SOCI 620: Quantitative methods 2

Jan 10

Probability models of social processes

1. Administrative
2. Estimating unemployment two ways
3. Bayes' rule
4. Grid approximation in R

Administrative

Installing software

- Script to test that everything is installed: https://soci620.netlify.app/labs/lab_01.R
- How many are still working on getting the software running?

Getting started with R

- Link to a good tutorial for getting started using R is posted on Teams (https://github.com/matloff/fasteR)
- Completing (and understanding) the first 8 lessons from this tutorial should give you a good foundation for the course

Labs

- It looks like Tuesdays around 1pm will work best for the largest number of people
- Labs will begin next week

Unemployment rate in Newfoundland and Labrador

- How do we say something about the proportion of residents who have no employment?
- Full census is impractical.
- Instead we use a probability model to *estimate* the proportion based on a sample (S).

$$S = (E, E, E, U, U, E, E, E, U, E)$$

$$n = 10$$
$$k = 3$$

Estimating unemployment

$$S = (E, E, E, U, U, E, E, E, U, E)$$

Frequentist estimation

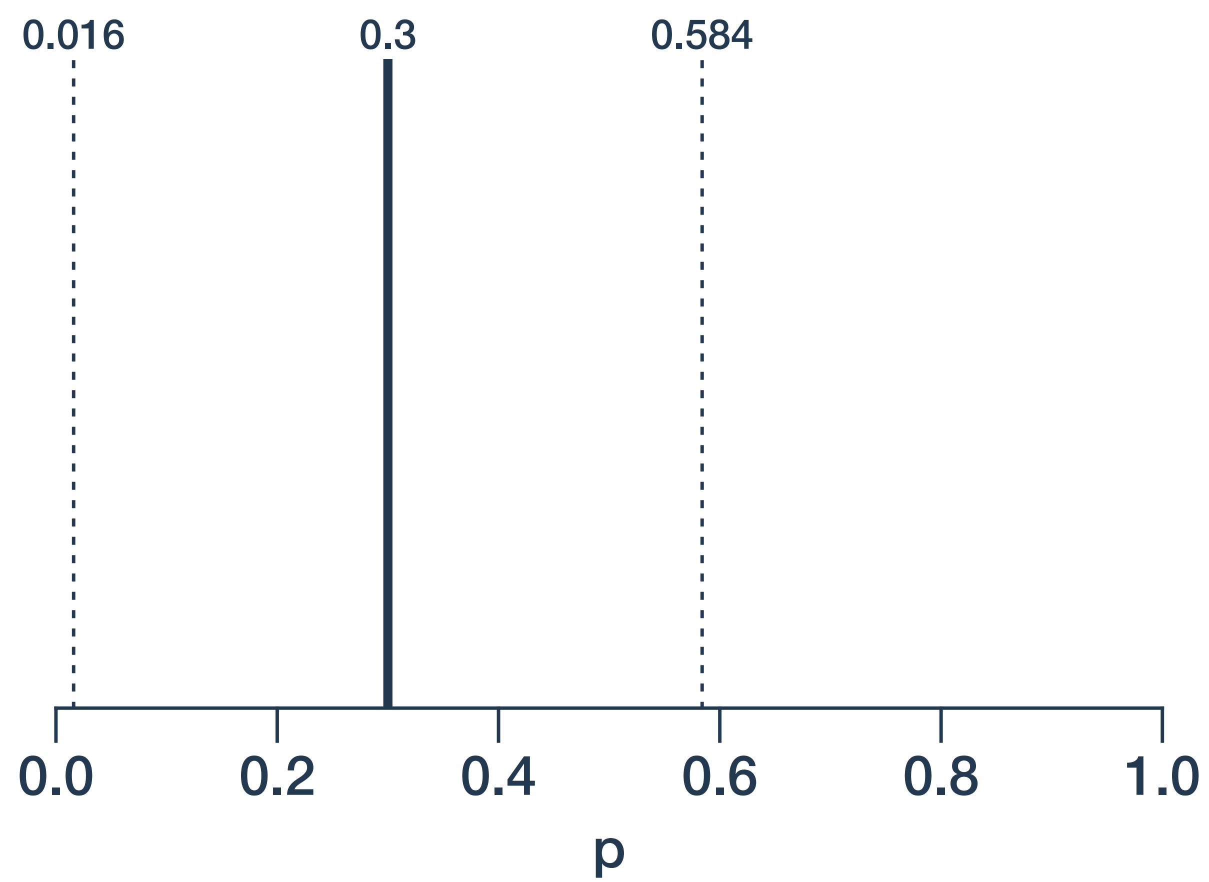
1. Pick an "estimator" (such as sample proportion)

