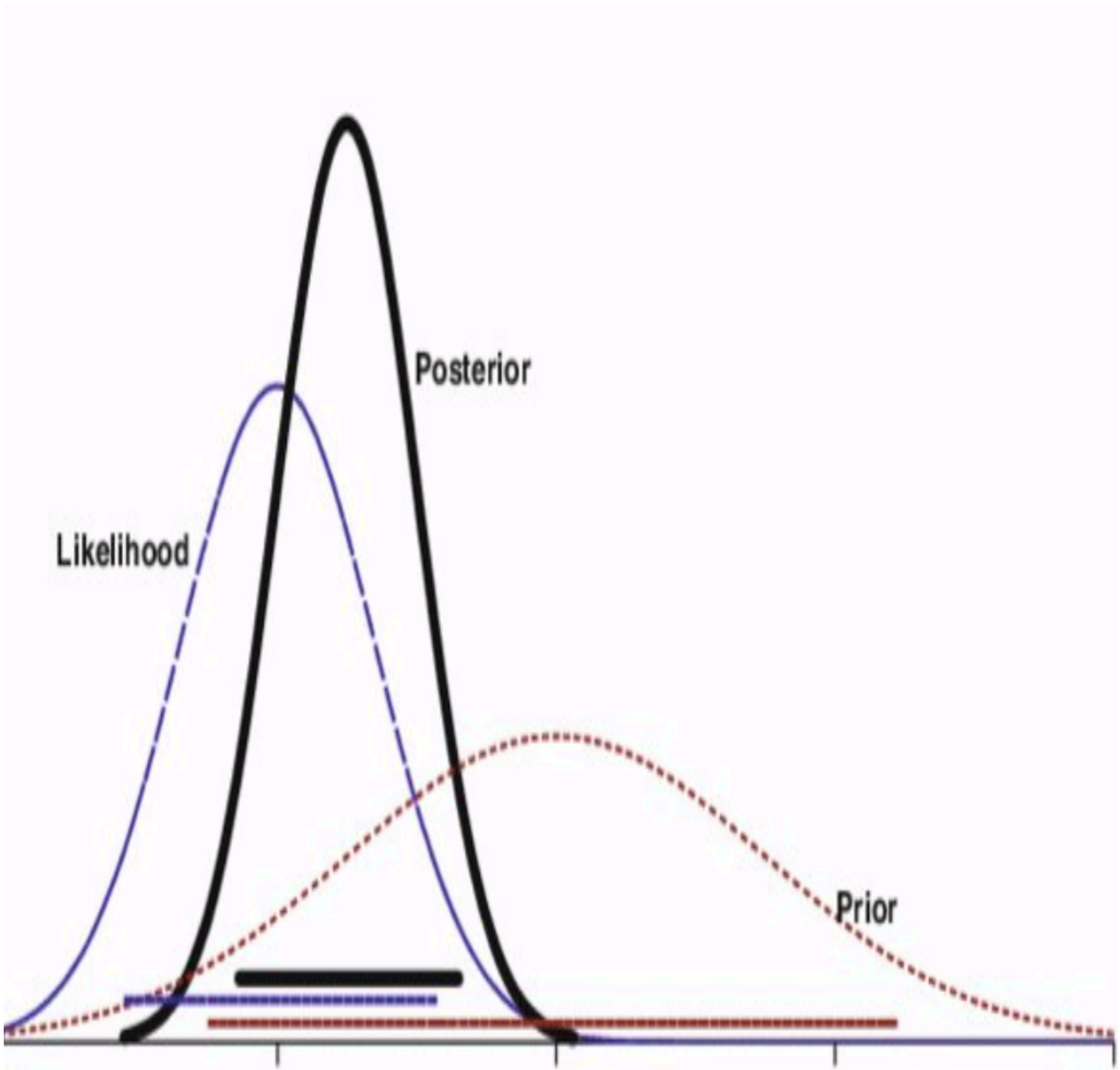


Applied Bayesian Statistics

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

- 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 $\propto$ $prior\ distribution \times likelihood\ function$

1^{st} stage of Bayesian model: Likelihood function

- **The data distribution – likelihood function**

  Before you select the students in your sample and interview them, you can regard each student's potential response as a Bernoulli random variable.

