# ISTA410/INFO510 Bayesian Modelling and Inference

Lecture 2 – Introduction to Bayesian modelling, Part 2

Dr. Kunal Arekar
College of Information
University of Arizona, Tucson

August 27, 2025

**A Bayesian and a Frequentist were to be executed. The judge asked them what were their last wishes.**
**The Bayesian replied that he would like to give the Frequentist one more lecture.**
**The judge granted the Bayesian's wish and then turned to the Frequentist for his last wish.**
**The Frequentist quickly responded that he wished to hear the lecture again and again and again and again…**

# Example analyses:

In the population of these colorful balls, determine the percentage of red-colored balls – either 10% or 20%

Collect data – able to collect less sample due to limited budget (resources); data collection can be costly

Higher sample size = more trustworthy decision
But let's go ahead with less sample size (limited budget)

Sample size (n) = 5 balls (G, Y, P, B, R)
k = 1; no. of red balls

# Example analyses:

**Frequentist Method:**
Declare the null and alternative hypothesis.

Calculate the p-value and compare against the desired significance level.

Determine the valid hypothesis.

$H_0$ = 10% red balls
$H_A$ = > 10% red balls

Significance Level — 0.05

$$\begin{aligned} P\text{-value} &= P(K \geq 1 \mid n=5, p=0.10 \\ &= 1 - P(k=0 \mid n=5, p=0.10) \\ &= 1 - 0.90^5 \\ &= 1 - 0.59 \\ &= 0.41 \end{aligned}$$

*High P value; failure to reject the $H_0$

# Example analyses:

**Bayesian Method:**

Evaluate the probabilities of both these models, as opposed to having to choose one of as our null and the other as alternative.

Hypotheses; $H_1$ = 10% red balls
$H_2$ = 20% red balls

Prior $P(H_1) = 0.5$
$P(H_2) = 0.5$

*Since we don't have a reason to believe that one is more likely than the other, our priors would be with equal probability.*

Likelihood:

$$P(k=1 \mid H_1) = \binom{5}{1} 0.10 \times 0.90^4 \approx 0.33$$

$$P(k=1 \mid H_2) = \binom{5}{1} 0.20 \times 0.80^4 \approx 0.41$$

Posterior for H1:

$$P(H_1 \mid k=1) = \frac{P(H_1) \times P(k=1 \mid H_1)}{P(k=1)} = \frac{0.5 \times 0.33}{0.5 \times 0.33 + 0.5 \times 0.41} = 0.45$$

Posterior for H2: $P(H_2 \mid k=1) = 0.55$

| | ***Frequentist*** | ***Bayesian*** | |
|---|---|---|---|
| ***Obs. Data*** | ***P(k or more \| 10% red)*** | ***P(10% red \| n,k)*** | ***P(20% red \| n,k)*** |
| n =5, k=1 | 0.41 | 0.45 | 0.55 |
| n=10, k=2 | 0.26 | 0.39 | 0.61 |
| n=15, k=3 | 0.18 | 0.34 | 0.66 |
| n=20, k=4 | 0.13 | 0.29 | 0.71 |

If we increase the sample size, every time the frequentist approach fails to reject the null hypothesis.

Bayesian method always yields a higher posterior for the second model where P is equal to 0.20

# Frequentist vs Bayesian approach to probability

**Frequentist:** Relative frequency of events in a hypothetical sequence of events.
E.g., Rolling fair six-sided die infinite number of times, P(x=4) will be 1 in 6 or 1/6
P(drop packets) = 1/10000; if we lose 1 in 10000 packets

**Bayesian:** the probability represents your own perspective, it's your measure of uncertainty; it takes into account what you know about a particular problem. You may have a particular information about the events in question that help you change your perspective about it.
E.g., P(fair) – probability that the die is fair. If you have some information about fairness of the die, then your probability might differ from someone else's without that information

# Frequentist approach

## Key features

- Objective Probability - probabilities are seen as inherent properties of the phenomena under consideration; objective chance of an event based on how often it happens
- Parameter Estimation - parameters (characteristics of a population) are assumed to be fixed and unknown constants.
- Hypothesis Testing - null and alternative hypotheses; p-value is a central concept in hypothesis testing.
- Confidence Intervals - Confidence intervals provide a range of values within which a population parameter is likely to fall with a certain level of confidence.

## Limitations and criticism

- Subjectivity in Null Hypothesis Choice - different researchers may formulate different null hypotheses for the same research question.
- Interpretational Challenges - Confidence intervals can be misinterpreted, leading to a misunderstanding of the probabilistic nature of these intervals.
- Fixed Parameter Assumption - population parameters are fixed and do not have probability distributions.