2. Generate *point estimate* of *p*

$$\hat{p} = \frac{3}{10} = 0.3$$

3. Use approximation of the sampling distribution to quantify uncertainty

$$\hat{\sigma} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.145$$

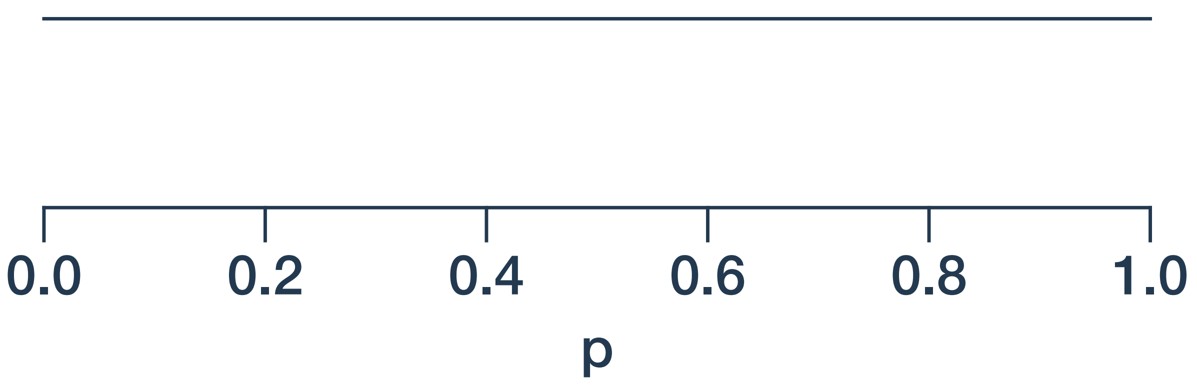


Estimating unemployment

$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

1. Pick a prior (such as uniform distribution)



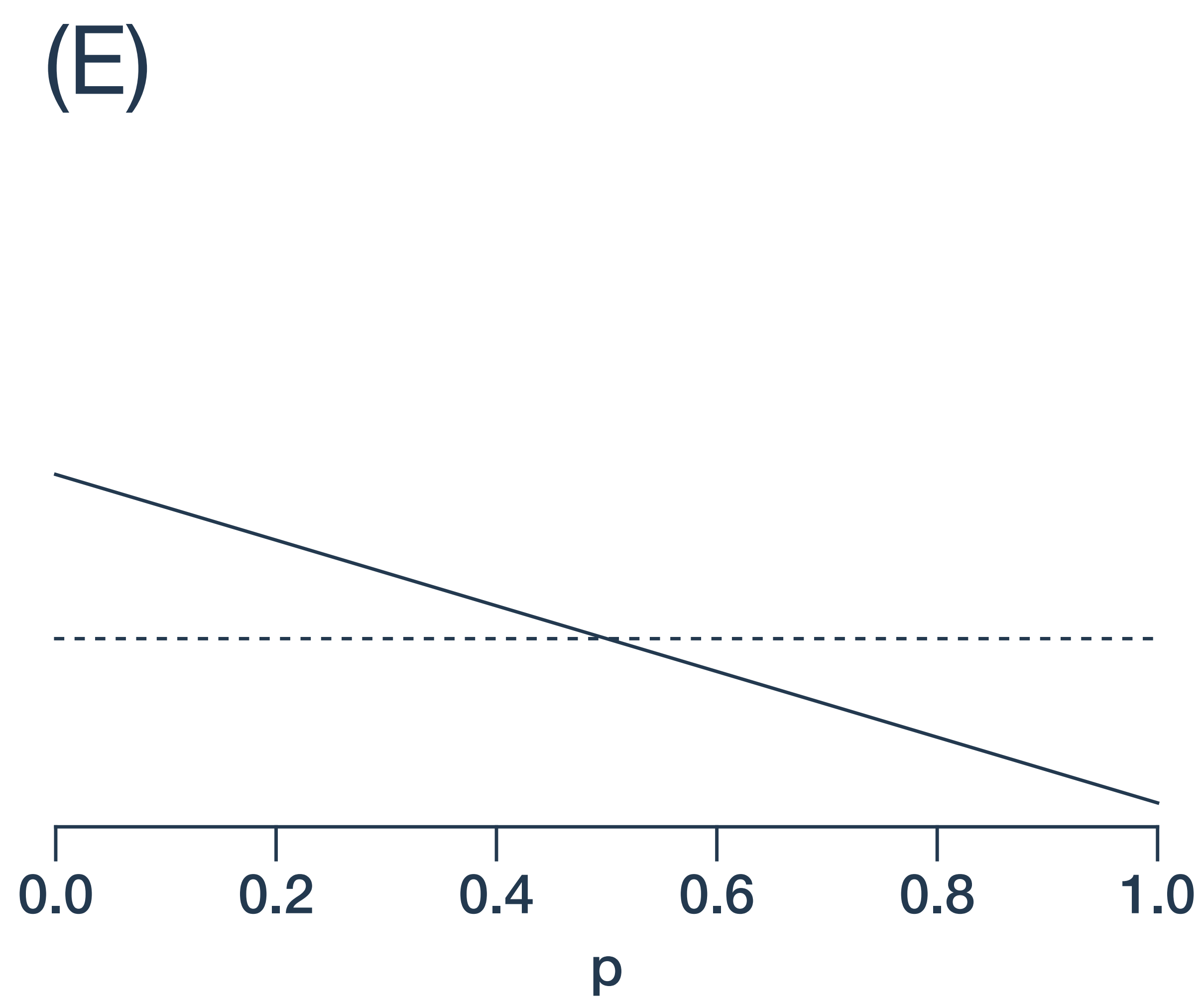
Estimating unemployment

$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

1. Pick a prior (such as uniform distribution)