$$x \sim Bernoulli(\pi)$$

$$p(x) = \pi^{x}(1-\pi)^{1-x}, \qquad x = 0,1$$

- Define a random variable Y as the count of the number of successes (yeses) in your sample. Y meets the conditions of a binomial random variable — it is the count of the number of successes in n-independent Bernoulli trials, all with the same success probability. We can write

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

1^{st} stage of Bayesian model: Likelihood function

$$Y \sim Binomial(n, \pi)$$

$$P(y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \quad y = 0, 1, 2, \ldots, n$$

- What is $\binom{n}{y}$?
- What are the possible values of y ? (Of course, you won't find out the value that y takes on in your own survey until you actually draw the $n = 50$ students and interview them.)
- If we knew π, we could use the binomial probability mass function to compute the probability of obtaining any one of the possible values y that the random variable Y could take on in our sample.

For example, if we magically knew that $\pi = 0.1$, then the probability of getting $y = 4$ yesses among the respondents in a random sample of 50 students would be

$$p(Y = 4 | \pi = 0.1) = \binom{50}{4} 0.1^4 \, 0.9^{46} = 0.181$$

Kernels and Normalizing Constants

- Before we proceed, we need to distinguish between the *kernel* of a function and the *normalizing constant in Bayesian statistics*.

- The *kernel* includes all terms that will change in value for different values of the **variable** of interest.

- The *normalizing constant* includes all terms that will NOT change in value for different values of the **variable** of interest.

- **What are the *kernel* and *normalizing constant* of the following function?**

$$P(y) = \binom{n}{y} \pi^{y} (1-\pi)^{n-y}, \quad y = 0, 1, 2, \ldots, n$$

The Likelihood Function

After you interview the 50 students, you will know how many said yes. That is, you will know which value y the random variable Y actually took on. Suppose this number turns out to be $y = 7$. Plug in $y = 7$ in the binomial function

$$L(\pi) = \binom{50}{7} \pi^7 (1 - \pi)^{43}, \qquad 0 < \pi < 1$$

When viewed in this way, the expression is called the likelihood function.

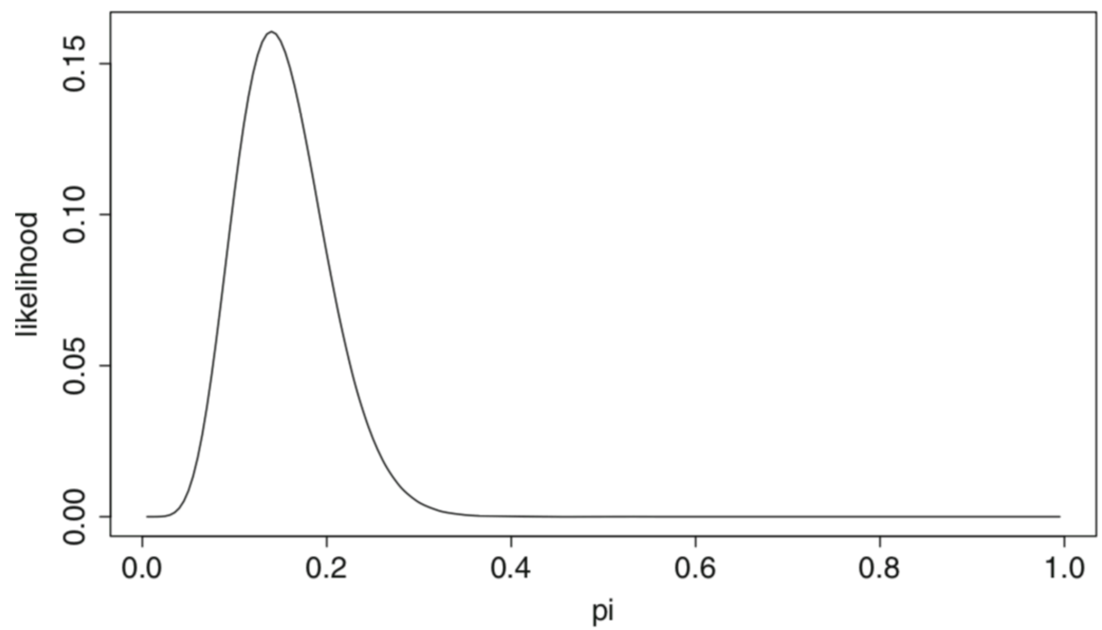
- This function is a function of π.
- What are the kernel and normalizing constant of this function?

