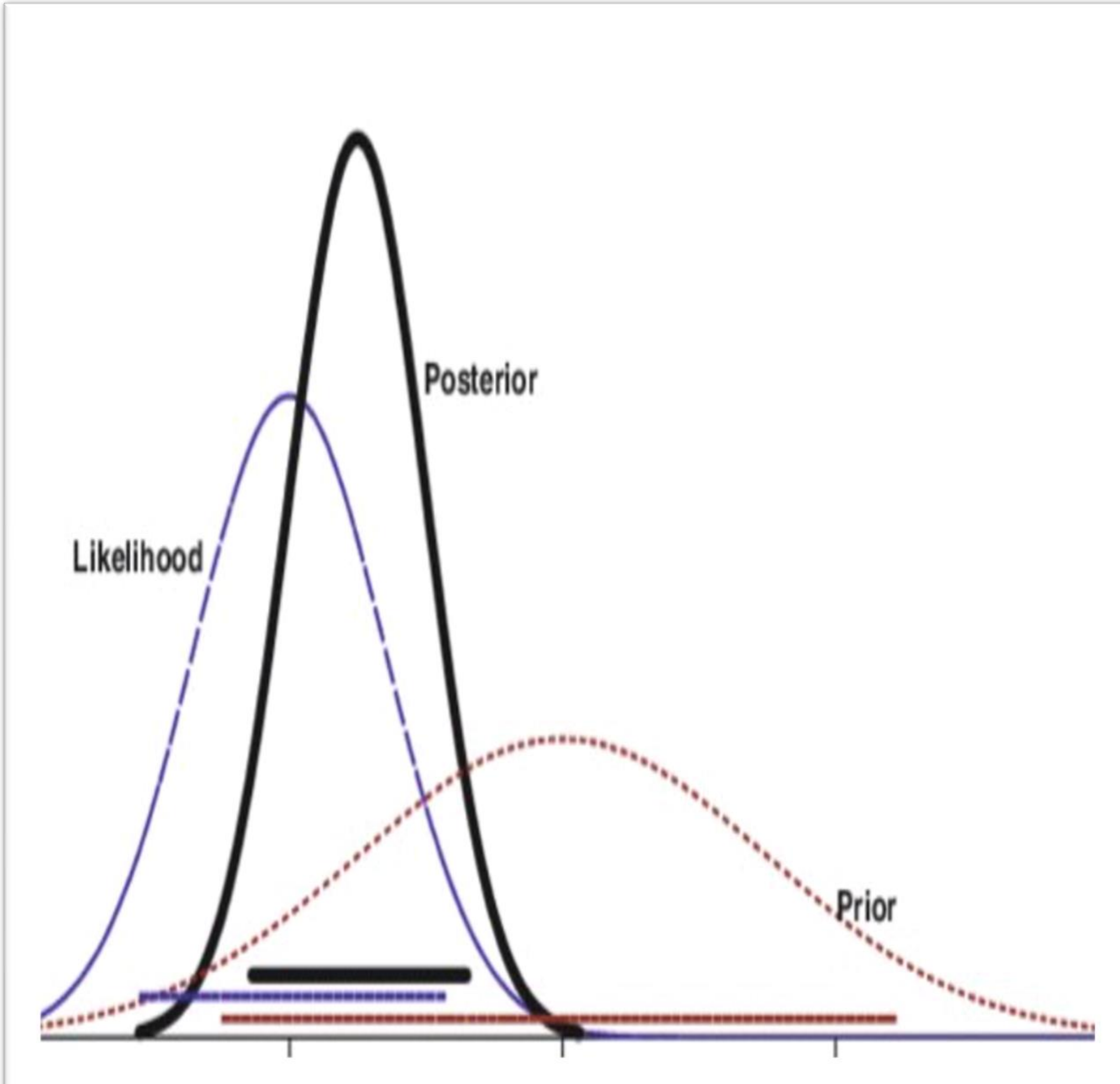


Applied Bayesian Statistics

Unit 3

Population Proportion Estimation

Hamdy F. F. Mahmoud, PhD
Collegiate Assistant Professor
Statistics Department @ VT
ehamdy@vt.edu

This Unit covers

- A Motivation Example
- Bayesian framework
  - **Stage 1: The data distribution**
  - **Stage 2: The Prior Distribution**
  - **Stage 3: Update the Prior: The Posterior Distribution**
- Computing and Graphing the Posterior Distribution
- What is to come?