2. Update prior with data (one at a time or all at once)



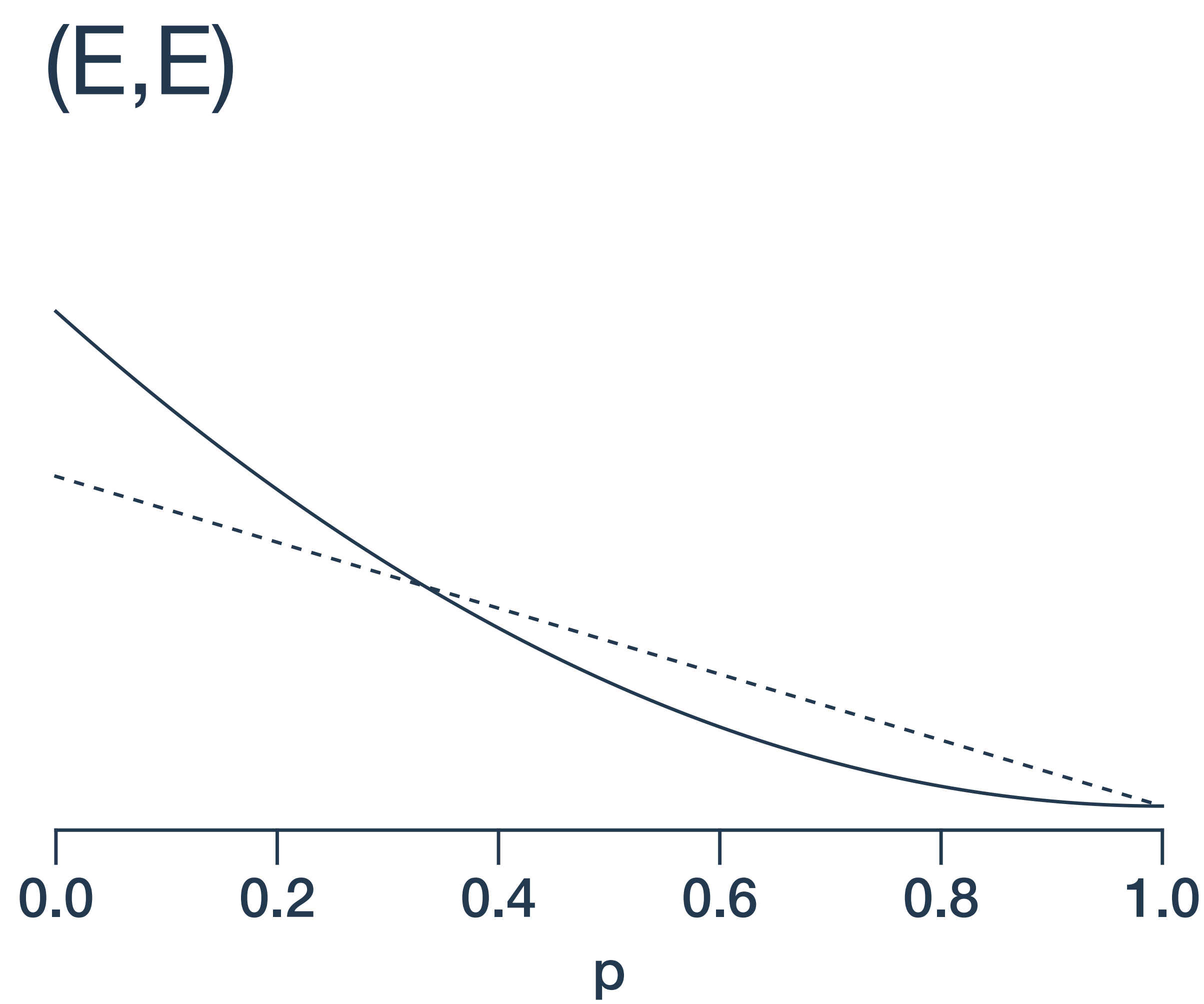
Estimating unemployment

$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

1. Pick a prior (such as uniform distribution)

2. Update prior with data (one at a time or all at once)

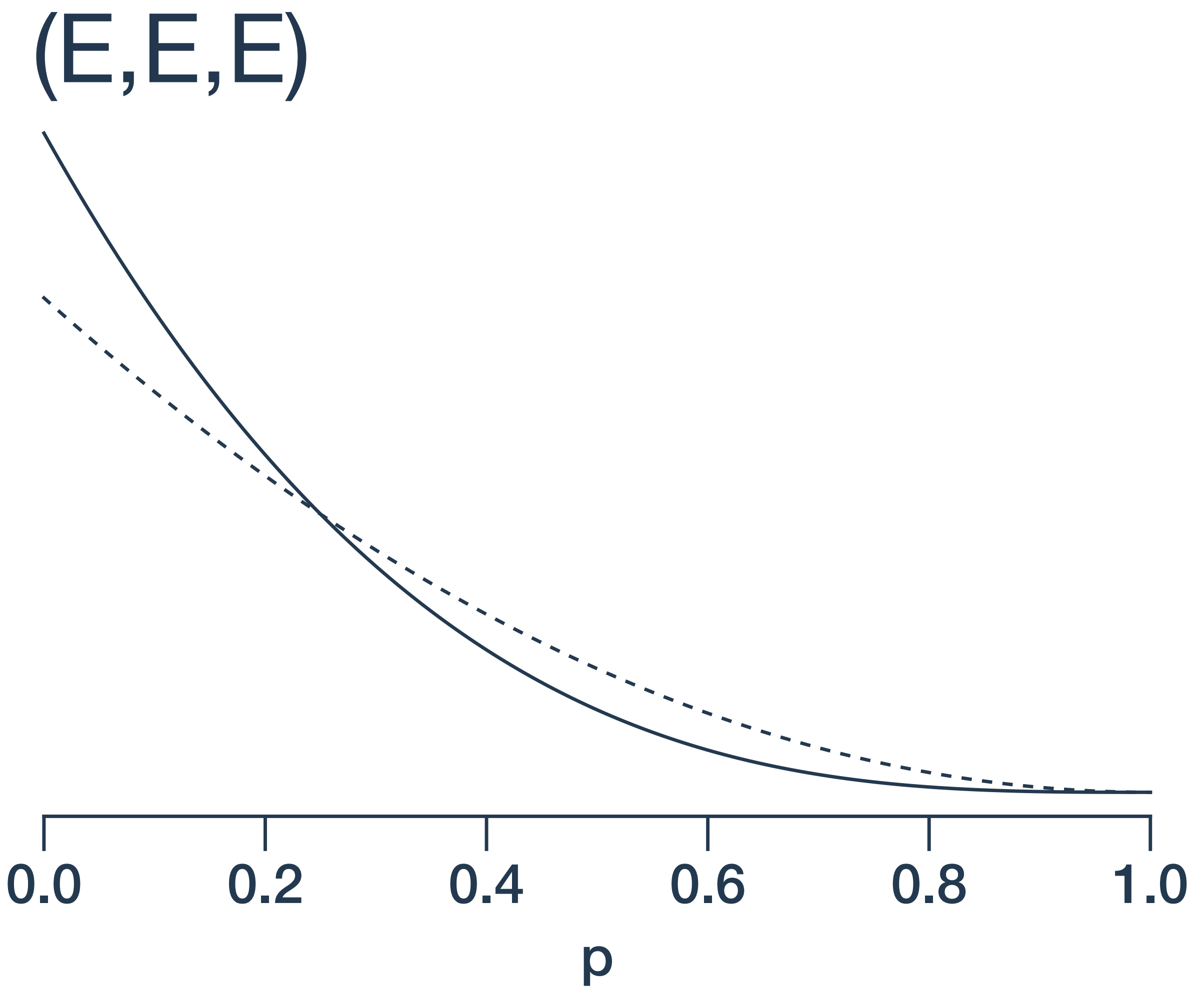


Estimating unemployment

$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

1. Pick a prior (such as uniform distribution)
2. Update prior with data (one at a time or all at once)



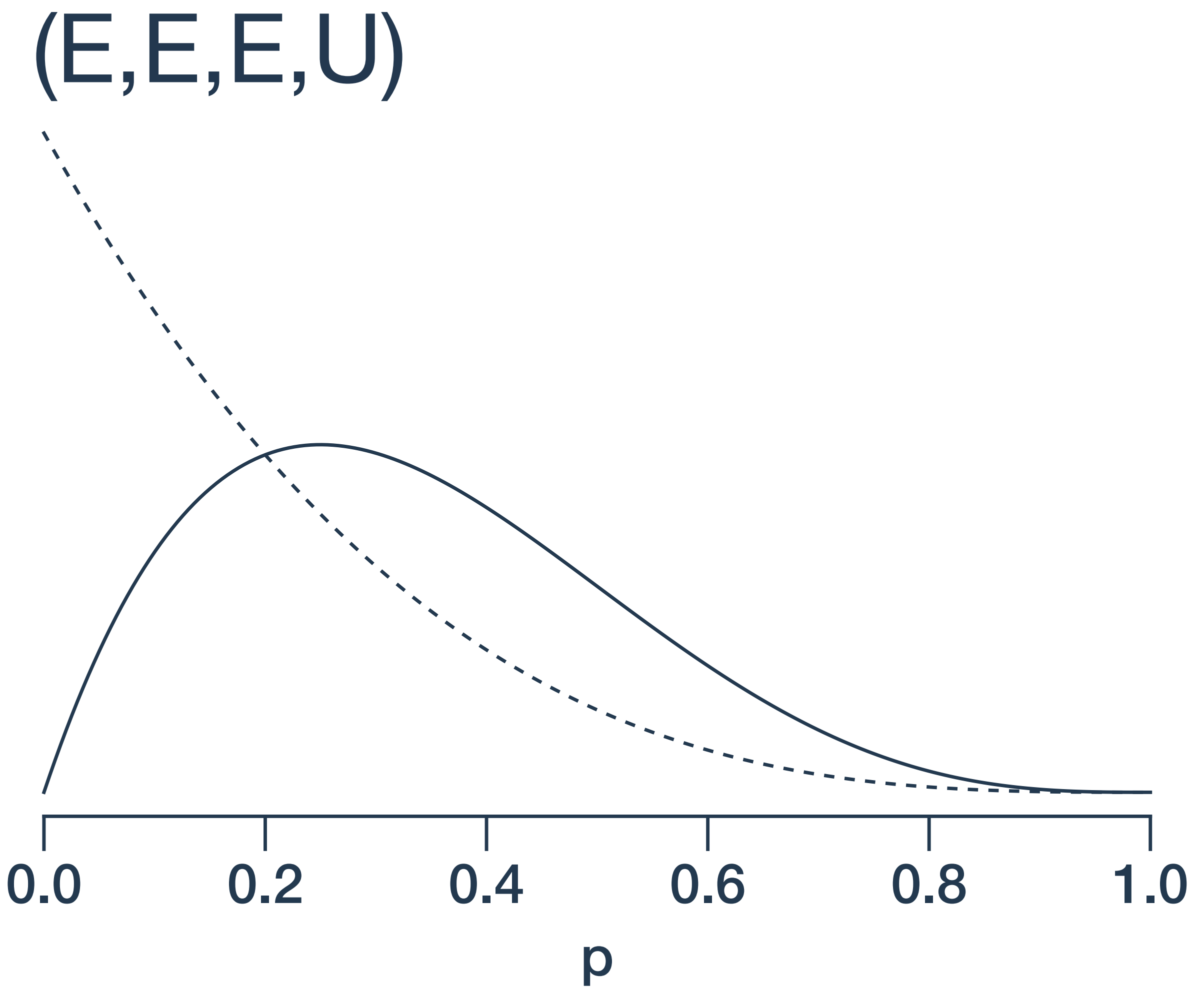