The Likelihood Function

- We could compute this likelihood for different values of π. Intuitively, values of π that give larger likelihood evaluations are more consistent with the observed data.
- Frequentist estimation approach uses only this likelihood function to estimate π, population proportion.

The *sample* proportion of yesses in your observed data is $\hat{\pi} = \frac{7}{50} = 0.14$

- Note that this is the value of π at which the likelihood function attained its maximum.



Binomial likelihood function with 7 successes in 50 trials

The Likelihood Function

- The normalizing constant does not depend on π, so we can write the likelihood function

$$L(\pi) = \binom{50}{7} \pi^7 (1-\pi)^{43}, \quad 0 < \pi < 1$$

as

$$L(\pi) \propto \pi^7 (1-\pi)^{43}, \quad 0 < \pi < 1$$

Practice:

Go to canvas, files, and find the problem sheet under Hands-on #2 that is under “Hands-on Sheets” folder.

2nd Stage of the Bayesian Model: The Prior

- To carry out a Bayesian analysis to learn about the unknown population proportion π, we need to assess our previous knowledge or belief about π ***before*** we observe the data from the survey. **Why?**

- The Bayesian approach to express prior knowledge about a population parameter is to put a probability distribution on the parameter—that is, to treat the unknown population parameter *as if* it were a random variable.

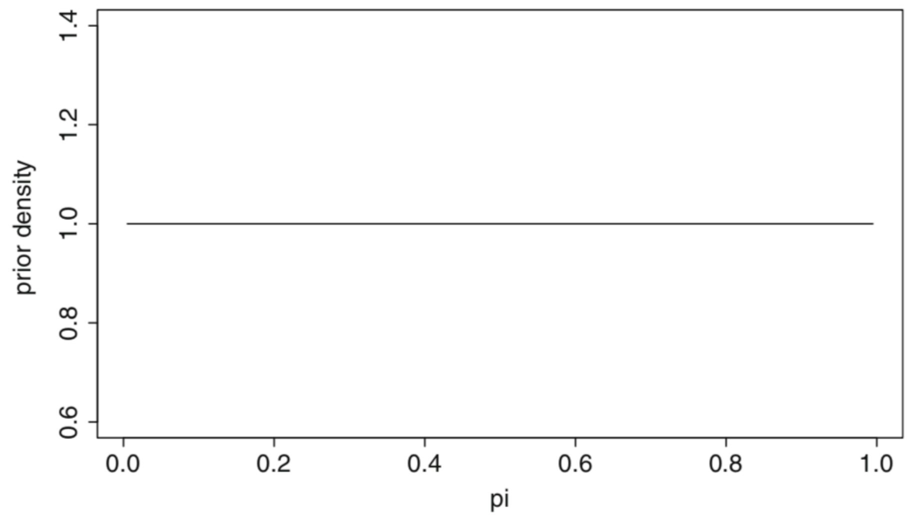
- Because it is a proportion, the parameter π hypothetically could take on any value in the interval (0, 1), although most of us realize that some ranges of values are much more likely than others.

2nd Stage of the Bayesian Model: The Prior

- A person who has little or no knowledge about university students might consider all values in (0, 1) equally plausible before seeing any data. A *uniform* density on (0,1) describes this belief (or state of ignorance!) mathematically and graphically.