# Bayesian approach

## Key features

- Subjective Probability - reflects an individual's degree of belief in the occurrence of an event.
- Bayes' Theorem - provides a framework for updating beliefs in light of new evidence.
- Prior Distribution - representing beliefs about the parameter before observing any data
- Posterior Distributions – obtained using Bayes' Theorem by combining prior beliefs with the likelihood of the observed data.
- Parameter Estimation - provides a probability distribution for the parameter of interest rather than a point estimate.

## Limitations and criticism

- Computational Challenges: Intensive; might take weeks, in rare cases months, to finish the analysis

- Impact of Priors: Subjective choice of priors, sensitivity analyses can be conducted to assess the impact of different prior choices. – to test the robustness of the Bayesian analysis.

# Difference between Frequentist and Bayesian

**Philosophical Differences:**

- Probability Interpretation: Frequentists see probability as the objective chance of an event based on how often it happens, while Bayesians view it as a personal belief about how likely an event is.
- Parameter Approach: Frequentists consider parameters as fixed but unknown, while Bayesians treat them as uncertain and use probability to describe this uncertainty.
- Hypothesis Testing: Frequentists test how likely the data is if a certain hypothesis is true, whereas Bayesians compare how likely different hypotheses are given the data.

**Practical Differences:**

- Estimation: Frequentists provide single value estimates and confidence ranges for population parameters, while Bayesians offer full distributions that show uncertainty.
- Decision Making: Frequentists use p-values to make decisions about hypotheses, whereas Bayesians base decisions on the probability distribution of the parameter.
- Priors: Bayesians include prior information, which can be detailed or vague, reflecting existing knowledge or uncertainty.

**Interpretational Differences:**

- Confidence vs Credible intervals: Frequentist confidence intervals show the range where the true parameter would fall in repeated samples, while Bayesian credible intervals show the most likely range for the parameter based on the data.
- p-values vs. Posterior Probabilities: Frequentists use p-values to assess how unusual the data is under the null hypothesis, while Bayesians use posterior probabilities to measure how likely a hypothesis is given the data.

*The decision to use frequentist or Bayesian methods usually hinges on the specific problem at hand, the presence of prior knowledge, and the researcher's underlying philosophical perspective.*

## Three pieces of the Bayesian model

*Likelihood*

*prior probability*

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|A')P(A')}$$

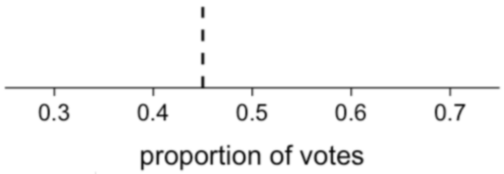
*sum of the product of all probabilities of mutually exclusive hypotheses and corresponding conditional probabilities*