Estimating unemployment

$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

1. Pick a prior (such as uniform distribution)

2. Update prior with data (one at a time or all at once)



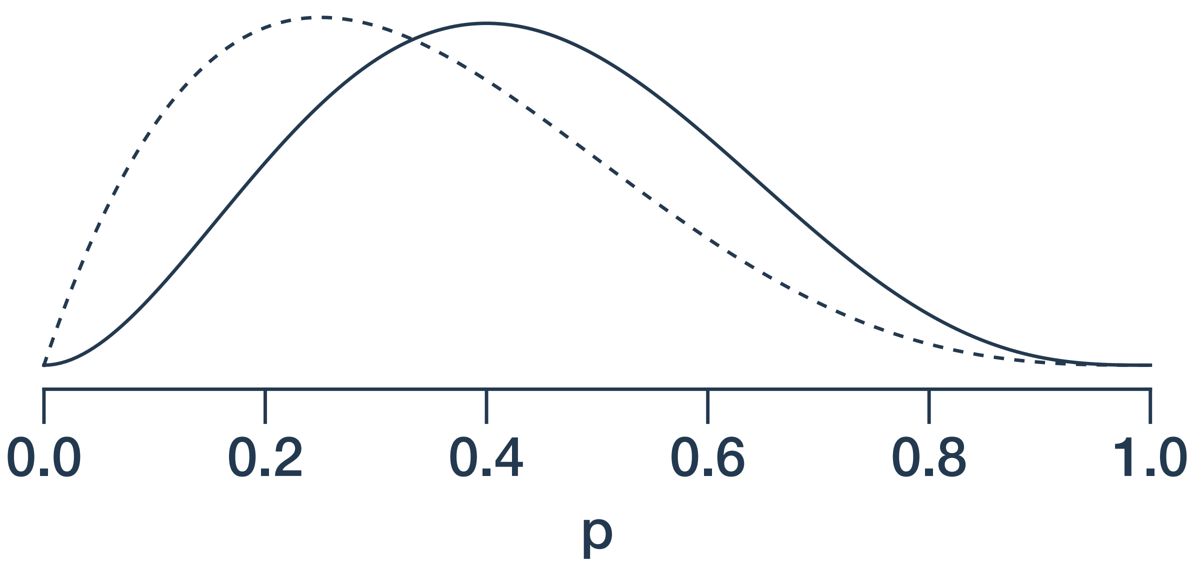
Estimating unemployment

$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

1. Pick a prior (such as uniform distribution)

2. Update prior with data (one at a time or all at once)



Estimating unemployment

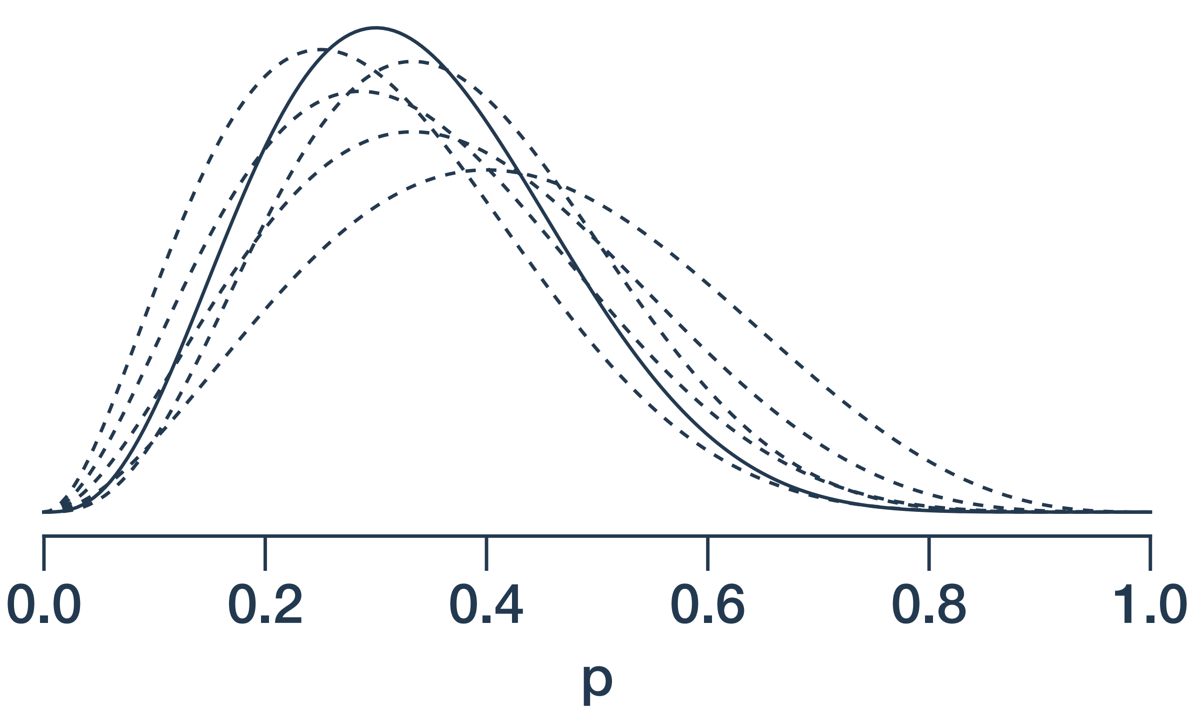
$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

