

# One Parameter Model: Estimation and Inference for Proportions

Edps 590BAY

Carolyn J. Anderson

Department of Educational Psychology



©Board of Trustees, University of Illinois

Fall 2019

# I Overview

- ▶ Pictures of how Bayes Theorem works
- ▶ Binomial distribution (likelihood, model for data)
- ▶ Beta distribution (prior and posterior)
- ▶ Explore combinations of beta and binomials
- ▶ Analytic results for a estimation and inference of a proportion
- ▶ Posterior prediction
- ▶ Up-dating given new information
- ▶ Practice
- ▶ Grid method
- ▶ Comparing Two Proportions & a little Monte Carlo
- ▶ Follow-up practice

Depending on the book that you select for this course, read either Gelman et al. p29–30 or Kruschke Chapters 5 & 6.

# I Pictures of How Bayes Theorem Works

The posterior distribution of  $\theta$  is a compromise between the prior and the posterior. Recall,

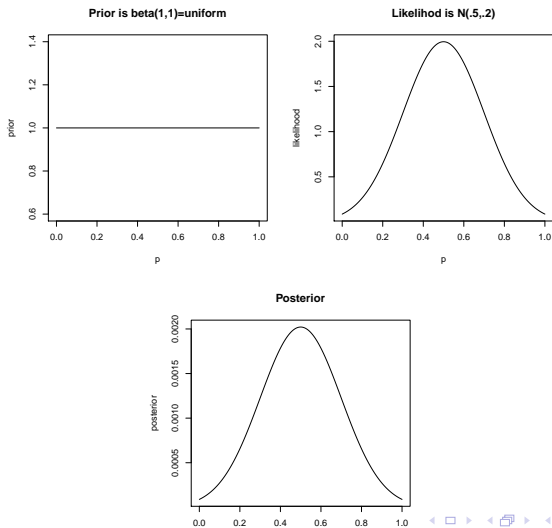
$$\begin{aligned} p(\theta|y) &= \frac{p(y|\theta)p(\theta)}{p(y)} \\ &\propto p(y|\theta)p(\theta) \end{aligned}$$

where  $p(y)$

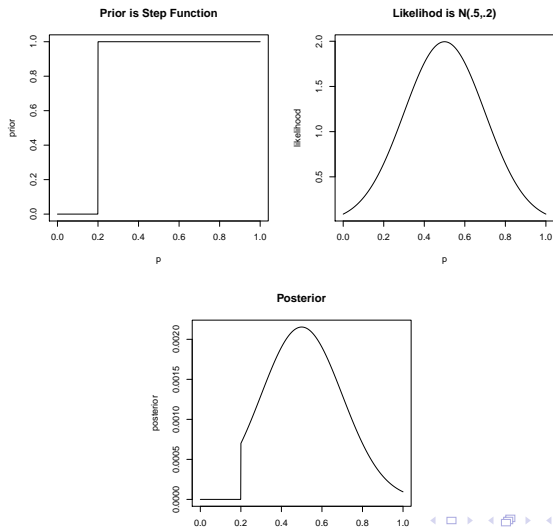
- ▶ ensures probability sums to 1 & is constant for given problem.
- ▶ “average” of numerator.
- ▶ For discrete  $y$ :  $p(y) = \sum_{\theta \in \Theta} p(y|\theta)p(\theta)$
- ▶ For continuous  $y$ :  $p(y) = \int_{\theta} p(y|\theta)p(\theta)d(\theta)$

Where  $y$  = number of “successes” out of  $n$  tries and  $\theta$  = the unknown probability of success.