$$\pi \sim U(0,1)$$

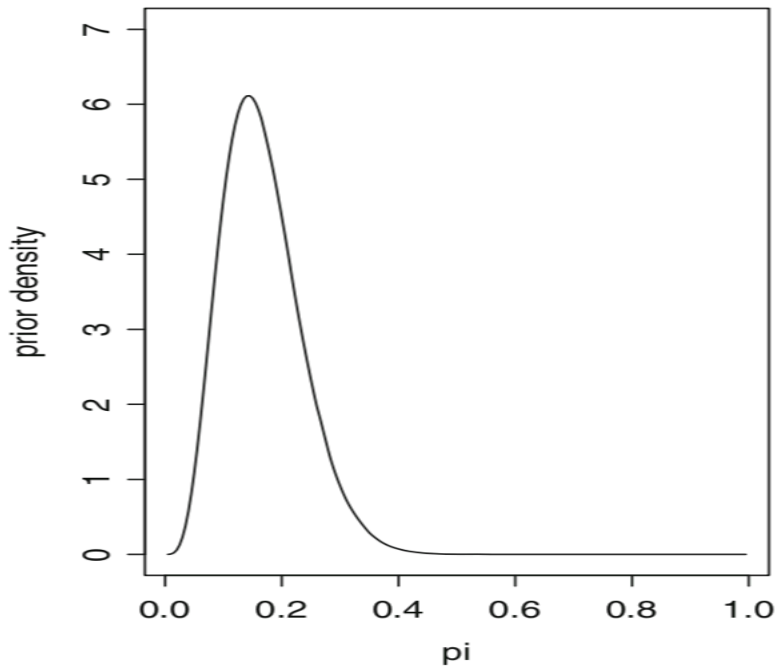
$$P(\pi) = 1, \qquad 0 < \pi < 1$$

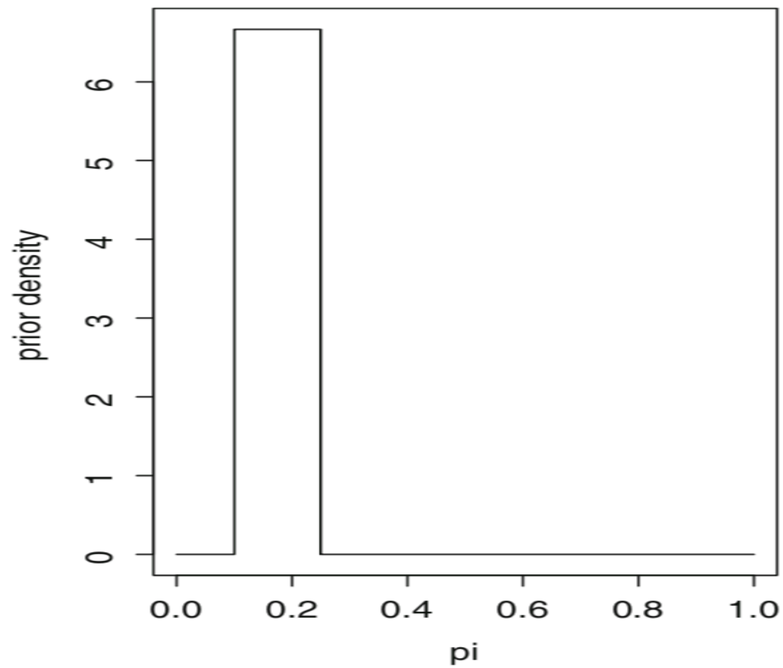


- This continuous uniform distribution is called a "vague" or "**noninformative**" prior. It says that if we pick any two intervals within (0,1) that are of equal width— say (0.2,0.29] and (0.80,0.89] — there is equal probability that π lies in each of them.

Other possible prior distributions

- If a person has knowledge or belief regarding the value of π, his or her prior will be **informative**.