1. Pick a prior (such as uniform distribution)

2. Update prior with data (one at a time or all at once)

(E,E,E,U,U,E,E,U,E)



Estimating unemployment

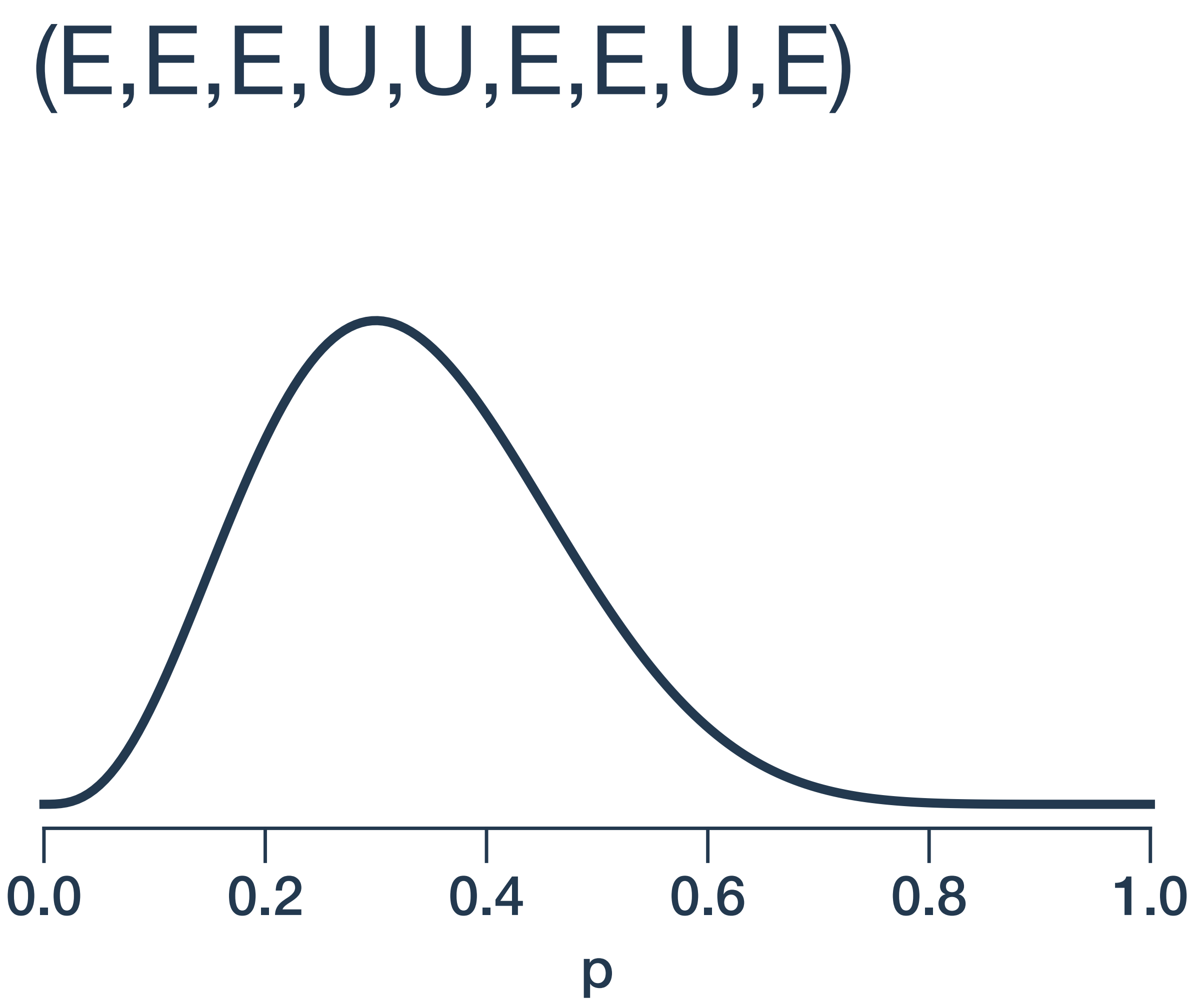
$$S = (E, E, E, U, U, E, E, E, U, E)$$

Bayesian estimation

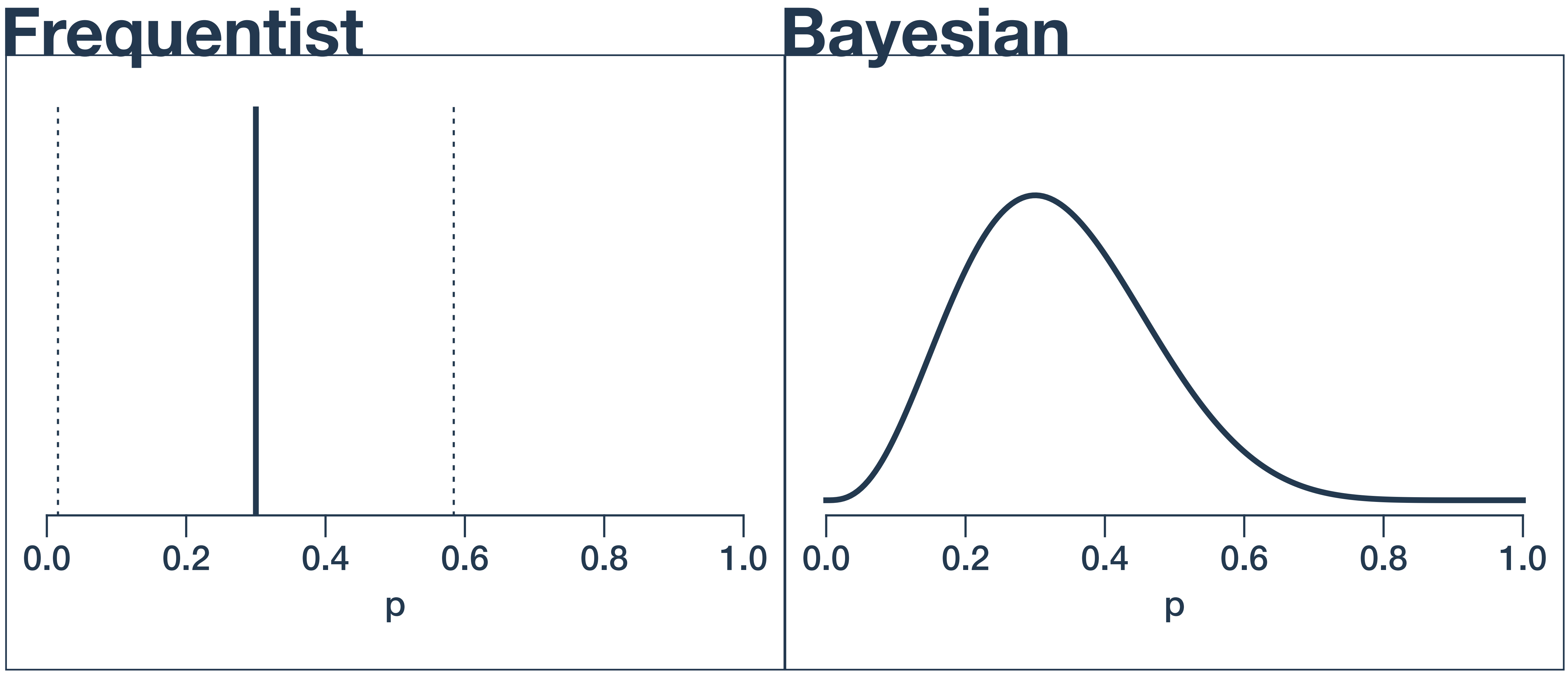
1. Pick a prior (such as uniform distribution)