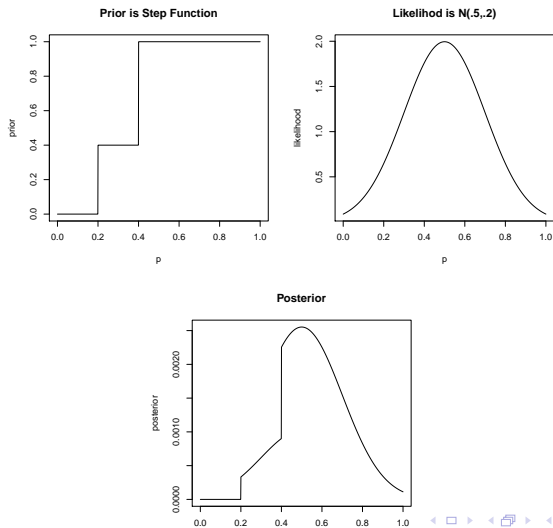
# I Example 1: No Prior Information



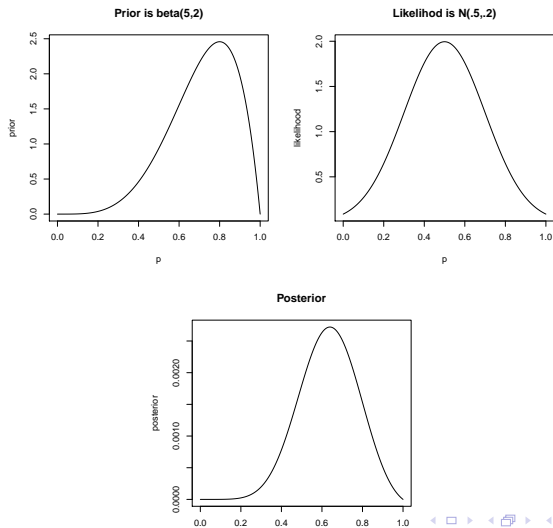
# I Example 2: Prior is Step Function



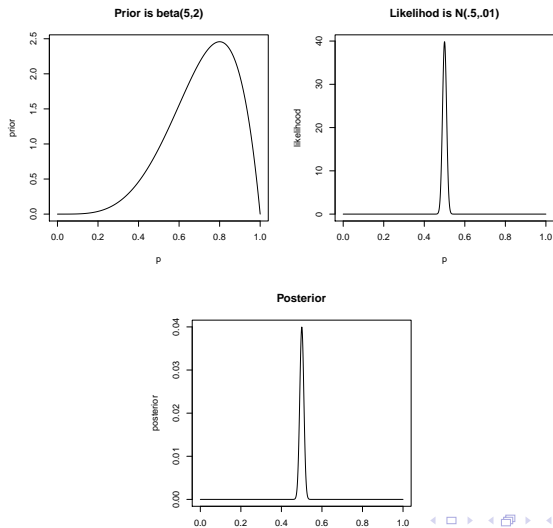
# I Example 3: Prior is a 2 Step Function



# I Example 4: Prior is Smooth Function



# I Example 5: Lots of data





# I Impact of prior and likelihood on posterior

Impact of prior and likelihood (data) on the posterior:

- ▶ Example 1: When you have no prior information (“ignorance”), the posterior distribution depends only on sampling model (i.e., the likelihood, the data)
- ▶ Example 2: With a **step function**, the posterior when prior value of the density equals 0, but once prior density equals 1, the posterior is the same as likelihood.
- ▶ Example 3: With a **2 step function**, this is like the one step function, except for the 2nd step where the prior density is non-zero, the posterior between these corresponding  $y$ s is linear in shape with non-zero slope.
- ▶ Example 4: With a **smooth prior and likelihood**, the posterior is a “compromise” between them. For example, the mode of the posterior is between the mode between the prior and likelihood.
- ▶ Example 5: With **lots of data**, the posterior is mostly influenced by the likelihood.

# I The Binomial

For proportions, a natural probability mass is the Binomial. The **Binomial will be our likelihood** for the probability of a dichotomous variable; that is, our sample model.

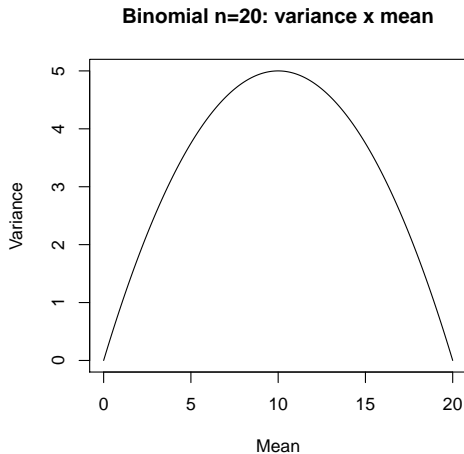
Let  $y =$  number of “successes” out of  $n$  independent trials where the probability of a success is  $\theta$ . The Binomial is

$$\begin{aligned} p(y|\theta) &= \binom{n}{y} \theta^y (1 - \theta)^{n-y} \\ &= \frac{n!}{y!(n-y)!} \theta^y (1 - \theta)^{n-y} \quad \text{for } y = 0, 1, \dots, n \end{aligned}$$

where

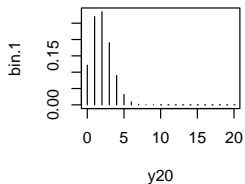
- ▶ Mean  $\mu = n\theta$
- ▶ Standard deviation  $\sigma = \sqrt{n\theta(1 - \theta)}$

# I The Binomial: Variance as function of mean

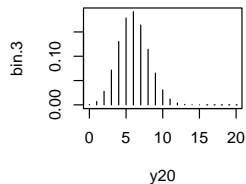


# I The Binomial varying $\theta$

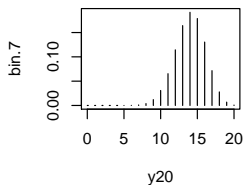
**n=20, p=.1**



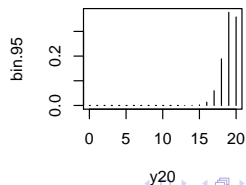
**n=20, p=.3**



**n=20, p=.7**

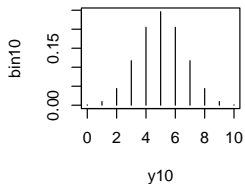


**n=20, p=.95**

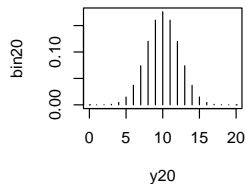


# I The Binomial varying $n$

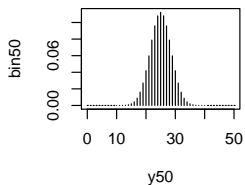
**$n=10, p=.5$**



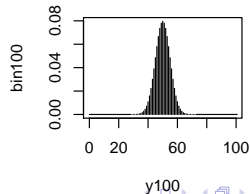
**$n=20, p=.5$**



**$n=50, p=.5$**



**$n=100, p=.5$**



# I Prior: The Beta Distribution

Why?

