BAYESIAN STATISTICS

CLASS 1

GOALS FOR TODAY'S LECTURE

- Understand terminology
  - Prior distribution
  - Sampling distribution
  - Posterior distribution
  - Marginal distribution of the data
- Understand simple coding in Stan
- Run a simple analysis in Stan

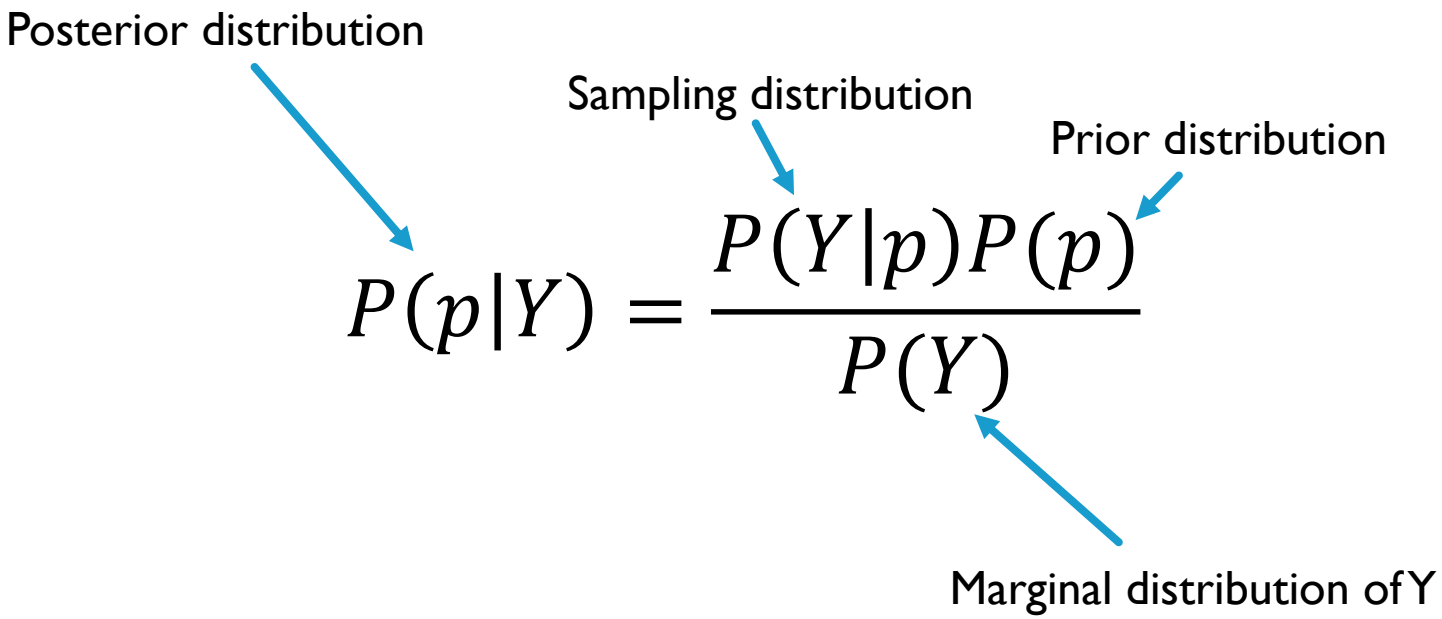
TERMINOLOGY

Sampling distribution: P(Y| parameters) (for example, $P(Y|\mu,\sigma^2)$ or P(Y|p) or $P(Y|\lambda)$)

Prior distribution: P(parameter) (for example, $P(\mu)$, $P(\sigma^2)$, P(p) or $P(\lambda)$)

Posterior distribution: P(parameters|Y) (for example, $P(\mu,\sigma^2|Y)$ or P(p|Y) or $P(\lambda|Y)$)

HOW IT ALL FITS TOGETHER (BAYES RULE)



HOW IT ALL FITS TOGETHER (BAYES RULE)