2. Update prior with data (one at a time or all at once)

3. Posterior distribution incorporates all the information we have about *p*

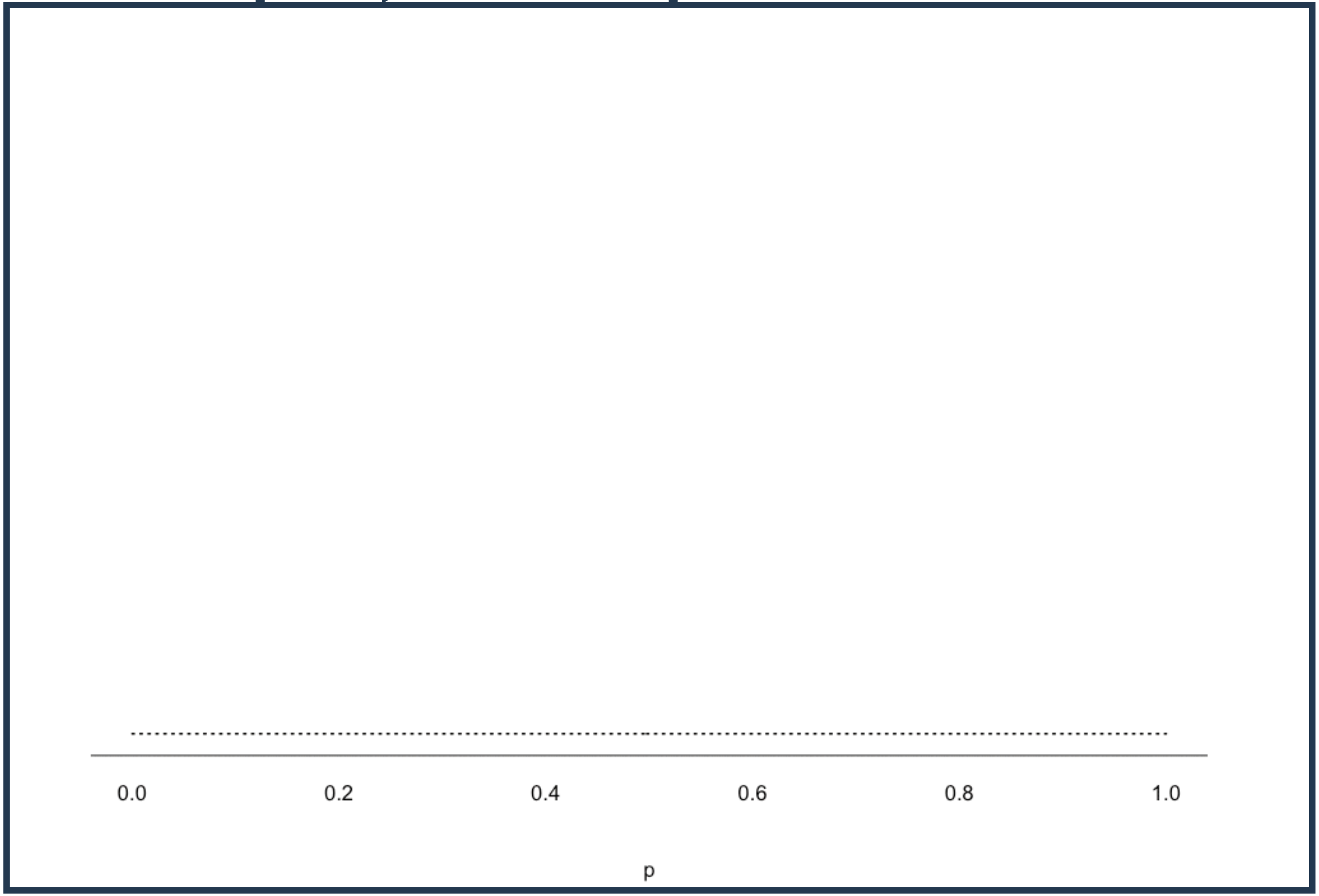


Comparing estimates



Bayesian updating

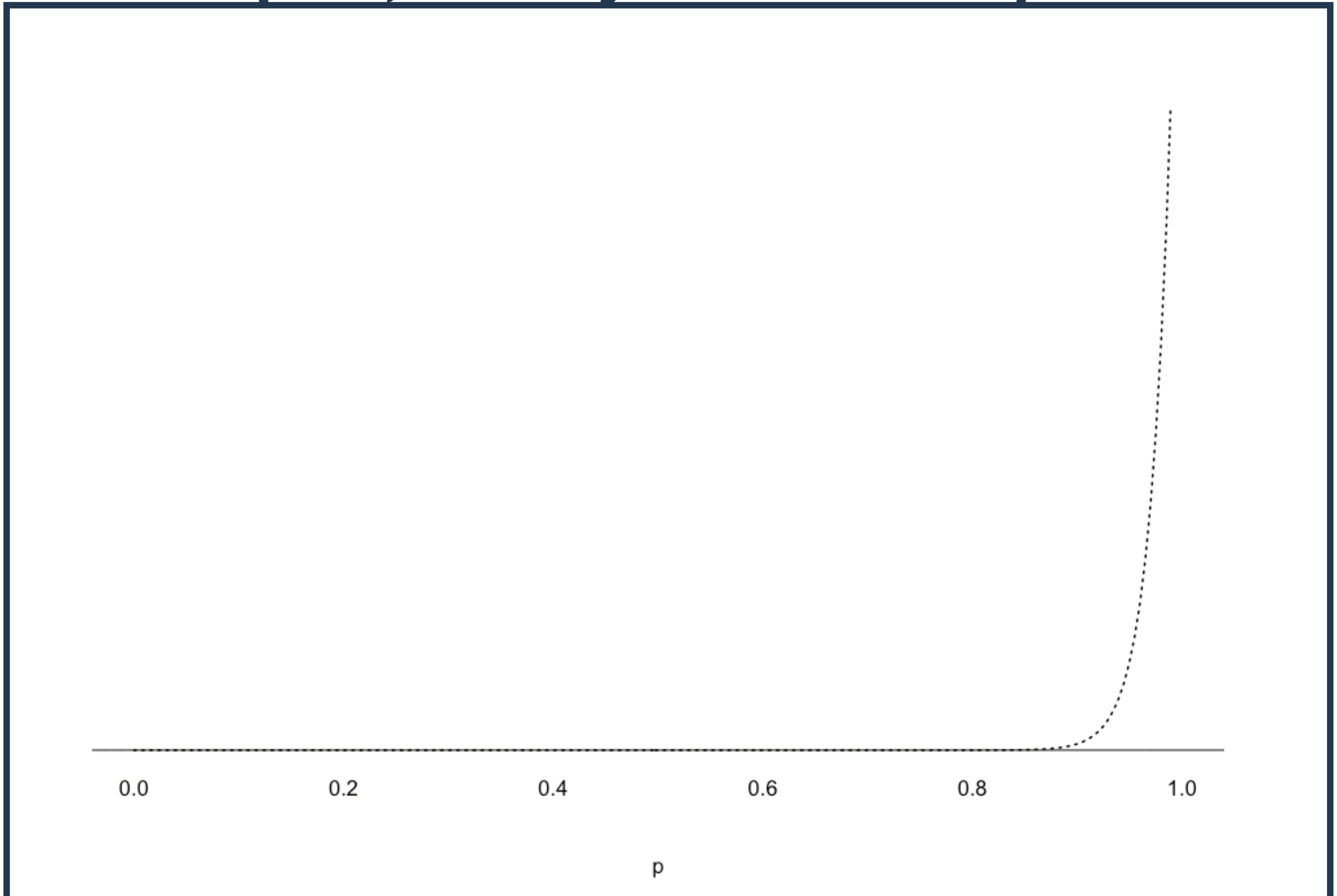
500 samples; uniform prior



https://soci620.netlify.app/media/p_uniformprior.gif

Bayesian updating

500 samples; heavily informative prior

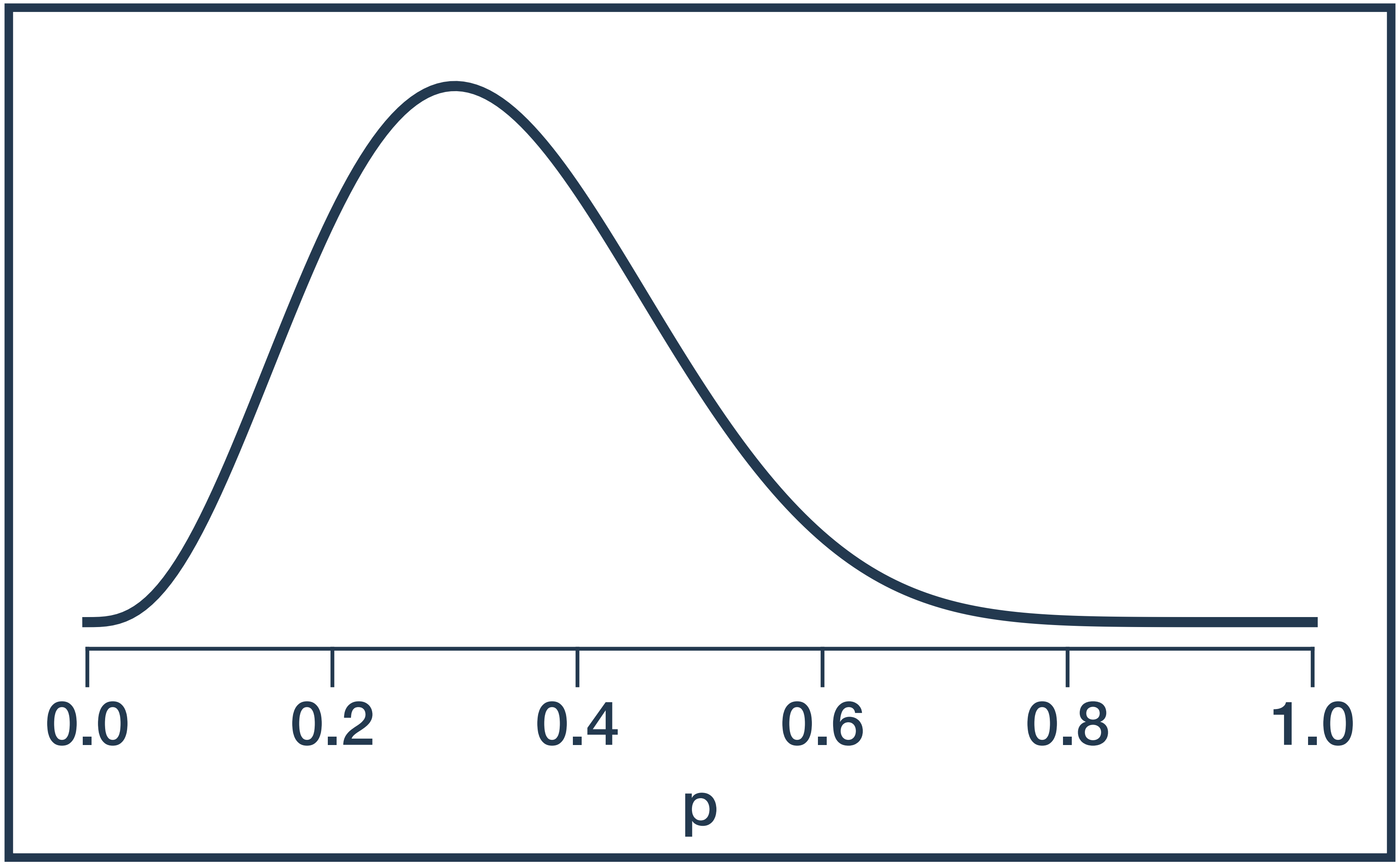


https://soci620.netlify.app/media/p_skewedprior.gif

Conditional probability

Varying → p **Fixed** → S

$$\Pr(p \mid S)$$



Bayes' rule

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

$$\Pr(p|S) = \frac{\Pr(S|p)\Pr(p)}{\Pr(S)}$$

Bayes' rule

Posterior probability

$$\boxed{\Pr(p|S)} = \frac{\Pr(S|p)\Pr(p)}{\Pr(S)}$$

Posterior probability is what we care about. It tells us everything we know about the unemployment rate (p) given what we've learned from our sample.

Bayes' rule

Prior probability

$$\Pr(p|S) = \frac{\Pr(S|p)\boxed{\Pr(p)}}{\Pr(S)}$$

The prior probability is everything we claim to know about the unemployment rate (p) *before* we ask anybody about their employment.

Bayes' rule

The evidence is just the average probability of seeing our sample across all possible values of p, normalizing our posterior. It is often the hardest part of a model to calculate (but fortunately we can usually ignore it).

$$\Pr(p|S) = \frac{\Pr(S|p)\Pr(p)}{\boxed{\Pr(S)}}$$

↑ **Evidence**

Bayes' rule

Likelihood

$$\Pr(p|S) = \frac{\boxed{\Pr(S|p)}\Pr(p)}{\Pr(S)}$$

The likelihood is where our *model* lives.