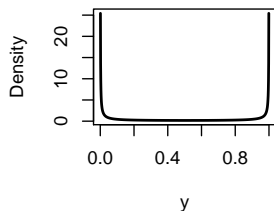
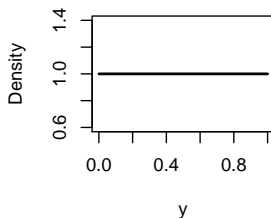
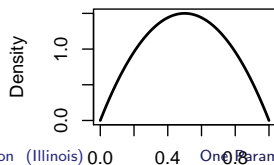
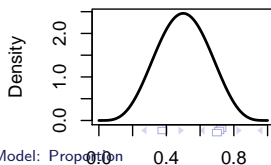
- ▶ It has the same possible values as a probability:

$$0 \leq x \leq 1$$

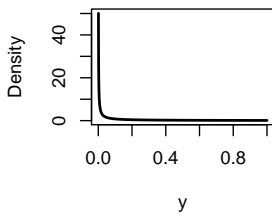
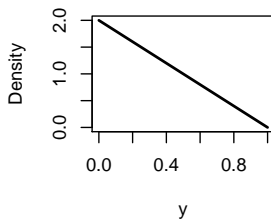
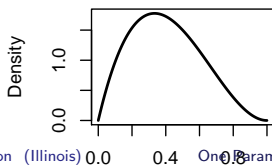
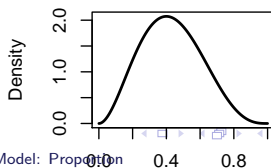
- ▶ The density function can take on many different shapes.
- ▶ With a Beta prior and Binomial Likelihood, the posterior will be a Beta distribution.

Definition: A Conjugate prior distribution is one where the product of the likelihood and prior (i.e.,  $p(y|\theta) \times p(\theta)$ ) yields a posterior distribution for  $\theta$  that has the same form as the prior.

# I Examples of Beta Distribution: $a = b$

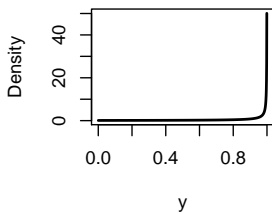
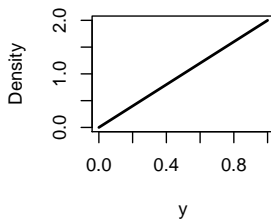
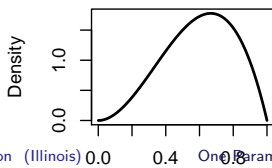
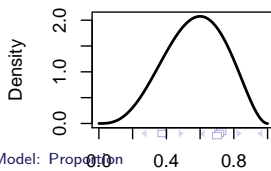
**Beta(0.1,0.1)****Beta(1,1)=Uniform(0,1)****Beta(2,2)****Beta(5,5)**

# I Beta with Right Skew: $a < b$

**Beta( 0.1 , 1 )****Beta( 1 , 2 )****Beta( 2 , 3 )****Beta( 3 , 4 )**



# I Beta with Left Skew: $a > b$

**Beta( 1 , 0.1 )****Beta( 2 , 1 )****Beta( 3 , 2 )****Beta( 4 , 3 )**

# I Facts about the Beta Distribution

- ▶ It depends on two parameters,  $a$  and  $b$ , where  $a$  is *like* the number of successes and  $b$  *like* the number of failures.
- ▶ Summary from figures:
  - ▶  $a = b \rightarrow$  symmetric.
  - ▶  $a > b \rightarrow$  left skew.
  - ▶  $b < a \rightarrow$  right skew.
  - ▶ As  $a$  and  $b$  get larger, more centered around the middle.
  - ▶ What happens when  $a < 1$  and  $b < 1$  or  $a < 1$  and  $b > 1$ ?  
Run function from web-site.

$$\mu = \frac{a}{b+a}$$

$$\text{mode} = \omega = \frac{(a-1)}{(a+b-2)} \text{ for } a, b > 1$$

$$\text{"concentration"} = \kappa = a + b$$

# I More notes on The Beta Distribution

Useful facts for Beta Distribution:

$$\mu = \frac{a}{b+a}$$

$$\text{mode} = \omega = \frac{(a-1)}{(a+b-2)} \text{ for } a > 1 \text{ and } b > 1$$

$$\text{median} \approx \frac{a-1/3}{a+b-2/3} \text{ for } a > 1 \text{ and } b > 1$$

$$\sigma^2 = \frac{ab}{(a+b)^2(a+b+1)} \text{ for } a \text{ and } b > 1$$

# I The Beta Distribution (continued)

The Beta distribution is

$$\text{Beta}(\theta|a, b) = p(\theta|a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{(a-1)}(1-\theta)^{(b-1)}$$

where

- ▶  $\Gamma()$  is the “gamma function”.
- ▶  $\Gamma(a) = \int_0^\infty t^{(a-1)} \exp(-t) dt$
- ▶  $\Gamma(a+b)/(\Gamma(a)\Gamma(b))$  is like the binomial coefficient for the binomial distribution.

Recall the Binomial distribution is

$$p(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

# I The Beta-Binomial

$$\begin{aligned}
 p(\theta|y, n) &= ((\text{Binomial likelihood}) \times) \text{Beta distributon} \\
 &= p(y|\theta)p(\theta) \\
 &= \binom{n}{y} \theta^y (1 - \theta)^{n-y} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{(a-1)} (1 - \theta)^{(b-1)} \\
 &= \frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)} \theta^{(y+a-1)} (1 - \theta)^{n+b-y-1} \\
 &\rightarrow \text{Beta}((a+y), (b+n-y)) \\
 &= \text{Beta}(a^*, b^*)
 \end{aligned}$$

where  $a^* = a + y - 1$  and  $b^* = b + n - y - 1$ .