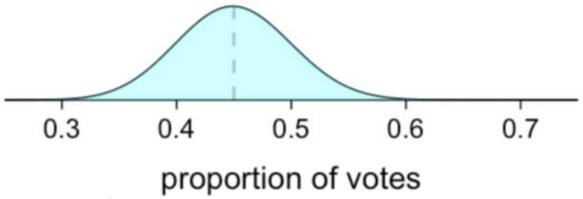
*Posterior probability*

Suppose you are running for an election at a public office

Older data suggest you have support from 45% voters

Uncertain due polling errors and fluctuation in support

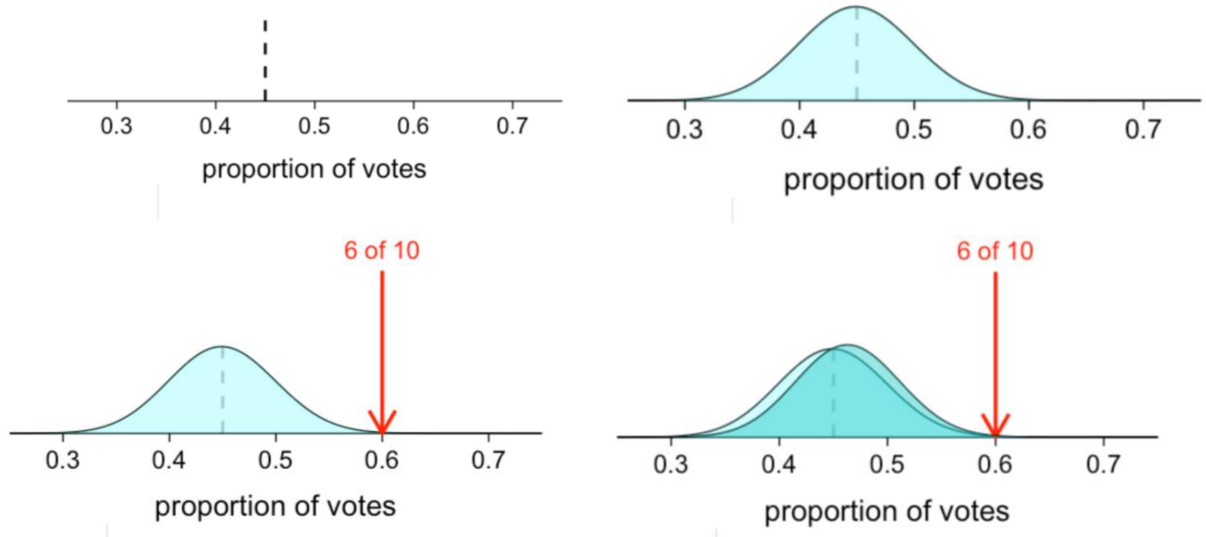


To gain batter insight, you conduct a small poll of 10 voters