Unemployment

Building a model

- Pretend we already know the proportion, call it *p*.
- A probability model tells a story about what *S* might look like, assuming we know *p*.

- **Reverse the logic of your question:**

*In **reality** we know S and want to learn about p.*

$$\Pr(p|S)$$

*In **our model** we know p and want to describe S.*

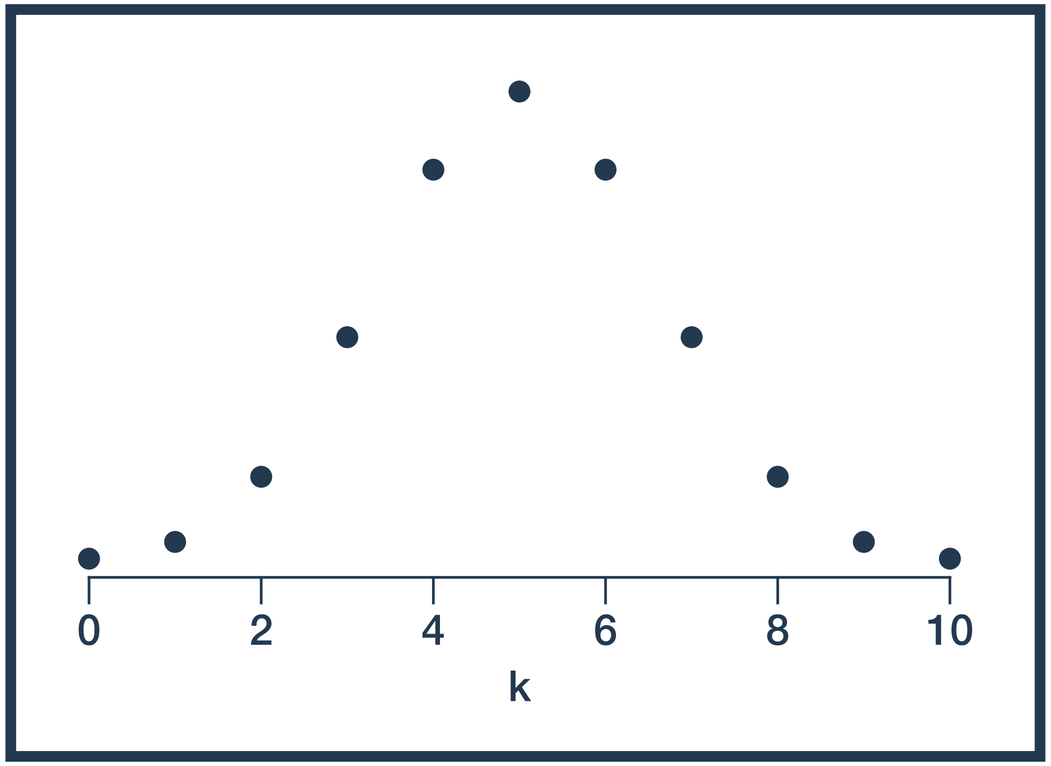
$$\Pr(S|p)$$

Binomial distribution

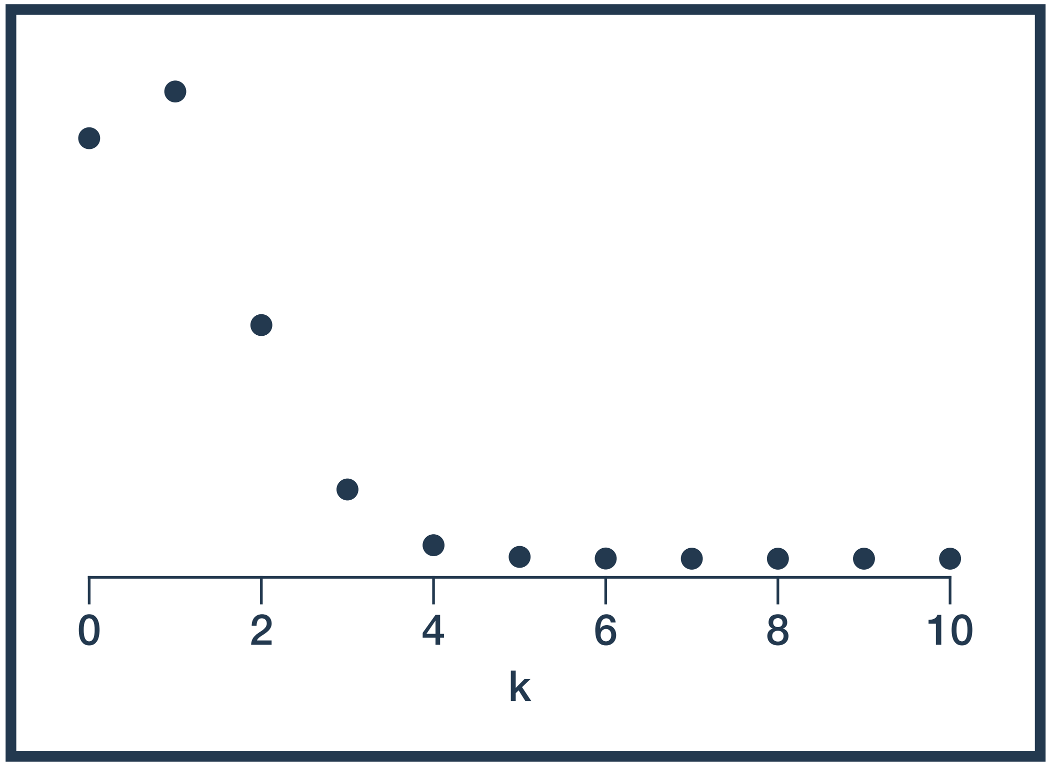
Binomial distribution
The probability of getting *k* 'successes' in *n* trials if the probability of success is *π*.

$$\Pr(k | n, \pi) = \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k}$$

***Bin*(*n*=10, *π*=0.5)**



***Bin*(*n*=10, *π*=0.1)**



Bayes' rule

Likelihood

$$\Pr(p|S) = \frac{\boxed{\Pr(S|p)}\Pr(p)}{\Pr(S)}$$

The likelihood is where our model lives. In this case, the probability of getting our sample (*S*), given the actual unemployment rate (*p*) is simply the binomial distribution we saw earlier: ***Bin*(*n*,*p*)**

Bayes' rule

Posterior probability

Likelihood

Prior probability

$$\Pr(p \mid S) = \frac{\Pr(S \mid p)\,\Pr(p)}{\Pr(S)}$$

Evidence

Bayes' rule

$$\Pr(p|S) \propto \Pr(S|p)\Pr(p)$$

$$\text{Posterior probability} \propto \text{Likelihood} \times \text{Prior probability}$$

Conditional probability

$$\Pr(p|S) \propto \Pr(S|p)\Pr(p)$$