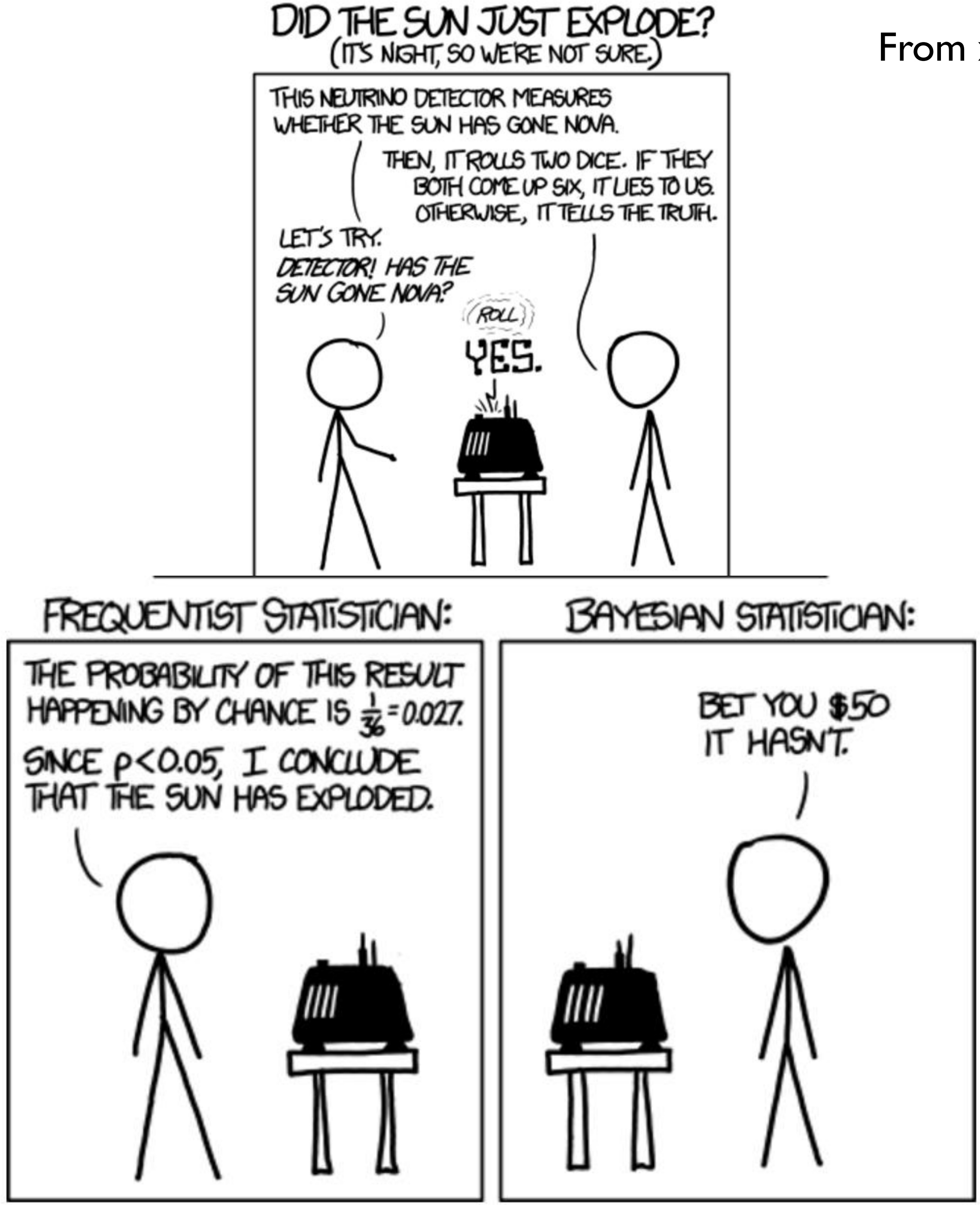


From xkcd.com

What is Bayesian Statistics?

Two different philosophies for statistics

Frequentist

Bayesian

Main difference is how they view population parameters (can view many problems in statistics in estimating and understanding parameters)

What is Bayesian Statistics?

- Bayes Theorem in Probability is the basis for Bayesian Statistics
- Bayes Theorem is named after Thomas Bayes (1701-1761)
- Thomas Bayes was an English Statistician and a Presbyterian minister

Review of Frequentist Ideology

- A population parameter is a FIXED, unknown quantity that we are trying to estimate.
- Estimate the population parameter with a point estimate (a statistic)

$$\mu \longrightarrow \bar{x}$$

$$p \longrightarrow \hat{p}$$

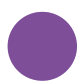
$$\beta_1 \longrightarrow b_1$$

How Bayesians view Parameters

- Remember, frequentist said that population parameters are *FIXED*, unknown numbers (true age of Americans is one number…don't know it, but we can get a sample to estimate it)
- Bayesians believe population parameters are *RANDOM VARIABLES*!!
- So, we are not estimating a number, we are estimating a distribution!!!! VERY different way of thinking about this!