Motivation Example: Tuition Raise (continue)

- Assume that Virginia Tech decided to raise tuition 20%, what is the proportion of students who would be likely to quit school?

- You do not have the time or resources to locate and interview all 27,730 (2018–19) students, so you cannot evaluate the population proportion, π. Instead, you will 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 20%. You wish to use your sample data to estimate the population proportion π and to determine the amount of uncertainty in your estimate.

Bayesian framework to estimate a population proportion

To estimate the population proportion, π, using Bayesian approach, you need to follow three stages:

1. Likelihood function – the data distribution
2. Prior distribution – the parameter distribution
3. Posterior distribution – the updated prior distribution

Posterior distribution $\propto$ $\textit{prior distribution} \times \textit{likelihood function}$

In Part I, we discussed different priors for the population proportion and drove the posterior distribution under the uniform prior. Here, we need to discuss the posterior distribution under conjugate priors.

Conjugate Priors

A common way to construct a prior distribution is to choose the prior from a parametric family of densities such that the prior has the same functional form as the likelihood function. Such a prior is called a conjugate prior.

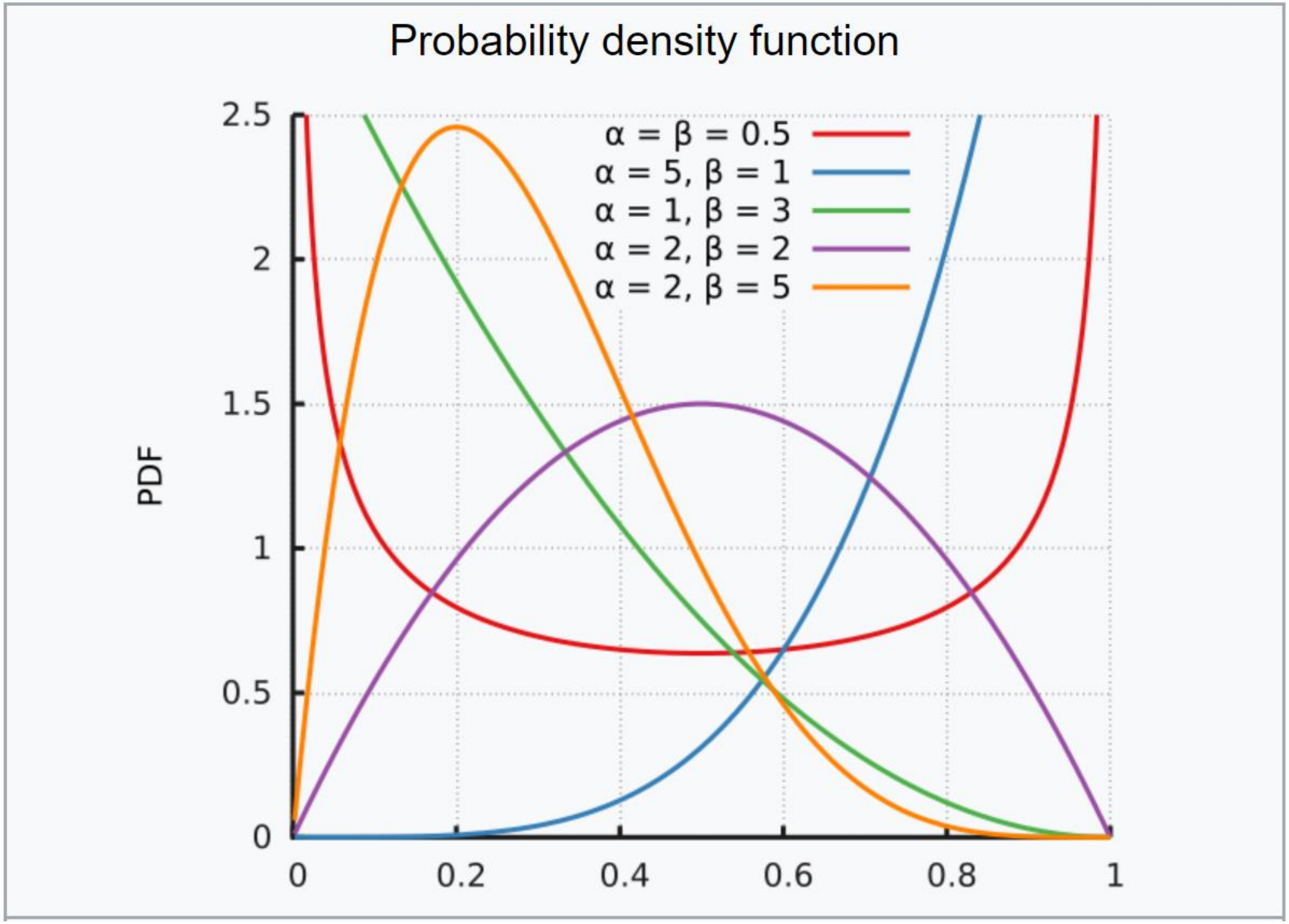
- Regarding the binomial example, recall that π must lie between 0 and 1, and note how the parameter π appears in the binomial likelihood. There π is raised to a nonnegative power, and $(1-\pi)$ also is raised to a nonnegative power. *What density function is this?*
- A beta family of densities, with fixed parameters α and β and with the random variable called π would be written as follows:

$$\pi \sim Beta(\alpha, \beta)$$

$$p(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\,\Gamma\beta}\pi^{\alpha-1}(1-\pi)^{\beta-1}, \quad 0 < \pi < 1$$

- What are the kernel and normalizing constant of this density function?

Beta distribution at different parameter values



Computing the Posterior Distribution with a Conjugate Prior

$$p(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\,\Gamma\beta}\pi^{\alpha-1}(1-\pi)^{\beta-1}, \quad 0 < \pi < 1$$

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

- Recall the relationship of the posterior distribution to the prior and likelihood

$$p(\pi|y) \propto p(\pi)L(\pi; y)$$

- So in the case of a beta prior and a binomial likelihood,

$$p(\pi|y) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1} \cdot \pi^{y}(1-\pi)^{n-y}$$

$$\propto \pi^{\alpha+y-1}\,(1-\pi)^{\beta+n-y-1}$$

This is the kernel of what density function?

- This is the kernel of another beta density function.

$$p(\pi|y) = Beta(\alpha^*, \beta^*) = Beta(\alpha+y, \beta+n-y)$$

Write down the density function and its mean and variance of

$$p(\pi|y) = Beta(\alpha^*, \beta^*) = Beta(\alpha + y, \beta + n - y)$$

Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

Here are several ways to think about choosing the parameters of a beta distribution to express prior beliefs or knowledge about an unknown proportion.

- **Strategy 1:** Graph some beta densities until you find one that matches your beliefs.

- **Strategy 2:** If we have a previous study, we use $\alpha - 1$ = # of successes and $\beta - 1$ = # of failures. Why?

- **Strategy 3:** Solve for the values of α and β that yield:
  - The desired mean (The mean of a beta(α ,β) density is ($\alpha/\alpha+\beta$).
  - The desired equivalent prior sample size, which for a beta(α ,β) is $\alpha + \beta - 2$, explain!. When you use this method, you are saying that your knowledge about π is as strong as if you'd seen a previous sample consisting of $\alpha - 1$successes and $\beta - 1$ failures.

- **Strategy 4:** Choose values of α and β that produce a prior probability interval that reflects your belief about π.