Among those, 60% plan to vote for you

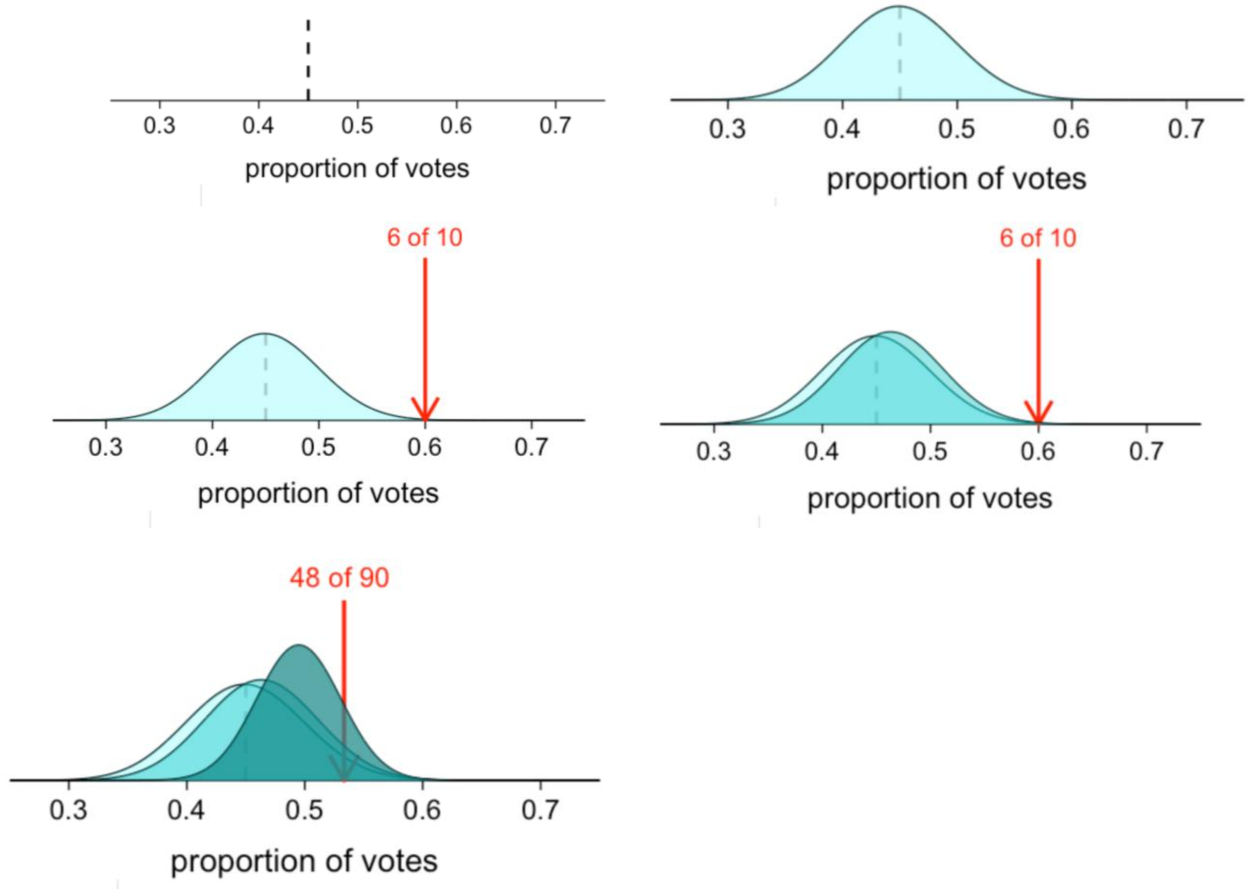
Updated posterior



You continue to collect data;
New poll 48 of 90 (53%) plan to vote for you

In the light of this new data, update posterior

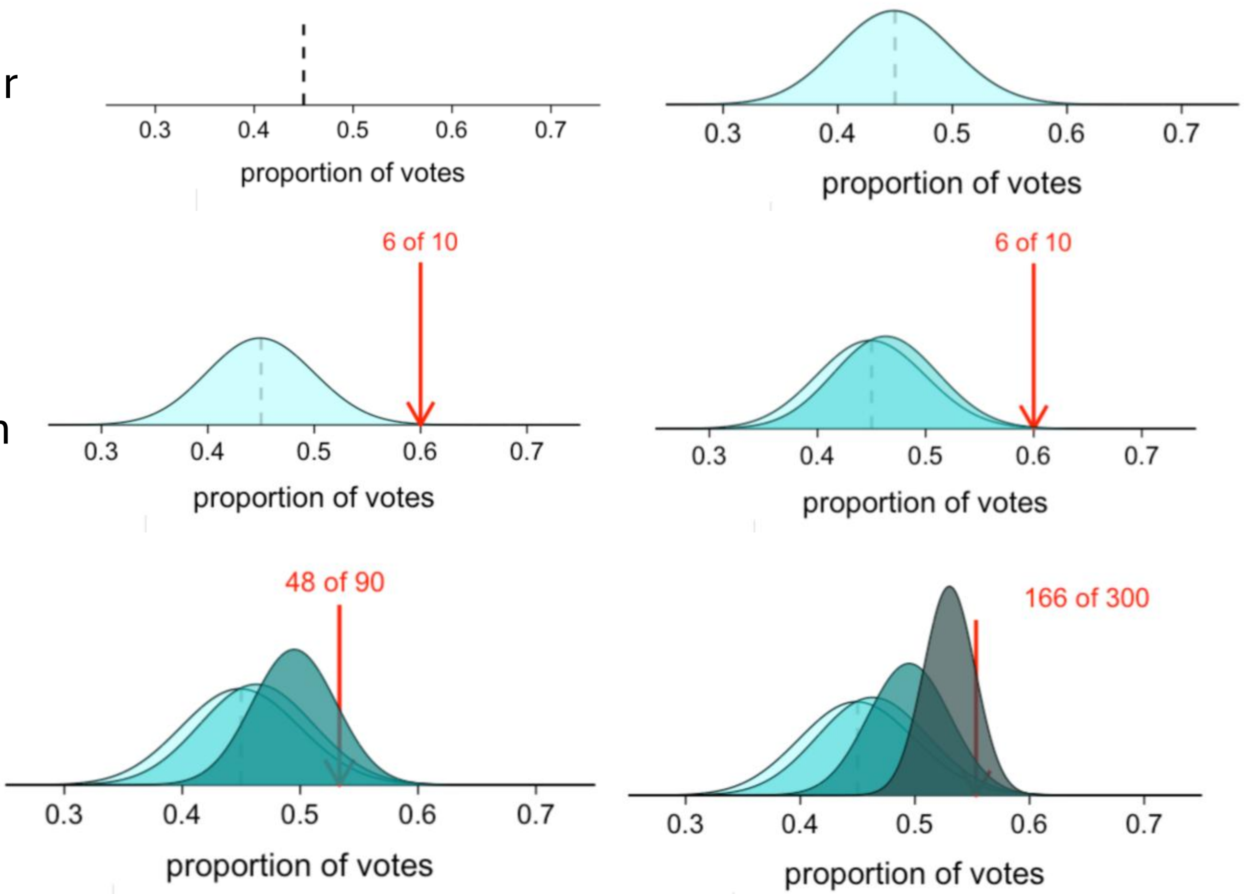
Posterior optimism about your election chances increase