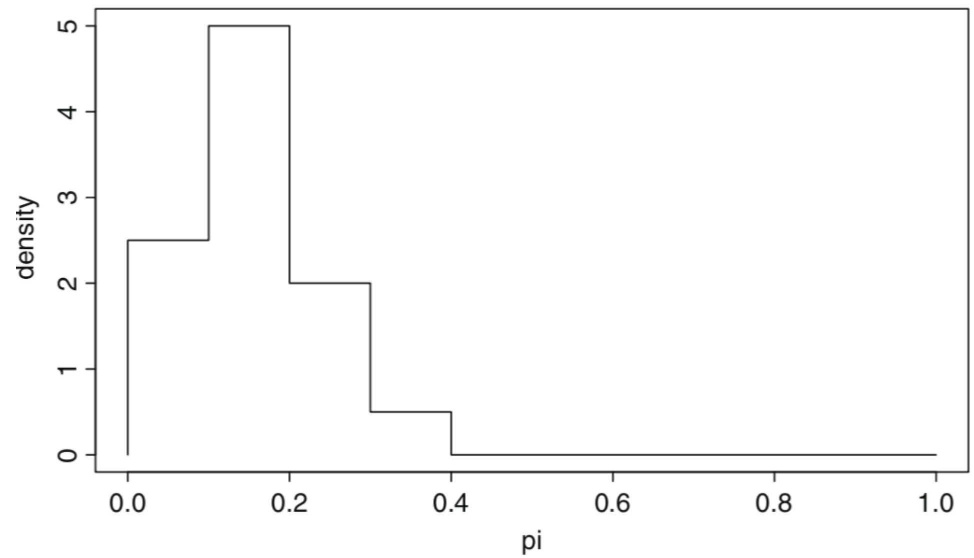


- Two different possible priors expressing the belief that π most likely lies between 0.1 and 0.25.

Other possible prior distributions

- If a person has knowledge or belief regarding the value of π, his or her prior will be **informative**.

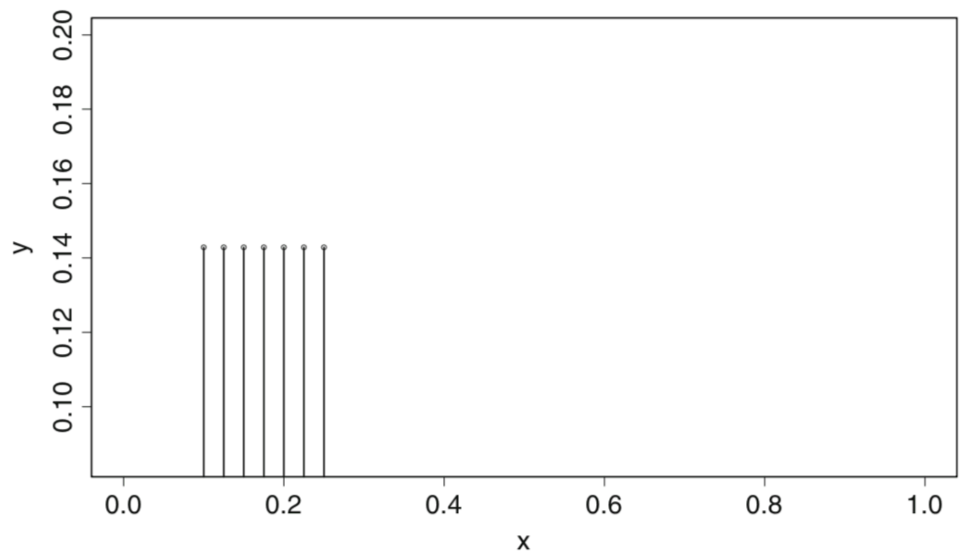
Histogram prior



- For a valid histogram prior, the areas of all the bars must sum to one.
- It represents the prior belief that the probability that π lies in the interval [0,0.1) is 0.25, in [0.1, 0.2) is 0.5, etc.

Other possible prior distributions

- All of the prior distributions we have mentioned so far treat the unknown parameter π as if it were a *continuous* random variable. In some application, even though the parameter of interest may in reality take on any value over a continuum, if very exact inference is not required, a discrete prior may be similar to work with and may adequately express available prior information.



A discrete uniform prior on π.

$$p(\pi) = \frac{1}{7}$$

$$\pi = 0.1, 0.125, 0.15, 0.175, 0.20, 0.225, 0.25$$

3rd stage: Updating the Prior, the Posterior Distribution

- Let's see what happens if we use the "**noninformative**" continuous **uniform prior** for our analysis.