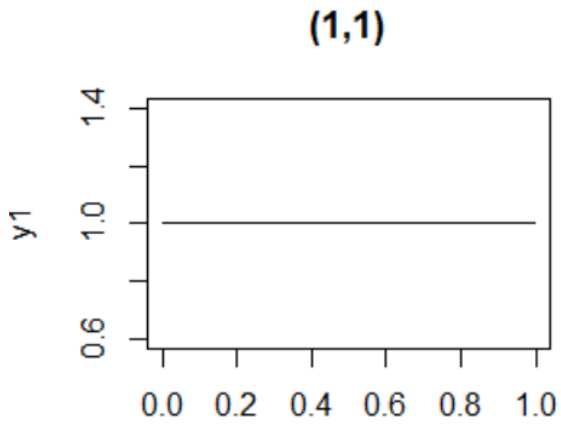
$$P(p|Y) \propto P(Y|p)P(p)$$

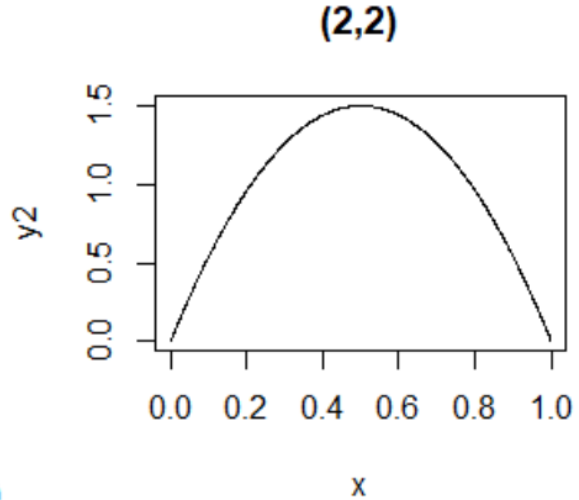
SIMPLE EXAMPLE

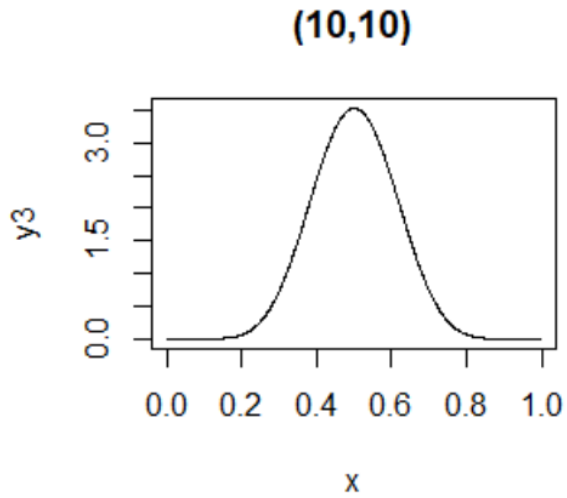
- Want to estimate the proportion of students at NCSU who voted in 2020 Democratic primary
- Sampling distribution:
- Parameters:
- Prior distribution:

EXPLORE THE BETA DISTRIBUTION

```
x<-seq(0.001,0.999,length=1000)
y1<-dbeta(x,1,1)
plot(x,y1,type='l', main='(1,1)')
y2<-dbeta(x,2,2)
plot(x,y2,type='l',main='(2,2)')
y3<-dbeta(x,10,10)
plot(x,y3,type='l',main='(10,10)')
```







MODEL INFO IN STAN

Data

Parameters

Transformed parameters (if needed)

Model

MODEL INFO IN STAN (SEPARATE FILE)

Data

This is where you define your data (integer, real, are there any bounds on information here?)

Parameters

Transformed parameters (if needed)

Model

MODEL INFO IN STAN (SEPARATE FILE)

Data

This is where you define your data (integer, real, are there any bounds on information here?)

Parameters

This is where you will define all of your parameters in the analysis (if not defined here, it will get confused)

Transformed parameters (if needed)

Model

MODEL INFO IN STAN (SEPARATE FILE)

Data

This is where you define your data (integer, real, are there any bounds on information here?)

Parameters

This is where you will define all of your parameters in the analysis (if not defined here, it will get confused)

Transformed parameters (if needed)

If you are transforming variables…for example in a linear regression

Model

MODEL INFO IN STAN (SEPARATE FILE)

Data

This is where you define your data (integer, real, are there any bounds on information here?)

Parameters

This is where you will define all of your parameters in the analysis (if not defined here, it will get confused)

Transformed parameters (if needed)

If you are transforming variables…for example in a linear regression

Model

This is where you will define your model

Be sure that your Stan programs ends in a blank line without any characters including spaces and comments.
Statements will end in semicolins;

CODE FOR EXAMPLE

```
data {
      int <lower=0> y;
      int <lower=0> n;
      }
parameters {
      real <lower=0, upper=1> p;
                  }
model {
      p ~ beta(1,1);
      y ~ binomial(n, p);
         }
```

NOTE: Be sure to save this file as .stan (not .txt…careful to make sure it is NOT saved as .txt)

IN CLASS EXAMPLE

- A political science student wants to estimate the proportion of students at NCSU who voted in the 2020 election. Identify the sampling distribution, the number of parameters and potential prior(s).
- The student gathered a sample of 150 students of which 100 indicated that they did vote.
- Get the posterior distribution(s) of the parameter(s) and find 95% probability interval(s). Assume a uniform prior for p.