You decide to do a final poll

166 of 300 (55%) plan to vote for you

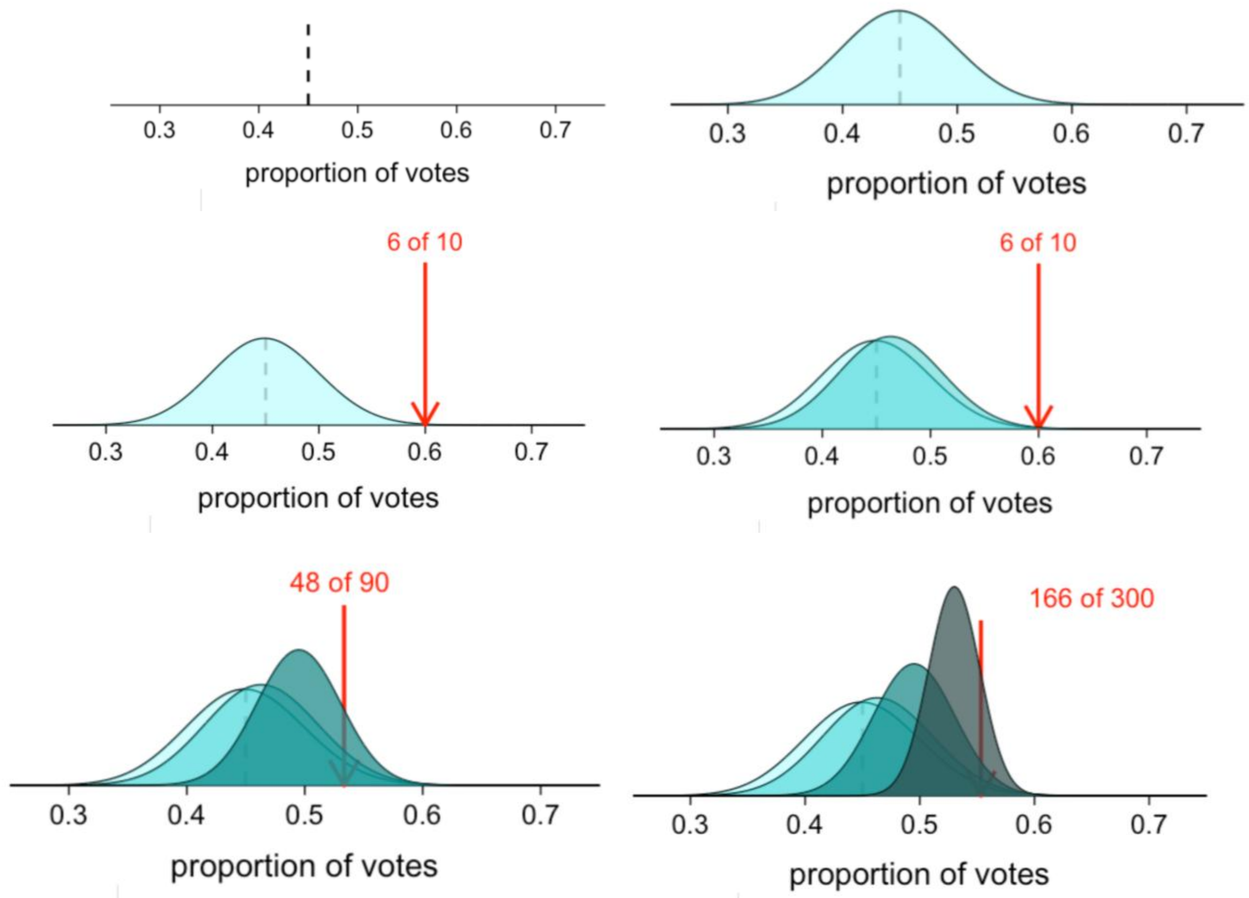
With such large sample size, more confident about the posterior probability aka chances of winning the election