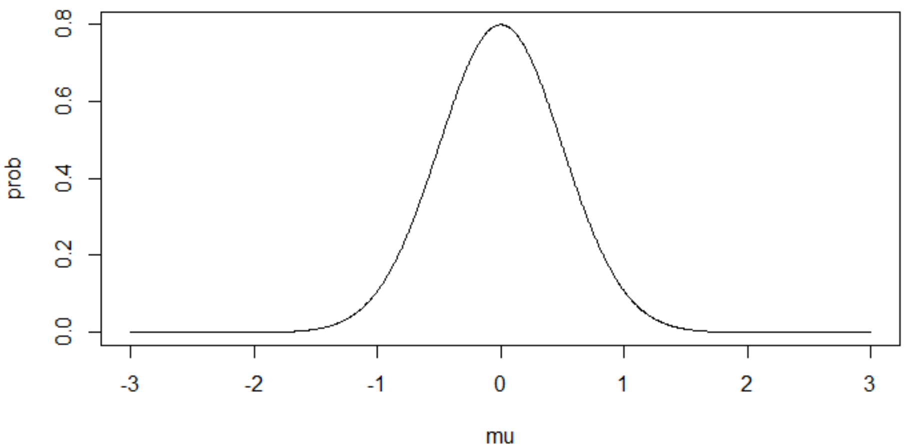
Frequentist versus Bayesian

μ



Frequentist

μ



Bayesian

Frequentist versus Bayesian

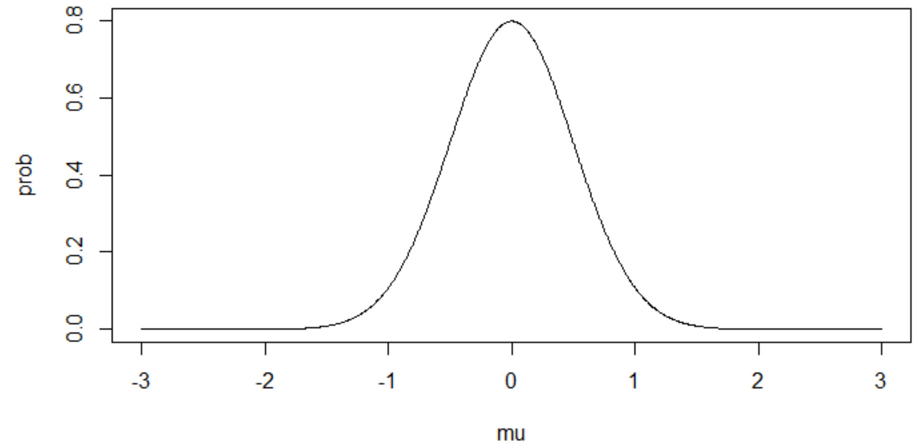
In both situations, we are using samples to obtain information about the population parameter

$\bar{x}$

Posterior distribution

μ

μ



Frequentist

Bayesian

Basics of Bayesian Statistics

The Basics (terminology)

- We need to assign a ***prior distribution*** to the parameter (can be an informative prior or noninformative or vague prior)
- The data that is collected is referred to as the ***sampling distribution***
- For example, let's say we looking at the mean age of Americans (μ)
- The prior distribution will be denoted by $p(\mu)$
- The sampling distribution will be denoted by ($p(Y|\mu)$
- We put this information together to get the ***posterior distribution*** ($p(\mu|Y)$)

One more time….

Prior

$p(\mu)$

Combined with data (sampling distribution)

$p(Y|\mu)$

Gives posterior

$p(\mu|Y)$

This is the goal

Accomplished through Bayes Theorem:

Prior distribution

Sampling distribution

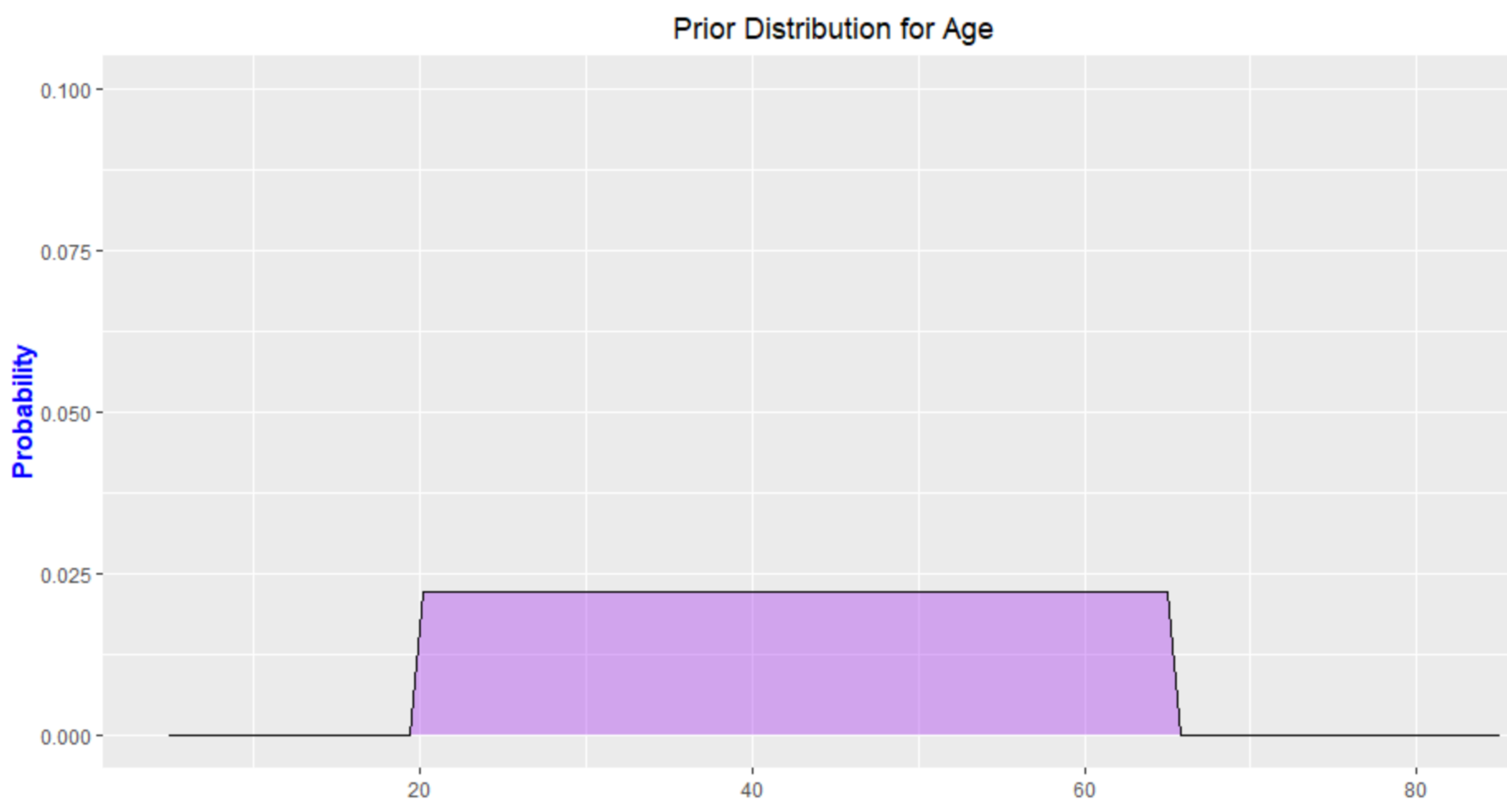
$$P(\mu|Y) = \frac{P(Y|\mu)P(\mu)}{P(Y)}$$