## I Some Pictures of beta binomials

Play with R function that I wrote: `beta.binomial`

- ▶ Download from web-site (file name is “function\_beta\_binomial.txt”)
- ▶ Copy and run in console
- ▶ Use: `beta.binomial(y,n,a,b)`
- ▶ Output: figure of prior ( $p(\theta) = \text{beta}(a, b)$ ), likelihood  $p(y|\theta)$ , and the posterior  $p(\theta|y)$

OR

- ▶ Download and install package “LearnBayes”
- ▶ Use the command `triplot(prior, data, where=“_____”)`
  - ▶ `prior` = vector with scale parameters of beta
  - ▶ `data` = vector with  $y$  and  $n - y$ .
- ▶ For example

*`triplot(c(3,2),c(7,3),where=“topleft”)`*

# I Mean of the posterior as a compromise

The mean of the posterior equals

$$\begin{aligned}
 E(\theta|y) &= \frac{y + a}{n + a + b} \\
 &= \frac{y}{n + a + b} + \frac{a}{n + a + b} \\
 &= \left(\frac{n}{n}\right) \frac{y}{n + a + b} + \left(\frac{a + b}{a + b}\right) \frac{a}{n + a + b} \\
 &= \underbrace{\frac{y}{n}}_{\text{data}} \times \underbrace{\frac{n}{n + a + b}}_{\text{weight}} + \underbrace{\frac{a}{a + b}}_{\text{prior}} \times \underbrace{\frac{a + b}{n + a + b}}_{\text{weight}} \\
 &= w \frac{y}{n} + (1 - w) \frac{a}{a + b}
 \end{aligned}$$

“Shrinkage” is a property of Bayesian estimates. In the following example,  $E(\theta) = 0.5$ ,  $y/n = .8236$ , and  $E(\theta|y) \approx .7895$ .

## I Example: Heights of US Presidents

Data are from web-site [en.wikipedia.org/wiki/Heights\\_of\\_presidents\\_and\\_presidential\\_candidates\\_of\\_the\\_United\\_States](http://en.wikipedia.org/wiki/Heights_of_presidents_and_presidential_candidates_of_the_United_States)

- ▶  $y = 14$ , the number times where the taller candidate won (if they were tied, I counted it as the taller winning) covering 1928-1996.
- ▶  $n = 17$ , the number of elections (note:  $14/17 = .8236$ )
- ▶ We'll go with uninformative priors; that is,  $\text{Beta}(1, 1) = \text{Uniform}(0, 1)$
- ▶ The model could be stated

$$\text{likelihood} \quad y \sim \text{Binomial}(\theta, n)$$

$$\text{prior} \quad \theta \sim \text{Beta}(1, 1)$$

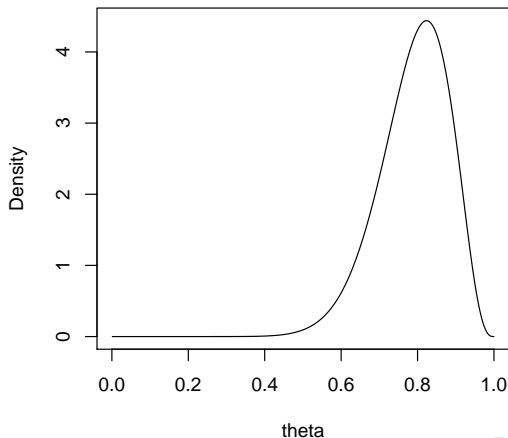
$$\text{posterior} \quad \theta|y \sim \text{Beta}(15, 4)$$

which gives us  $\text{mode} = .8235$ ,  $\text{mean} = .7895$ .



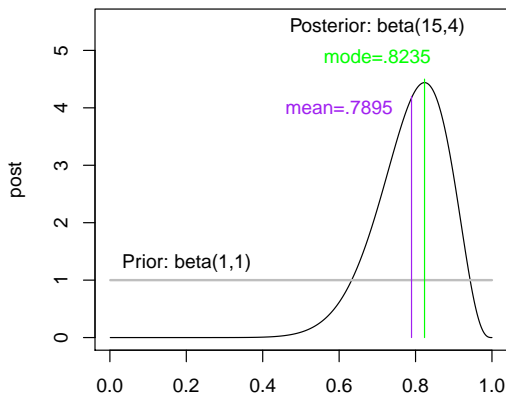
# I Posterior Distribution of Heights of US Presidents

Posterior:  $\text{Beta}(15,4)$



# I Posterior Distribution of Heights of US Presidents

## Posterior Distribution for Taller Candidate



# I Credible Intervals

- ▶ **Frequentist:** Confidence interval

$$Pr \{L(y) < \theta < U(y)|\theta\} = .95$$

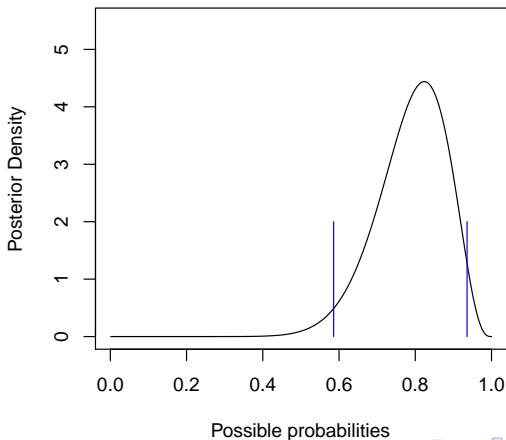
- ▶ Once we have data, the probability that the interval include  $\theta$  is 0 or 1.
- ▶ If we repeat process many times, then about 95% intervals constructed in the matter would include the true value of  $\theta$ .
- ▶ **Bayesian:** Credible interval

$$Pr(L(y) < \theta < U(y)|y) = .95$$

- ▶ There is a 95% chance that the interval from  $L(y)$  to  $U(y)$  includes the true value of  $\theta$ .
- ▶ This is conditioned on data (not unknown  $\theta$ ).