This example highlights the power of Bayesian models

Combines insights from prior model and observe data

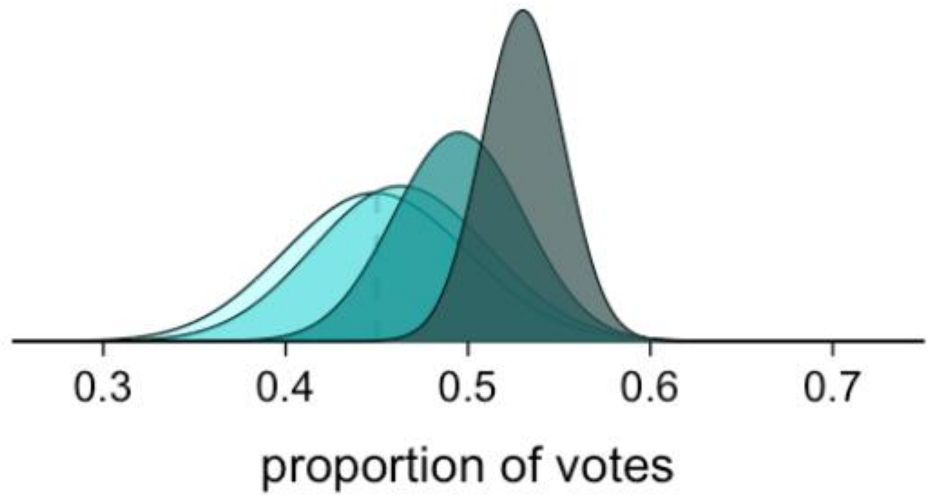
Updates as new data come in



**Bayesian posterior model**

- Combines insights from the prior model and observed data

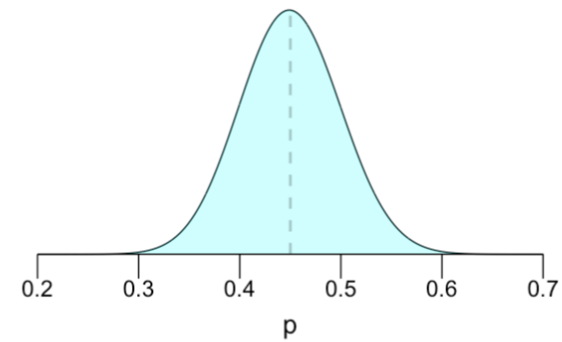
- Evolves as new data come in



# Three pieces of the Bayesian model

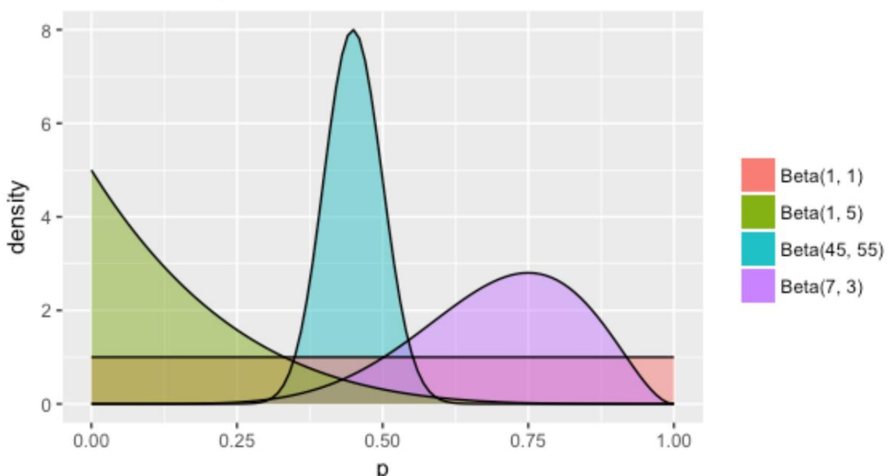
**Prior:**

P = proportion of a prior event (e.g. red balls, voters that support you)
P is between 0 and 1




Treating parameter P as a random variable
Thus, the prior model of P is simply a probability distribution

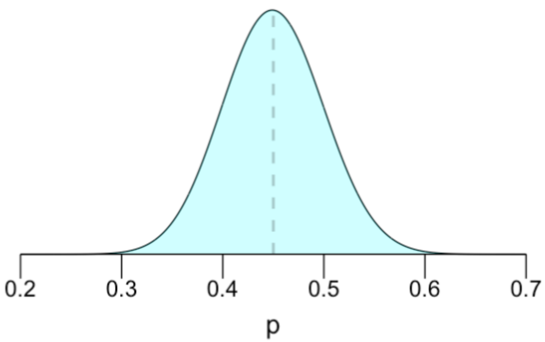
Beta distribution with shape parameters 45 and 55
$P \sim \text{Beta}(45,55)$