Normalizing constant

Example…

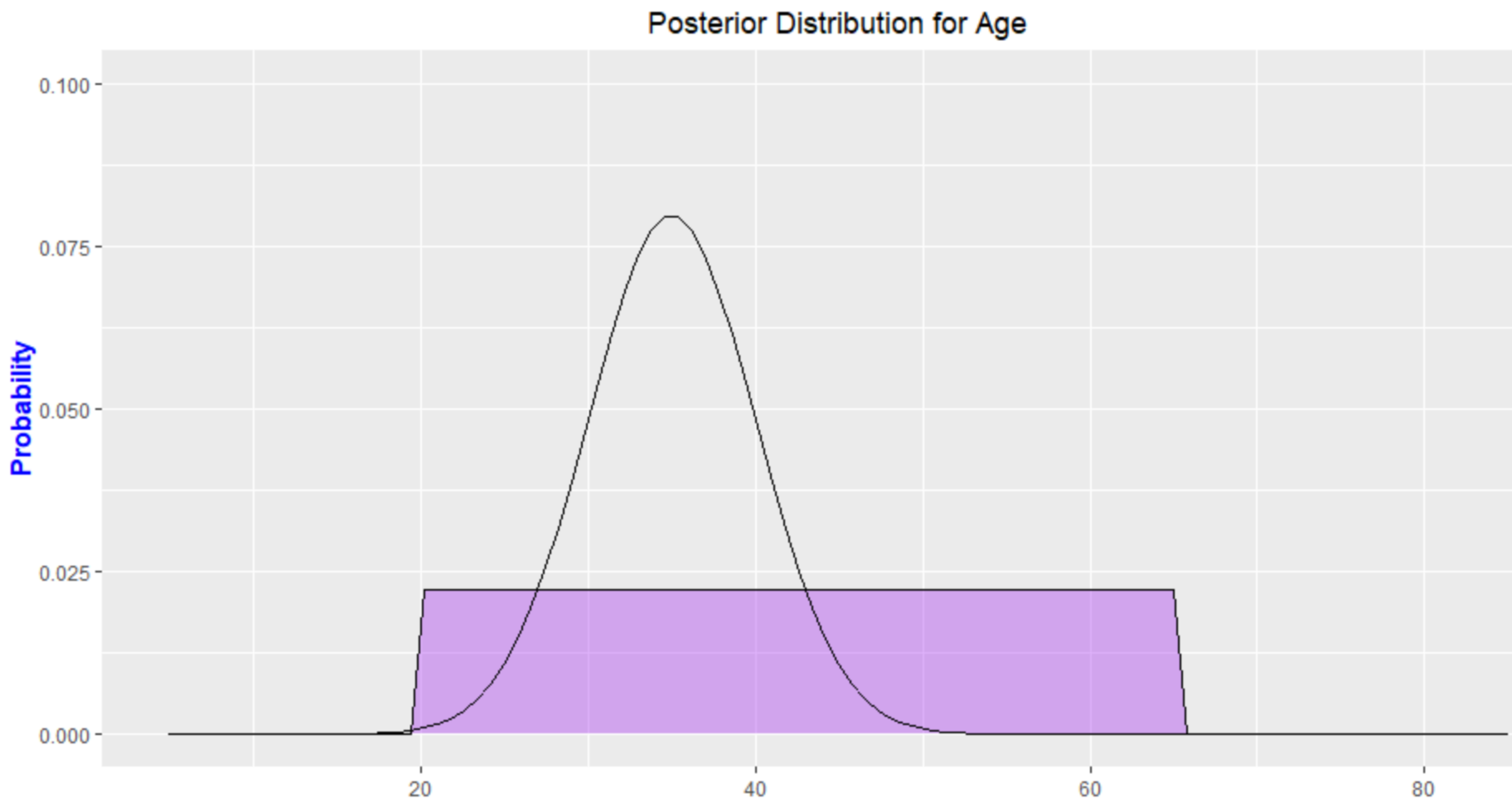
Average age of Americans (μ)

No idea of average age….Let's set Prior distribution for mean of Age (μ) as a Uniform(20,65)



Collect Data for the Sampling distribution

Put the Prior distribution and Sampling distribution together to get Posterior Distribution for mean of Age



How does this all work?

The process

STEP 1: Need to decide type of data being collected and decide what distribution the data will follow….this is very important because it defines the whole problem moving forward

Say for example, we can assume data being collected will follow a Normal distribution

STEP 2: What parameters are needed for this distribution? Pick appropriate priors.

Need to know a few basic distributions and their parameters

Posterior Distribution

STEP 3: Get data.

STEP 4: Use Sampling distribution (with data) and prior to get posterior distribution.

Most common sampling distributions:

Normal: continuous data that can take on negative and positive values

Gamma: continuous data that can only be positive (χ^2 is a special case of this)

Binomial: counting # success (number of females in a group)

Poisson: count data (number of accidents at an intersection)

Common priors:

- **<u>For Normal:</u>**
  - μ: Normal
  - σ^2: Gamma or Inverse-Gamma or Uniform
- **<u>For Gamma:</u>**
- α,β: Gamma or Inverse Gamma or Uniform
- **<u>For Binomial:</u>**
  - p: uniform(0,1) or Beta
- **<u>For Poisson:</u>**
  - λ: Gamma or Uniform
- You will need hyperparameters for each of these distributions. You can choose them so that these distributions are very spread out ($N(0,10^6)$)