Posterior distribution $\propto$ *prior distribution* $\times$ *likelihood function*

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

$$p(\pi|y) \propto p(\pi)\times \binom{n}{y}\, \pi^{y}(1-\pi)^{n-y}$$

$$p(\pi|y) \propto p(\pi)\times \binom{50}{7}\, \pi^{7}(1-\pi)^{50-7}$$

$$p(\pi|y) \propto p(\pi)\times \pi^{7}(1-\pi)^{43}$$

Using uniform prior, $p(\pi) = 1$, the posterior distribution will be

$$p(\pi|y) \propto 1\times \pi^{7}(1-\pi)^{43}$$

$$p(\pi|y) \propto \pi^{7}(1-\pi)^{43}$$

3rd stage: Update the Prior: The Posterior Distribution

$$p(\pi|y) \propto \pi^7(1-\pi)^{43}, \qquad 0 < \pi < 1$$

- This is the likelihood function!!
- The graph of this function is the one on slide #10.
- **This is the kernel of what distribution?**

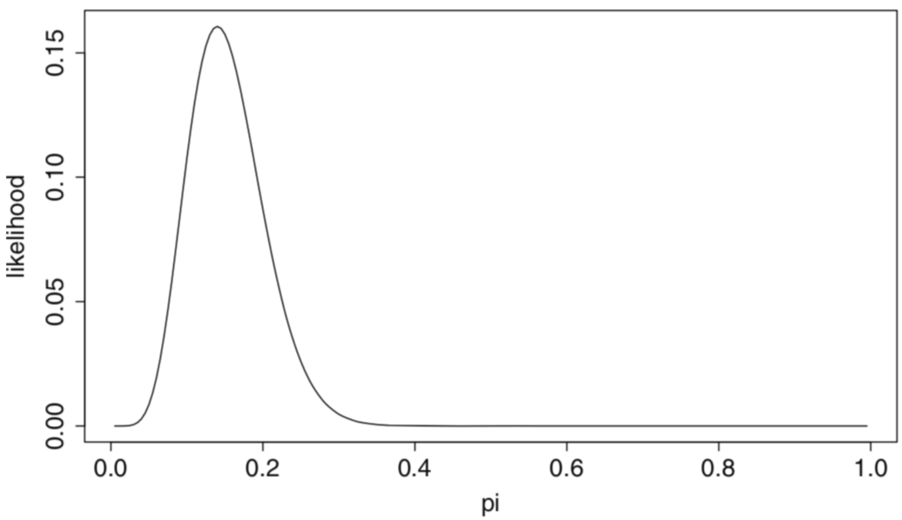
Beta(8, 44)!!!

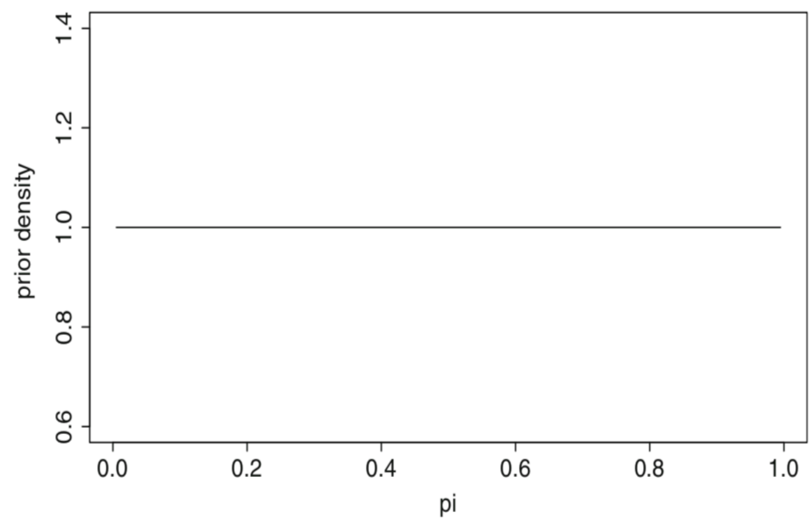
- **What is the mean (posterior mean or the Bayesian estimation of π)?**
  - $\frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha+\beta}} = \frac{\mathbf{8}}{\mathbf{8+44}} = \mathbf{0.15}$

Conclusions

- The maximum likelihood estimation of π – frequentist approach, $\hat{\pi} = \frac{7}{50} = 0.14$.

- The Bayesian estimation of π under Uniform prior distribution – the posterior mean of the posterior distribution function, 0.15





NEXT

- **What other possible prior distributions for a population proportion?**