Tuning the prior

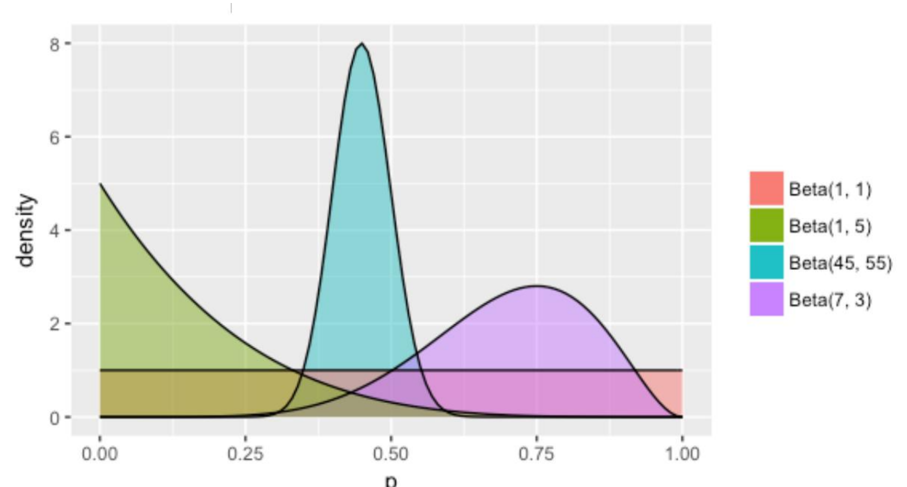
# Three pieces of the Bayesian model

**Prior:**




*Tuning* the Beta shape parameters produces alternative prior models of p

These range from models that reflect more pessimism about your election chances (here the Beta(1,5) in green) to models that reflect a complete lack of certainty about your chances (here the Beta(1,1) in red)



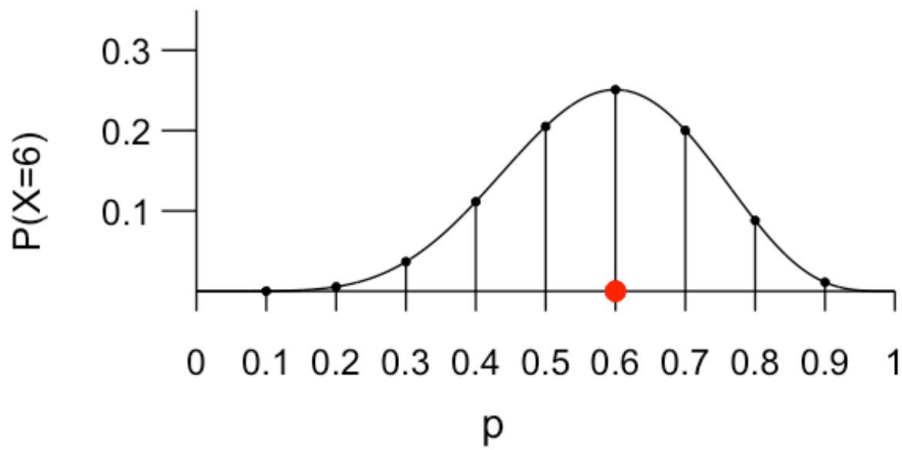

Tuning the prior

# Three pieces of the Bayesian model

**Likelihood:**

p = proportion of voters that support you

*Data –
X=6 out of n=10 voters plan to vote for you.




More likely to have observed this data if $p \approx 0.6$ than if $p < 0.5$
(underlying support – the probability that any given voter supports you)

X = the number of n polled voters that support you, (as a count of successes in n independent trials, each having probability of success p)

Conditional distribution of X given p
$X \sim Bin(n, p)$

Poll assumptions: voters are independent

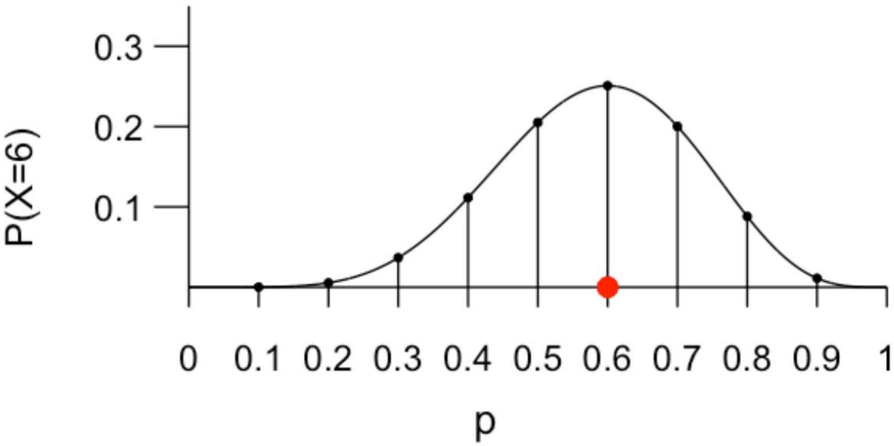
The Binomial model provides the tools needed to quantify the probability of observing *your* poll result under different election scenarios. This result is represented by the red dot: X = 6 of n = 10 (or 60% of) voters support you.

# Three pieces of the Bayesian model

**Likelihood:**

p = proportion of voters that support you

*Data –
X=6 out of n=10 voters plan to vote for you.