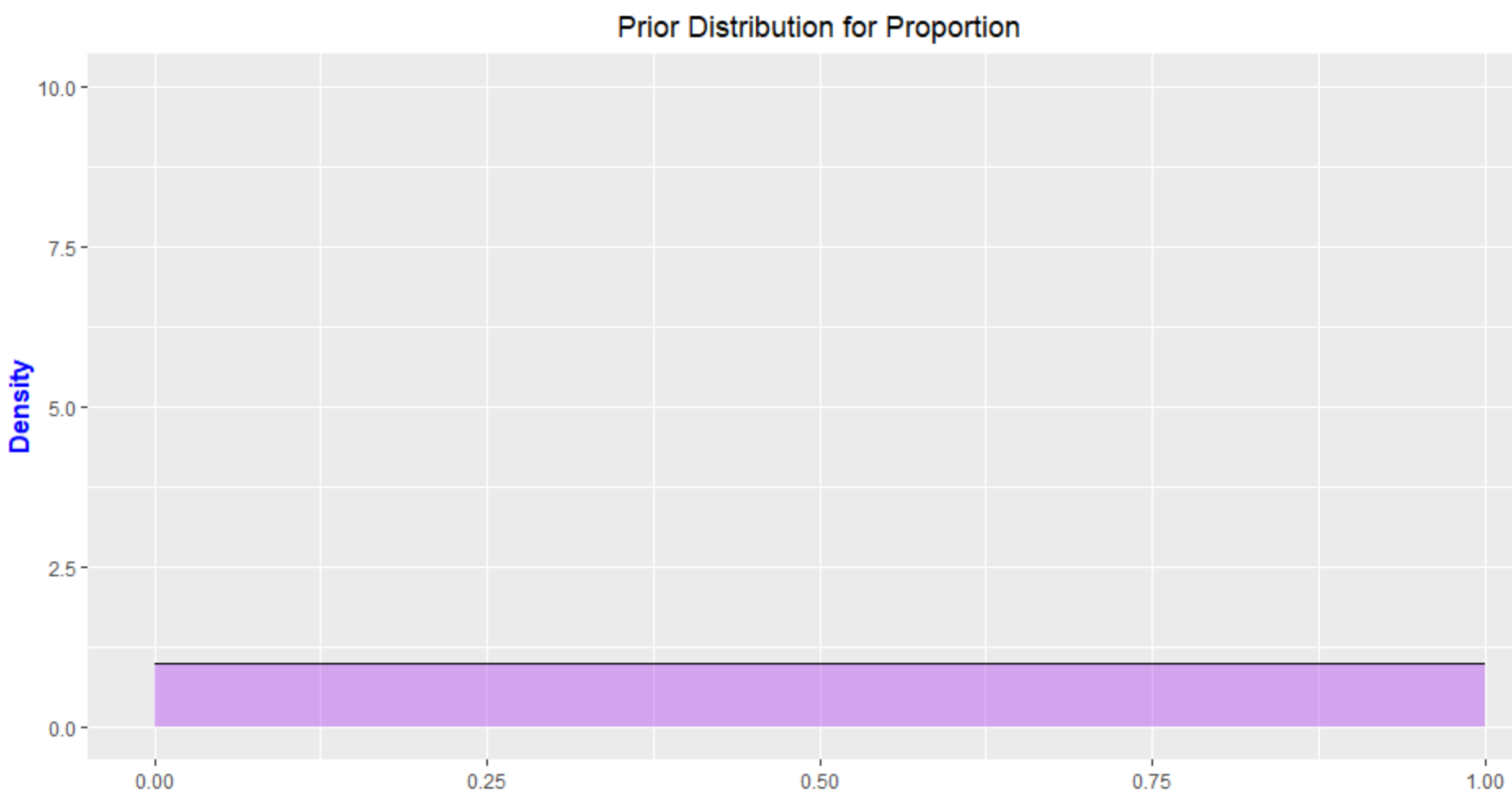
Let's try it:

Want to estimate proportion of Wake county residents who support a new bill being proposed. What type of data is being collected (distribution?)

Set up:

- Sampling distribution:
- How many parameters are there?
- Prior distribution for this parameter?