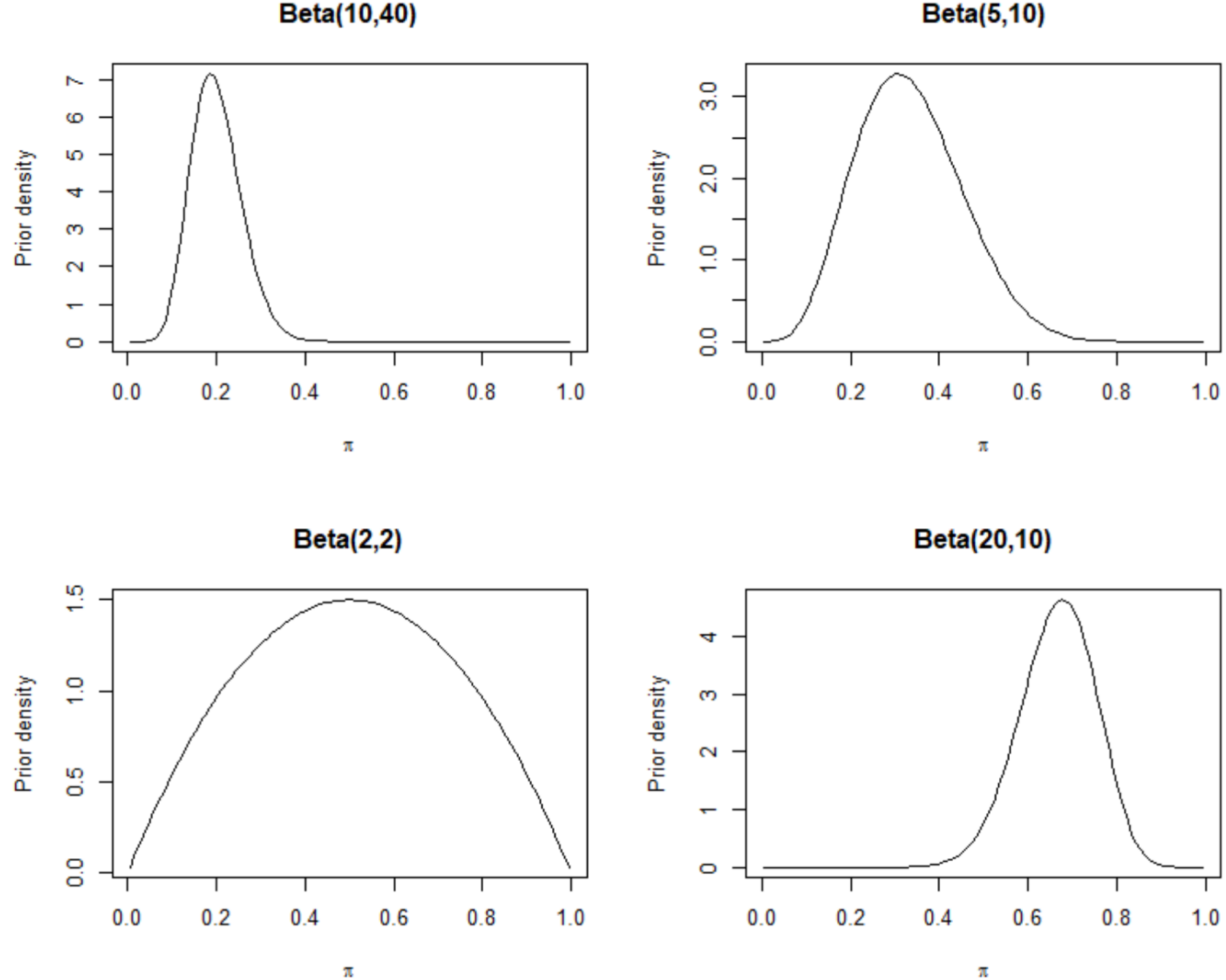
Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

- The new data must NOT be used in any way in constructing the prior density! We'll see shortly why that would make inference invalid.

- Let us apply the first four strategies to the quitting-school-because-of-rising tuition example. We are attempting to construct a reasonable prior before we see the results of the actual survey of 50 students.

Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

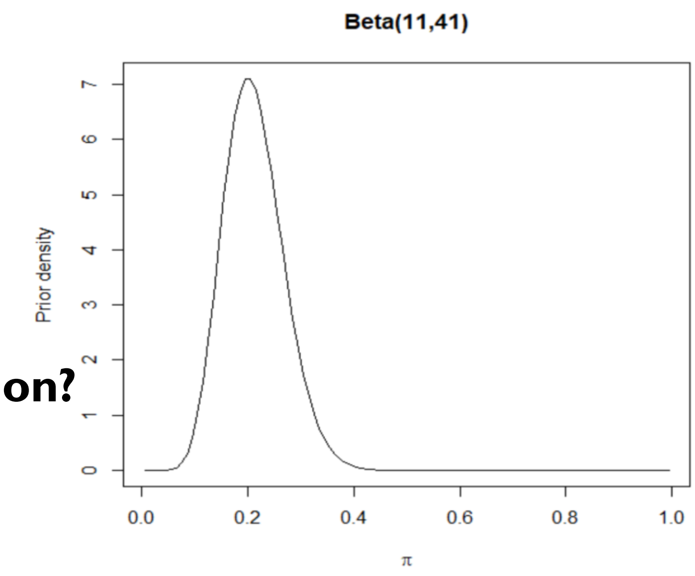
- **Strategy 1:** Graph some beta densities until you find one that matches your beliefs.



Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

- **Strategy 2:** If we have a previous study, we use $\alpha - 1$ = # of successes and $\beta - 1$ = # of failures.

We wish to use any relevant data available before we do our survey. Suppose that we read that such a survey has already been taken at Iowa State University, in which 50 students were interviewed and 10 said they would quit school; 40 said they would not. By strategy 2, this might suggest a *Beta(11, 41)* prior.



$$p(\pi) \propto \pi^{10}(1-\pi)^{40}$$

- **What is the mean of this density function?**

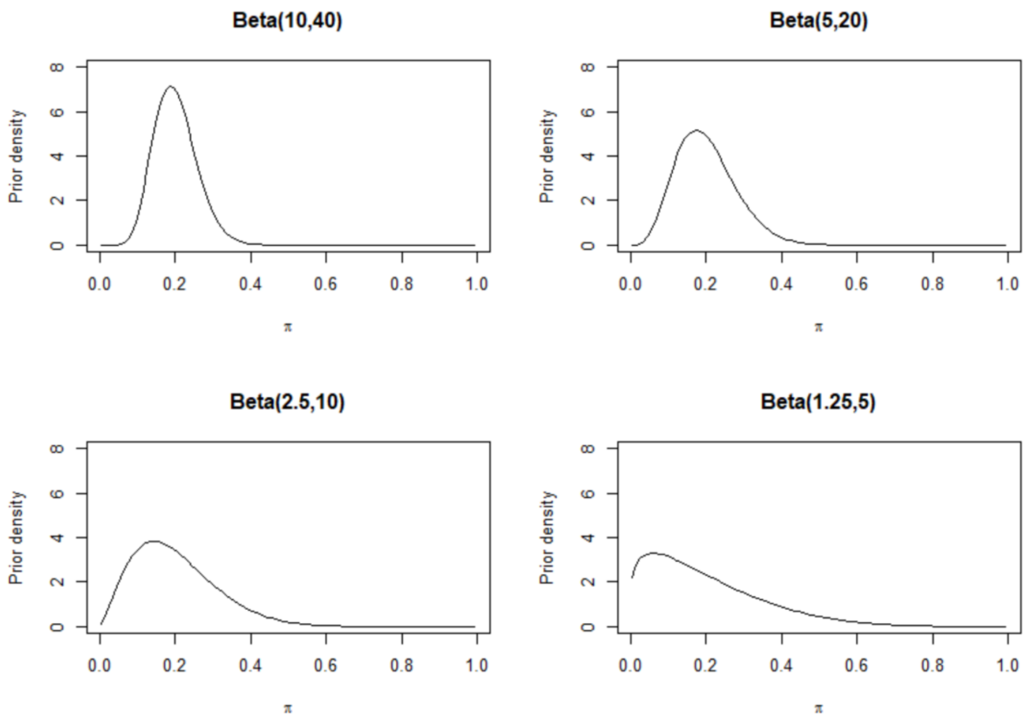
Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

- **Strategy 3:** Solve for the values of α and β that yield:
  - The desired mean (The mean of a beta(α ,β) density is ($\alpha/\alpha+\beta$).
  - The desired equivalent prior sample size, which for a beta(α ,β) is $\alpha+\beta-2$. When you use this method, you are saying that your knowledge about π is as strong as if you'd seen a previous sample consisting of $\alpha-1$ successesand $\beta-1$ failures.

Say that we want a prior mean of 0.2, the same as the sample proportion from the ISU data, but an "equivalent prior sample size" (remember, for a beta prior that is $\alpha+\beta-2$) that is smaller than 50. One possibility is to look at the graphs of several different beta distributions, all with the same mean 0.2 but with smaller and smaller equivalent prior sample sizes, and seek one that matches our beliefs.

Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

- **Strategy 3:** Solve for the values of α and β that yield:
  - The desired mean (The mean of a beta(α ,β) density is ($\alpha/\alpha+\beta$)).
  - The desired equivalent prior sample size, which for a beta(α ,β) is $\alpha+\beta-2$. When you use this method, you are saying that your knowledge about π is as strong as if you'd seen a previous sample consisting of $\alpha-1$successesand $\beta-1$ failures.