More likely to have observed this data if $p \approx 0.6$ than if $p < 0.5$
(underlying support – the probability that any given voter supports you)

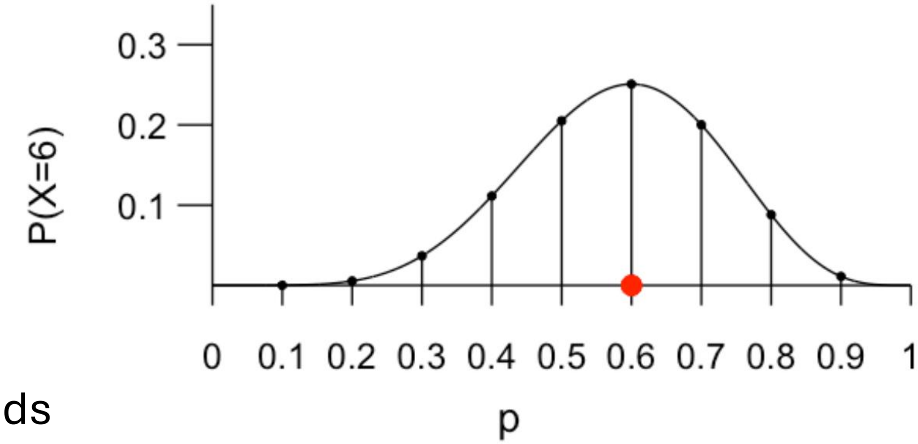
X = the number of n polled voters that support you, (as a count of successes in n independent trials, each having probability of success p)

Conditional distribution of X given p
$X \sim Bin(n, p)$

Poll assumptions: voters are independent

The Binomial model provides the tools needed to quantify the probability of observing *your* poll result under different election scenarios. This result is represented by the red dot: X = 6 of n = 10 (or 60% of) voters support you.

The likelihood function summarizes the likelihood of observing polling data X under different values of the underlying support parameter p.




the likelihood is a function of p that depends upon the observed data X

it provides insight into which parameter values are most compatible with the poll.

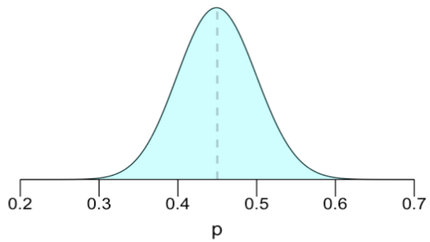
High likelihood = p is compatible with data

Low likelihood = p is not compatible with data

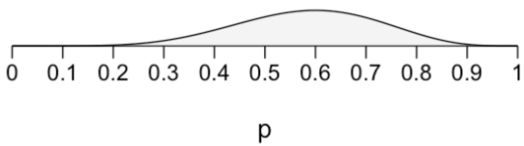
# Three pieces of the Bayesian model

**Posterior:**

P ~ Beta (45,55)

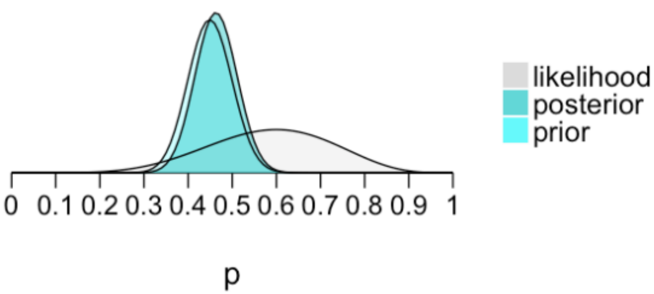



Conditional distribution of X given P
X ~ Bin (n, p)
X ~ Beta (10, p)




the prior contributes knowledge that you built prior to the most recent poll. The likelihood provides insight into the values of p that are most compatible with the current polling data.

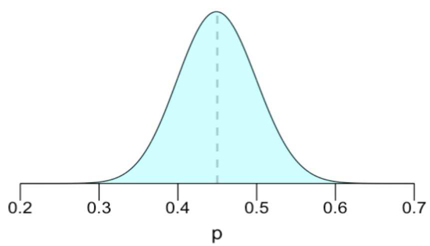
Bayes rule
Posterior $\propto$ Prior x Likelihood



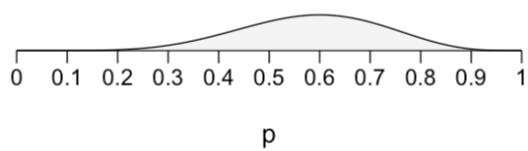
# Three pieces of the Bayesian model

**Posterior:**

P ~ Beta (45,55)




Conditional distribution of X
given P
X ~ Bin (n, p)
X ~ Beta (10, p)




the prior contributes knowledge that you built prior to the most recent poll. The likelihood provides insight into the values of p that are most compatible with the current polling data.

Bayes rule
Posterior $\propto$ Prior x Likelihood