# I Credible Intervals

**95% Credible Intervals: (.59, .94)**



# I Computing Credible Intervals

```
# For 95% credible interval for taller  
theta.95 ← c(.025,.975)  
ci ← qbeta(theta.95,15,4)  
ci  
0.5858 .9359
```

# I High Density Intervals

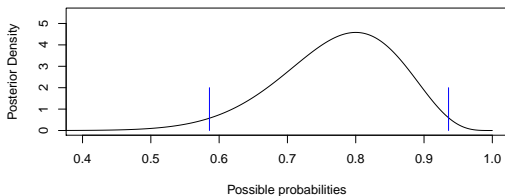
- ▶ The highest of the credible intervals are different.
- ▶ High density intervals are shortest ones with probability  $(1 - \alpha)$ .
- ▶ These have heights that are the same for both the left and right end point.
- ▶ They are harder to compute so we'll use package "HDInterval".

# I Finding High Density Intervals

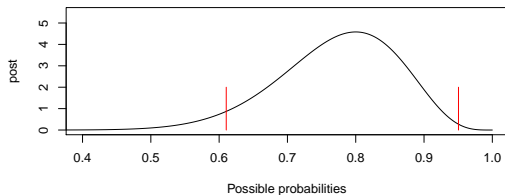
```
library(HDInterval)
inbeta ← rbeta(1E5,15,4)
hdi.p ← hdi(inbeta,credMass=.95)
hdi.p
      lower      upper
0.5992848 0.9297222
attr(,"credMass")
```

# I Compare HDI and Credible

95% Credible Interval: (.58, .94)



95% Highest Density Interval: (.60, .93)





## I Predicting New Observation

This include uncertainty about  $\theta$  and inherent randomness in sampling.

$$\begin{aligned} p(\tilde{y} = 1|y_1, \dots, y_n) &= \int_0^1 Pr(\tilde{y} = 1|\theta, y_1, \dots, y_n)Pr(\theta|y_1, \dots, y_n)d\theta \\ &= \int_0^1 Pr(\tilde{y} = 1|\theta)Pr(\theta|y_1, \dots, y_n)d\theta \\ &= \int_0^1 \theta Pr(\theta|y_1, \dots, y_n)d\theta \\ &\equiv E(\theta|y_1, \dots, y_n) \\ &= a^*/(a^* + b^*) \end{aligned}$$

Where  $a^*$  and  $b^*$  are from the posterior distribution of  $\theta$  given data. For our example where the posterior is Beta(15, 4) so  $Pr(\tilde{y}) = 15/(15 + 4) = 15/19 = .7895$

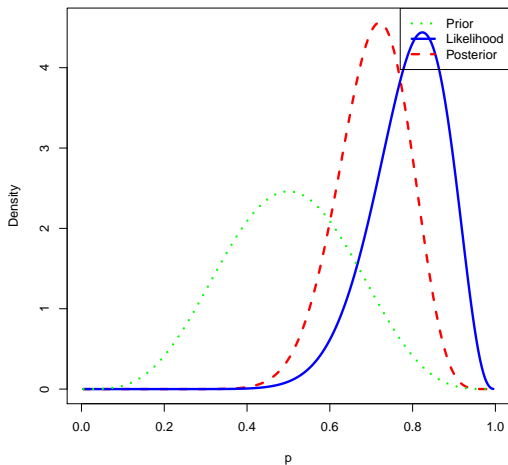
## I Using a prior Knowledge

Suppose I think that the probability of taller wining is  $1/2$ , and I have a fairly strong belief I could choose a beta

- ▶  $\theta = .5$ .
- ▶ Choose larger values for  $a$  and  $b$
- ▶ For our example, if we could use Beta(5,5).
- ▶ The posterior distribution would be Beta( $y + 5, n - y + 5$ ).
- ▶ For our example, Beta( $(14 + 5), (17 - 14 + 5)$ ) = Beta(20, 7)
- ▶ Mean of the posterior would be  $20/(20 + 7) = 20/27 = .7407$ .
- ▶ Mode of the posterior would be  
mode =  $(20 - 1)/(27 - 2) = .7600$ .