Assume non-informative prior

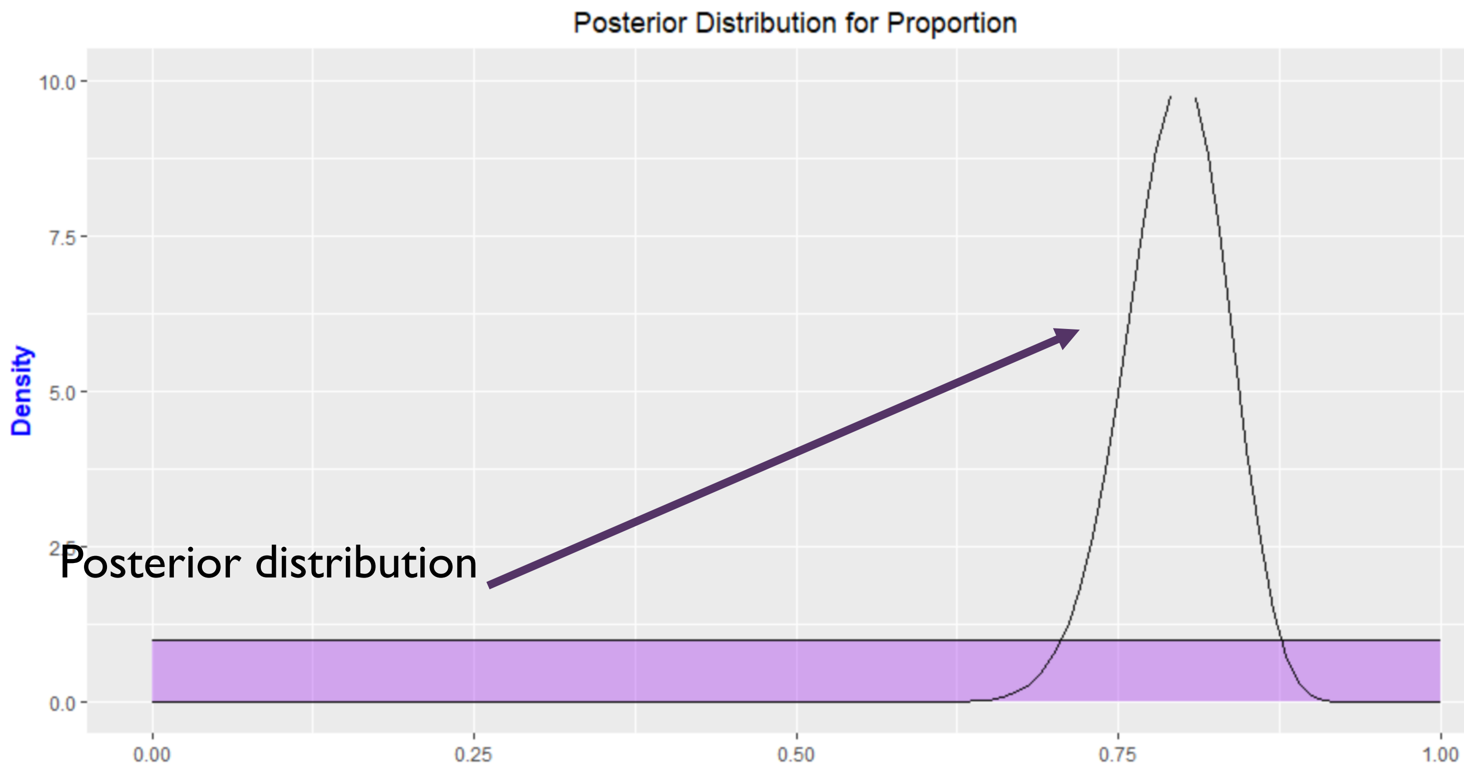


Data:

Assume out of a sample of 100, 80 supported the bill…

Posterior distribution:

Combining data (80 successes out of 100 trials and uniform(0,1) prior:



Can find some probabilities

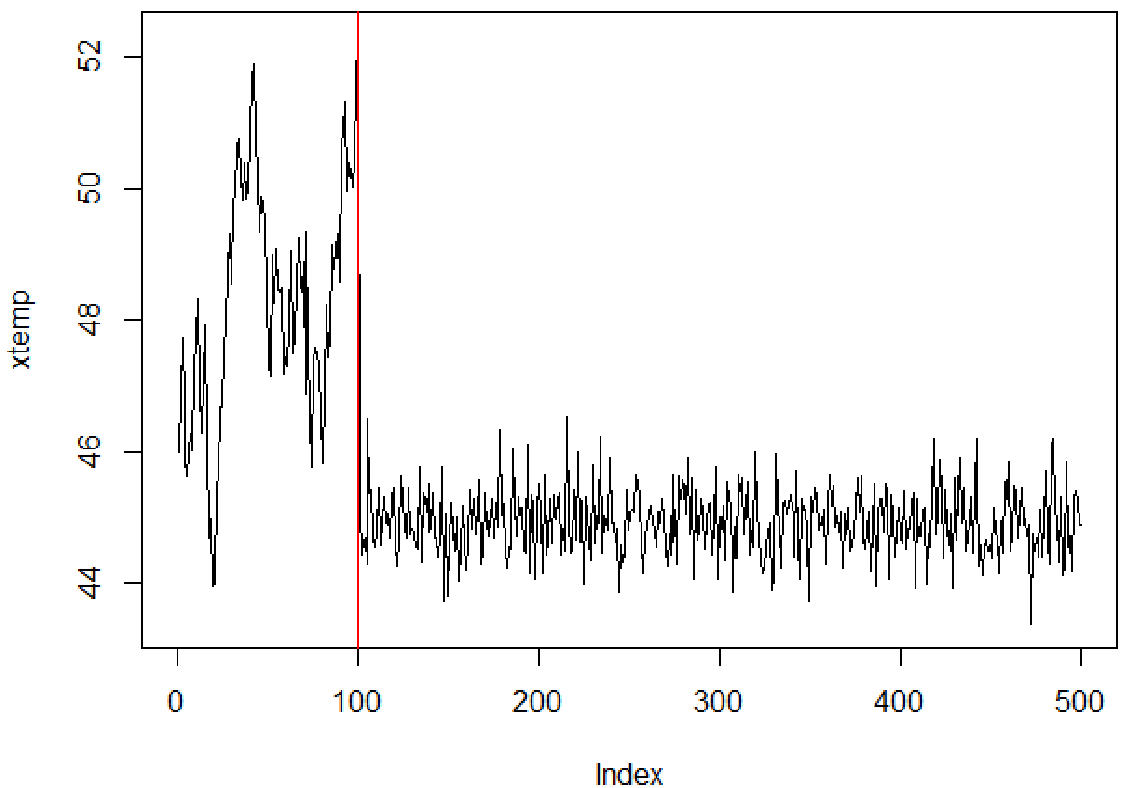
- $P(p>0.8)=0.46$
- $P(0.75<p<0.85)=0.79$

- 95% probability interval for p: 0.71,0.87

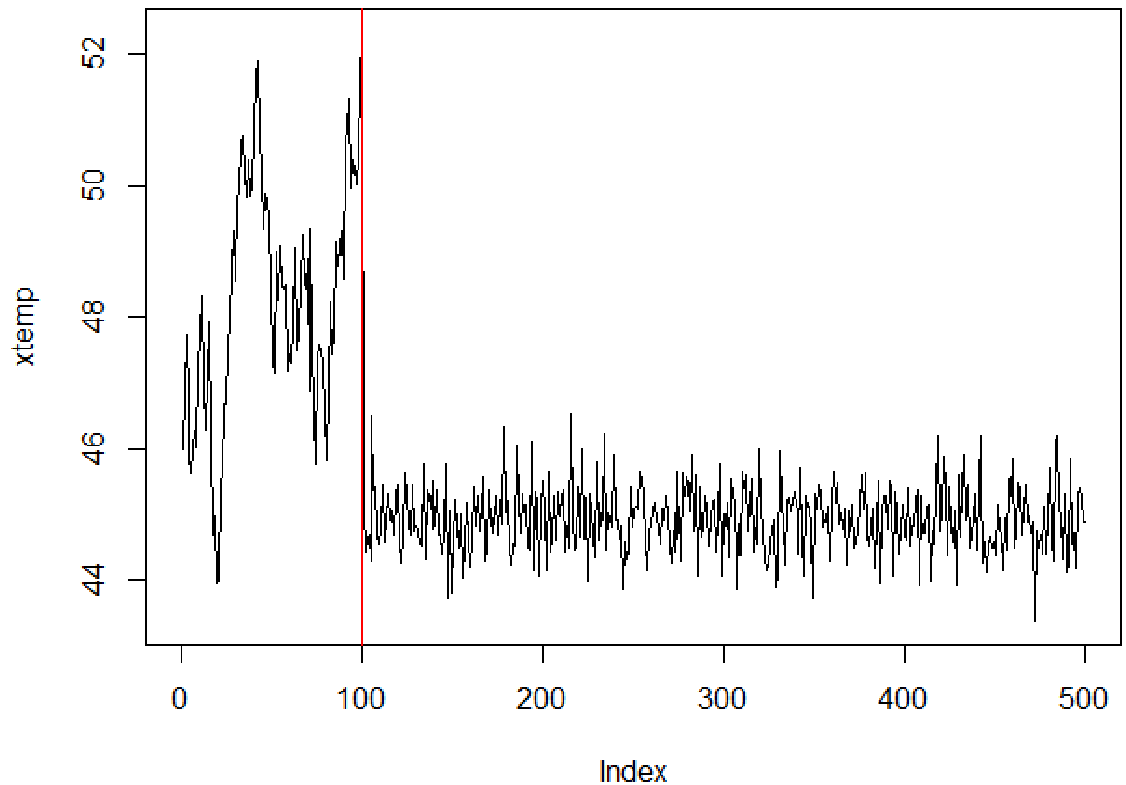
Posterior distribution

- With computers, we can simulate posterior distributions via Markov Chain Monte Carlo (MCMC) techniques

