26; b=26  # prior parameter values
y = 7; n=50   # no. of success; no. of Bernoulli trials
triplot(prior=c(a, b), data=c(y,n-y))
```



```
library(bayesrules)
plot_beta_binomial(alpha = 26, beta = 26, y = 7, n = 50)
```



Note: BayesRules! uses π instead of θ

Comments

- The posterior density always is in some sense a compromise between the prior density and the likelihood. It should (theoretically) have less variance than the prior.

- This translates to the posterior being more concentrated (i.e. "pointy") than either the prior or the likelihood.

- As the posterior combines the prior belief with the additional information from the data, it sensibly is more precise than either one of the two sources alone.