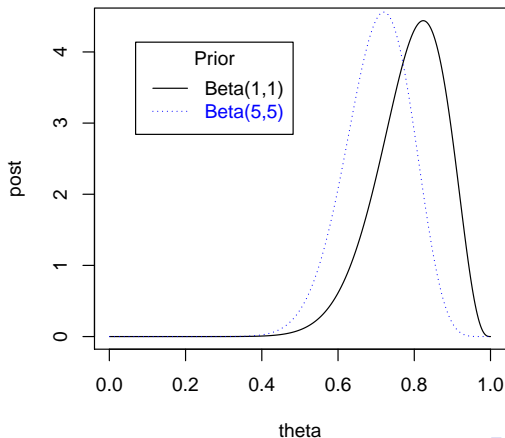
# I Using Informative Prior

Bayes Triplot, beta( 5 , 5 ) prior, s= 14 , f= 3



# I Different Posteriors using Informative Prior

Posterior: different priors



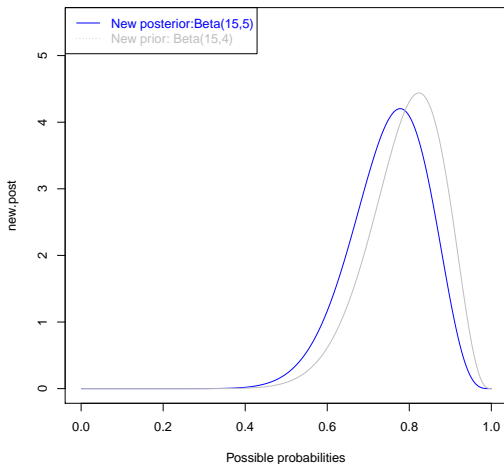
## I Up-dating and Using New Information

I only used elections up through 1996. How would our posterior change?

- ▶ Our posterior based on data is used as our prior to figure out what the new posterior would be.
- ▶ In 2000, the taller candidate lost: Bush vs Gore and Gore was taller.
- ▶ Current posterior is  $\text{Beta}(15, 4)$ , our newer posterior is  $\text{Beta}(15 + 0, 4 + 1) = \text{Beta}(15, 5)$
- ▶ New mean  $(15/20) = .75$   
new mode  $= (15 - 1)/(15 + 5 - 2) = .78$   
new median  $\approx (15 - 1/3)/(20 - 2/3) = .76$ .
- ▶ Data values are **Exchangeable**. The subscripts are just considered labels with no meaning (i.e., order doesn't matter).

# I Adding Another Observation

Posterior Distribution for Taller 1928–2000



## I Using Even More Information

Lets add elections 2000, 2004, 2008 and 2012 to our original data set:

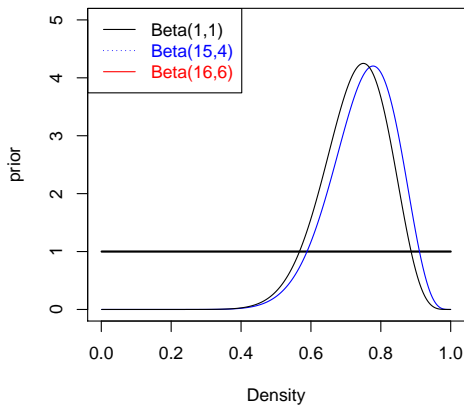
Year	Winner	cm	Loser	cm
2012	Obama	185	Romney	187
2008	Obama	185	McCain	175
2004	GWBush	182	Kerry	193
2000	GWBush	182	Gore	185

So started with  $y = 14$  and  $n = 17$  and a uniform prior, which gave us a posterior  $\text{Beta}(15, 4)$ . We use this as our prior. New data has 1 case where taller won and 3 cases where the shorter won.

Our new posterior is  $\text{Beta}(15 + 1, 4 + 3)$  with a mean of  $16/(16 + 7) = .6957$  and a mode of  $(16 - 1)/(16 + 7 - 2) = .7143$ ,

# I Learning with More Data

## Up-dating and Learning





# I Model Checking

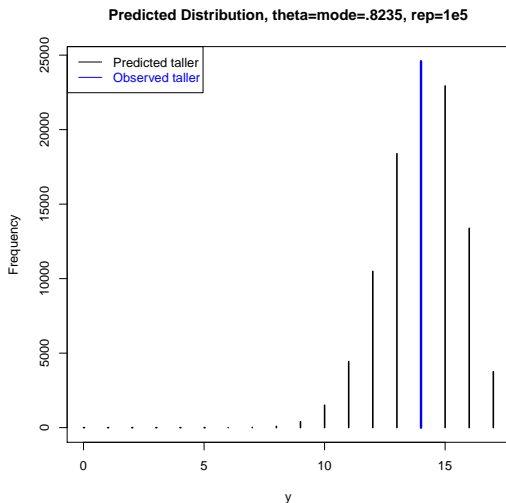
We'll check our original model:  $y = 14$ ,  $n - 7 = 3$ ,  $n = 17$ , uniform priors with  $\text{Beta}(15, 4)$  are posterior distribution for  $\theta$ .

We will draw 100,000 sample of size 17 from our posterior distribution ( $\theta = .8235$ , the mode)

- ▶ How well do random samples from the posterior match our data?
- ▶ Does the predicted maximum run length as our data?
- ▶ Does the number of switches predicted match the number in the data?

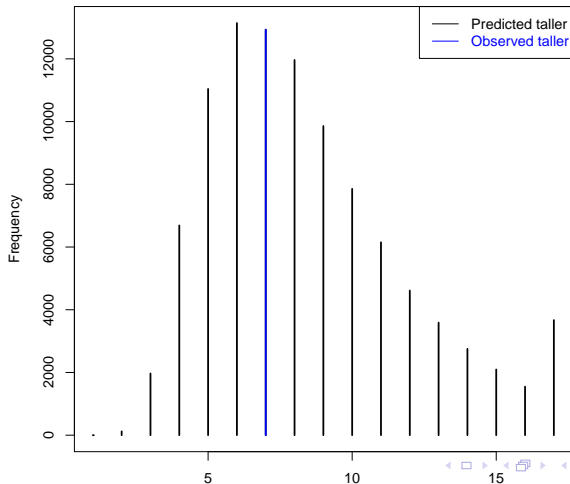
The latter two are crude ways to examine autocorrelation.