Want chain to "settle down" to posterior distribution



Chain finally settles down to the posterior distribution

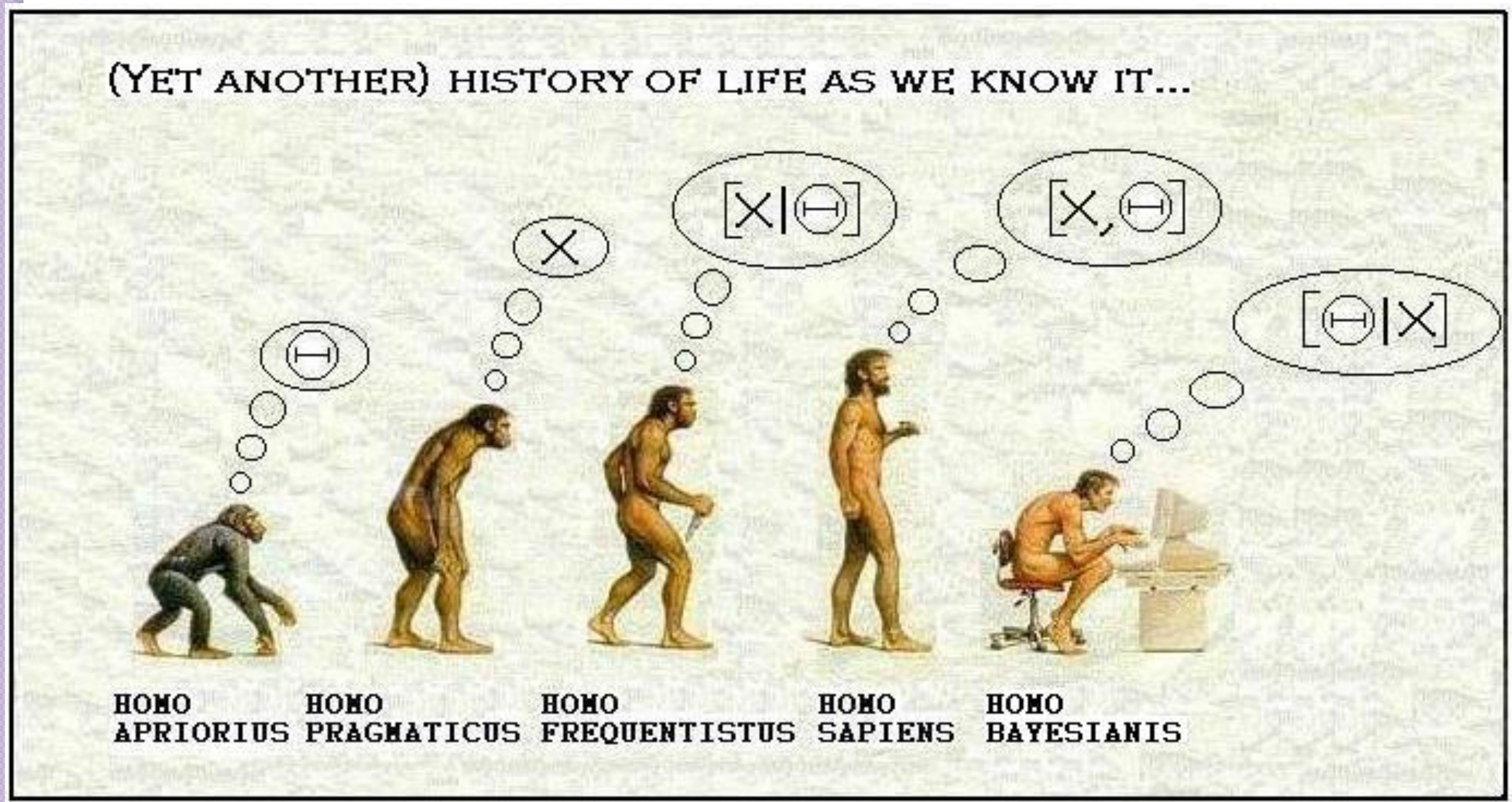


Burnin

Posterior samples

Common package to do this is STAN (STAN can also be run in Python):

- https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started



https://www2.isye.gatech.edu/~brani/isyebayes/jokes.html