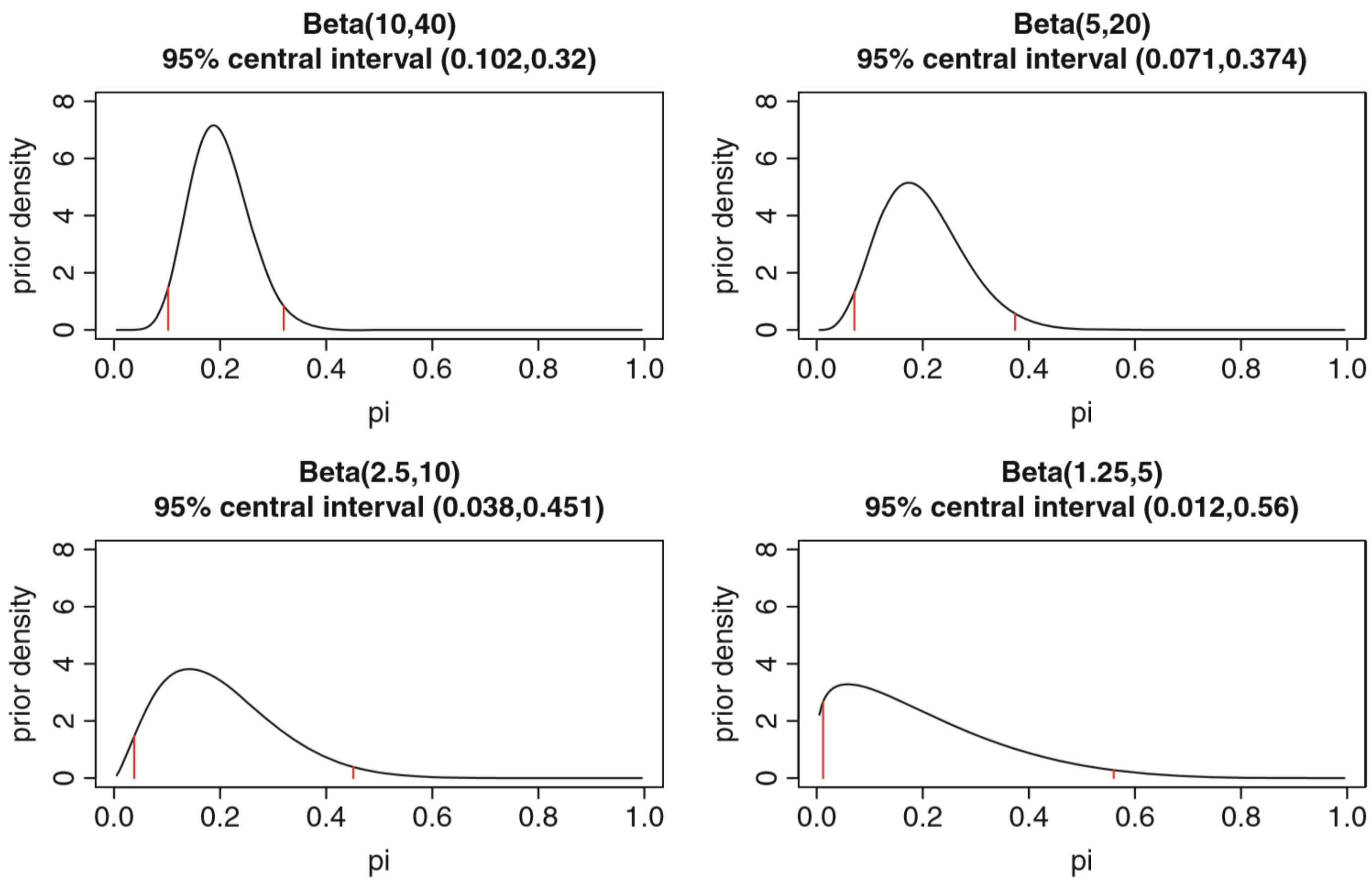


All of these prior densities have the same mean 0.2. **Which one do you choose?**

Note that when the sum of the two parameters is larger as in the first plot, the density curve is fairly concentrated over an interval near 0.2

Choosing the Parameters of a Beta Distribution to Match Prior Beliefs

- **Strategy 4:** Choose values of α and β that produce a ***prior probability interval*** that reflects your belief about π.