# I Data and Random Sample from Posterior



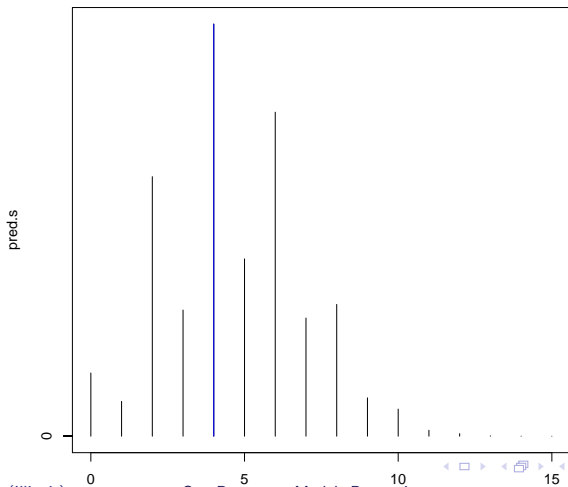
# I Max Run Length

### Longest Predicted Run Length



# I Number of Switched (need to check)

Predicted and Observed Switches



# I Practice

Just about everything we did in this lecture; however, using data from book *Applied Bayesian Statistics* by M.K. Cowles.

“On March 15, 2002, the *Iowa City Press Citizen* carried an article about the intended 19% tuition increase to go into effects at the University of Iowa (UI) for the next academic year. Let’s revisit that time and suppose that you wish to send the regents and the stat legislature some arguments against this idea. To support your argument, you would like to tell the regents and legislators what proportion of current UI students are likely to quit school if tuition is raised that much. . . . pick a simple random sample of  $n = 50$  students from the student directory and ask each of them whether she or he would be likely to quit school if tuition were raised by 19%.” 7 students said that would have to quit school.

# I Practice

Using a uniform prior,

- ▶ What is the posterior distribution? (R not required)
- ▶ What is the mean of the posterior distribution? (R not required)
- ▶ What is the median of the posterior? (R not required)
- ▶ Plot the posterior. Are the mean and mode where you expect they would be? (R: seq, dbeta, plot or use either beta binomial function or triplot)

# I Practice

- ▶ Using a uniform prior (continued),
  - ▶ Plot the credible intervals on same graph as posterior. (R: seq, dbeta, qbeta, )
  - ▶ Plot the high density intervals. (R: seq, hdi, dbeta, plot)
- ▶ What is the probability that a new sampled student would have to quite school (R: not required)
- ▶ What is your new posterior distribution? (R: not required)
- ▶ Draw random samples from the new posterior and compare with your observed data. (R: seq, rbinom, plot, lines)

# I Normal Approximation

- ▶ Since  $n$  is large, the distribution of  $y$ , which is a sum of  $n$  observations, becomes approximately normal.
- ▶ For  $y$  binomial, can use
  - ▶  $\hat{p} = y/n$  as an estimate of the mean
  - ▶  $\sqrt{n\hat{p}(1 - \hat{p})}$  as an estimate of the standard deviation.
- ▶ Also compute exact distribution (i.e., use `binom.test(y,n,conf.level=.85)`)
- ▶ Compare the estimates of  $\theta$  from Bayesian and frequentist methods.



# I Grid Method of Estimation of $p(\theta|y)$

1. Define grid (i.e., points to use to estimate  $\theta$ )

```
grid <- seq(0,1,length.out=50)
```

2. Compute value of prior at each point on grid; e.g., for beta prior (need to put in numbers for a and b)

```
prior <- dbeta(grid,a,b)
```

3. Compute value of likelihood at each point on grid; e.g., for  $n = 20$  and  $y = 14$

```
likelihood <- dbinom(y,n,prob=grid)
```

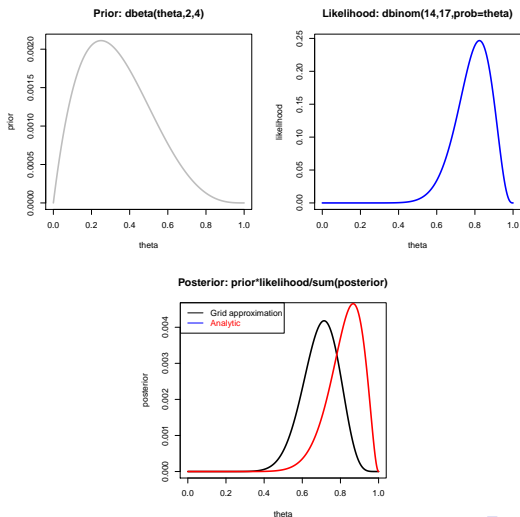
4. Compute unstandardized posterior

```
posterior.unsd <- prior*likelihood
```

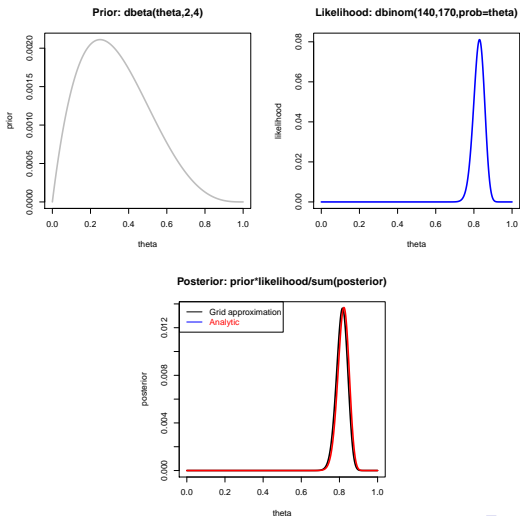
5. Standardize posterior

```
posterior <- posterior.unsd/sum(posterior.unsd)
```

# I Example: Grid approximation



# I Larger Sample



# I Using `grid.BetaBin`

Download the function “`grid.BetaBin`” from the web-site and run it to examine the impact of

- ▶ sample size
- ▶ prior (i.e.,  $a$  and  $b$  of the Beta distribution)
- ▶  $y$

# I Comparing 2 proportions

We can compute functions of  $\theta$  and obtain their posterior distributions via Monte Carlo.

Possible functions of interest for a proportion could be

- ▶ Ratio:  $\theta/(1 - \theta)$
- ▶ Difference between two:  $\theta_1 - \theta_2$
- ▶ Whether  $\theta_1 > \theta_2$

We will focus on  $\theta_i$  from two different populations.

We can accomplish this by using Monte Carlo simulation (sampling) from the posterior distributions for each population.

# I Monte Carlo

from wikipedia:

[Monte Carlo methods](#) (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results... In principle, Monte Carlo methods can be used to solve any problem having a probabilistic interpretation.

[A Monte Carlo algorithm](#) is an algorithm for computers which is used to simulate the behavior of other systems. It is not an exact method, but a heuristical one, typically using randomness and statistics to get a result.

For us, we will sample from the posterior distribution (we did this previously for model checking).

# I Monte Carlo and 2 Proportions

- ▶ Let  $\theta_1$  be the unknown probability from population 1 (e.g., taller candidate winning election 1928–2016).
- ▶ Let  $\theta_2$  be the unknown probability from population 2 (e.g., taller candidate winning election 1789–1924).
- ▶ Find the posterior distributions for each population.
- ▶ Take  $S$  samples from the posteriors:

$$\begin{array}{cc}
 p(\theta_1^{(1)}|y) & p(\theta_2^{(1)}|y) \\
 p(\theta_1^{(2)}|y) & p(\theta_2^{(2)}|y) \\
 \vdots & \vdots \\
 p(\theta_1^{(S)}|y) & p(\theta_2^{(S)}|y)
 \end{array}$$

We now have  $S$  independent samples of joint posterior distribution of  $(\theta_1, \theta_2)$

- ▶ Compute (for example)  $Pr(\theta_1 > \theta_2)$  by counting how many times this happened out of the  $S$  samples.

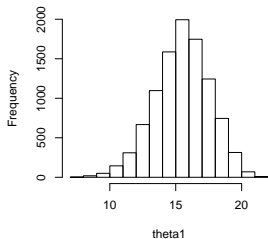
## I Example of Two

- ▶ Question: Are US presidents more likely to be taller than the other candidate in more recent elections?
- ▶ For 1928–2016:  $y_1 = 16$ ,  $n_1 = 22$ , so posterior is  $\text{Beta}(17, 6)$  and mode  $\theta_1 = .7273$ . (uniform priors)
- ▶ For 17789–1924:  $y_2 = 15$ ,  $n_2 = 30$ , so posterior is  $\text{Beta}(16, 16)$  and mode  $\theta_2 = .5161$ . (uniform priors)
- ▶ Drew 10000 samples from posterior to get distributions of  $\theta_1$ s and  $\theta_2$ s.
- ▶  $Pr(\theta_1 > \theta_2) = .4973$ .
- ▶ Because of simulation can get answer a bit different every time, but basic result is the same.
- ▶ Before going through the R code, a look at the simulated distributions.

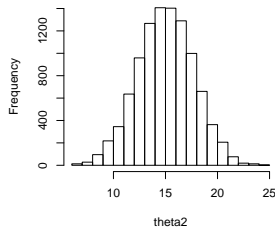


# I Monte Carlo Distributions 1e4

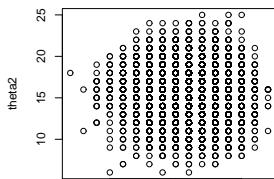
1e4 samples from posterior 1928–2016



1e4 samples from posterior 1879–1924



scatter plot of simulated values



$r(\theta_1, \theta_2) = .0035$   
 $\Pr(\theta_1 > \theta_2) \sim .5$

# I How to do this in R

- ▶ Find posteriors for both groups

- ▶ Take many random draws from each posterior:

```
# Monte carlo
```

```
S= 1e4
```

```
theta1 ← rbinom(S,n.new,theta.new)
```

```
theta2 ← rbinom(S,n.old,theta.old)
```

- ▶ Tally up

```
# Find proportion of times that theta1 > theta2
```

```
mean(theta1>theta2)
```

```
# Another way...
```

```
new.greater ← ifelse(theta1>theta2,1,0)
```

```
table(new.greater)
```

# I Practice

Return the problem about drop out due to tuition increase. Suppose that at the Iowa State University a random sample of 50 students were asked whether they would have to quit school if tuition increased by 19%. Of the 50, 10 said they would have to drop out. Is the proportion at Iowa State larger than that at University of Iowa?

# I What We Covered

- ▶ Beta and Binomial Distributions.
- ▶ Conjugate prior.
- ▶ Exchangeability.
- ▶ Inference for a proportion.
- ▶ Credible intervals & High Density intervals.
- ▶ Predicting new observation.
- ▶ Up-dating posterior distribution given new data.
- ▶ How Learning occurs.
- ▶ A little model checking.
- ▶ Monte Carlo
- ▶ Comparing two proportions