95% prior intervals for beta densities with different parameter values

Graphing the prior, likelihood, and posterior distributions

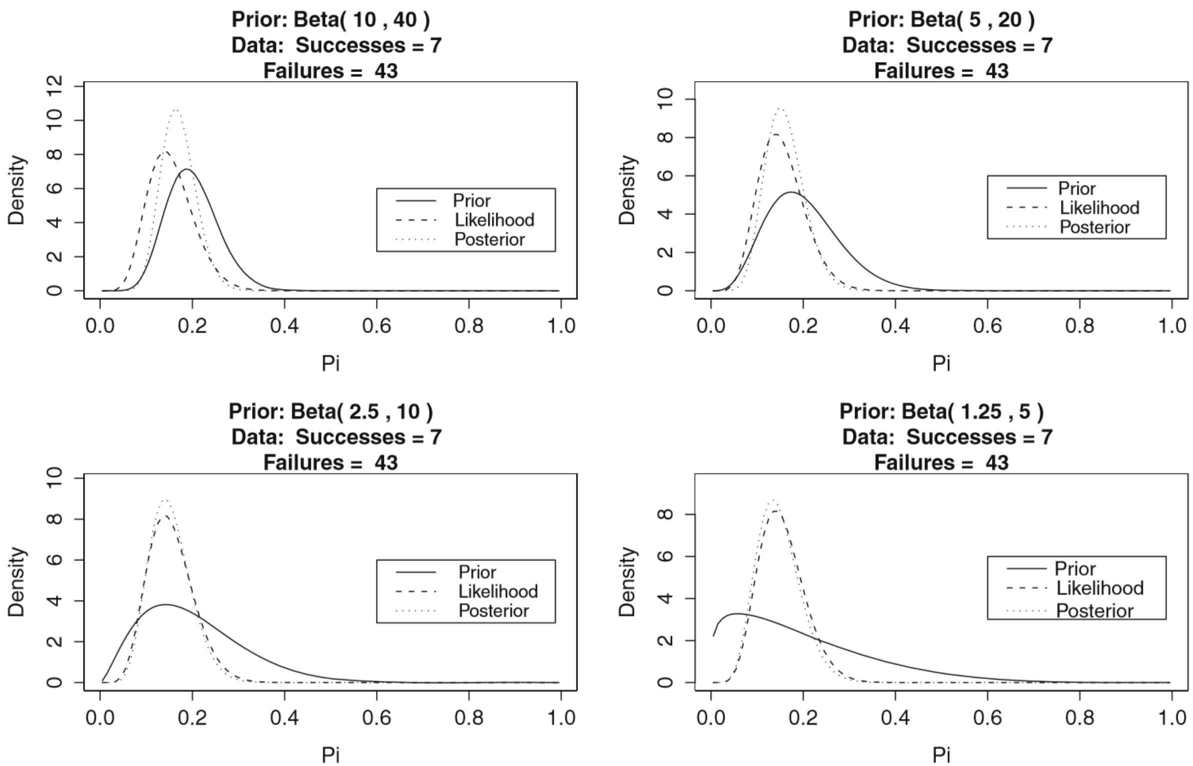
- Suppose you chose the beta(10,40) prior because it best represented your beliefs. You then gathered your own data on n = 50 of VT students, and found y=7 “successes” and n−y=43 “failures.” Then your posterior distribution of π given your beta prior and the new data is

$$p(\pi|y) \propto \pi^{\alpha+y-1}(1-\pi)^{\beta+n-y-1}$$
$$\pi^{17-1}(1-\pi)^{83-1}$$

This is the kernel of a beta(17, 83) density.

- **Another possibility:** if the beta(1.25, 5) prior better represented your previous knowledge, then your posterior distribution for π , given my prior and your data, would be a beta(8.25,48).

Graphing the prior, likelihood, and posterior distributions



Prior densities, normalized likelihoods, and resulting posterior densities

- Note that in all cases, the posterior density is more concentrated than either the prior density or the likelihood. This makes sense: When we combine our previous knowledge with the additional information in the current data, our knowledge becomes more precise than when we consider either one of the two sources alone
- The less informative the prior is (in the case of a beta prior, the smaller its parameter values), the more the data, as expressed in the likelihood dominates the posterior density

Practice

Go to canvas, files, and find the problem sheet Hands-on #3 that is in “Hands-on Sheets” folder.

We still need to know:

- How do we graph a density function?
- How do we use a posterior density function to estimate a population parameter, construct a confidence interval, perform a test of hypothesis about the population parameter, and make predictions?

- **In addition, we need to use a statistical software!**
- **You should bring your computer.**

  So let us install and learn R! – Go to Lab #1