USAID
FROM THE AMERICAN PEOPLE

Bayesian Analysis

USAID MENA Advanced MEL Workshop

Session Objectives

- Understand how to derive Bayes' Rule from the laws of probability
- Understand how to interpret Bayes' Rule in the context of a data analysis
- Understand how thinking like a Bayesian follows the scientific method and should guide how we think about the world

Session Objectives

Bonus content:

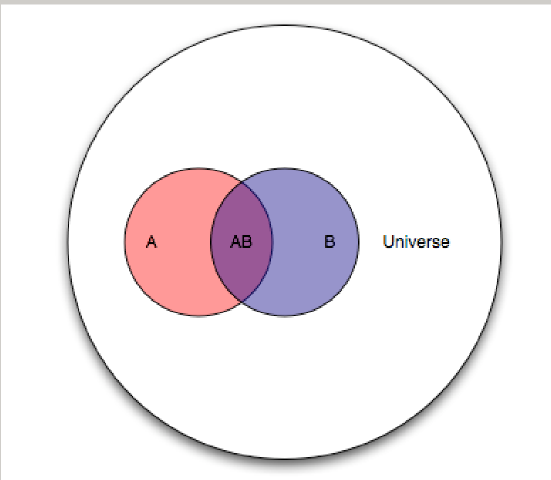
- Setting Bayesian priors as qualitative research
- Naive Bayes'
- Expectation Maximization

Level Set

Deriving Bayes' Rule

Bayes' Rule

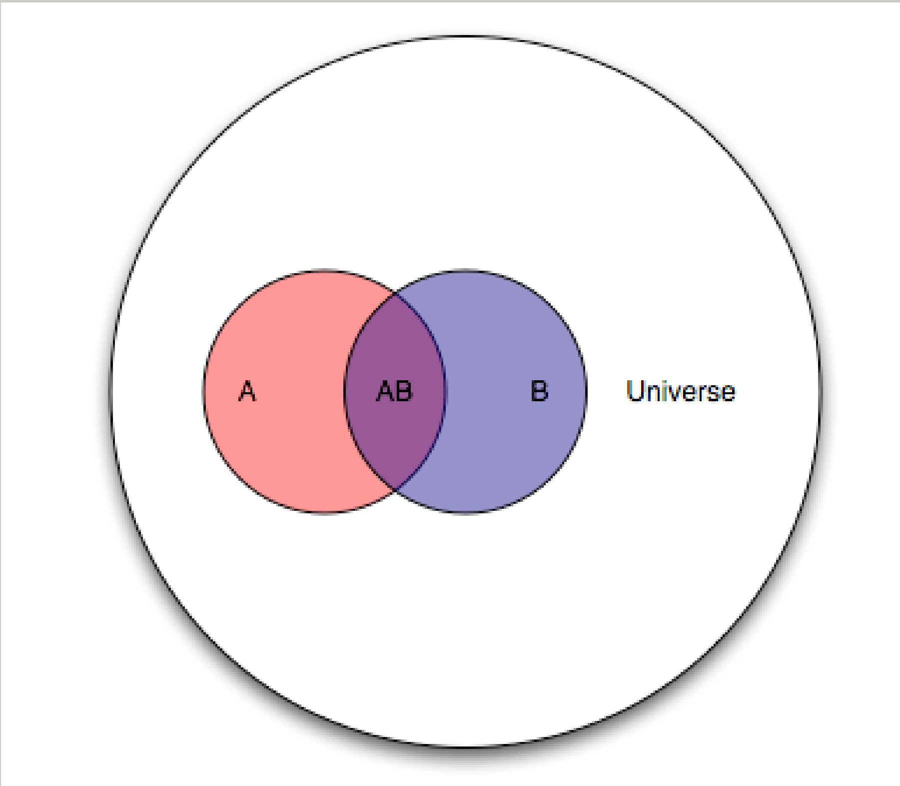
Consider two overlapping events A and B occurring within a universe U.



$$P(A) = \frac{A}{U}$$

$$P(AB) = \frac{AB}{U}$$

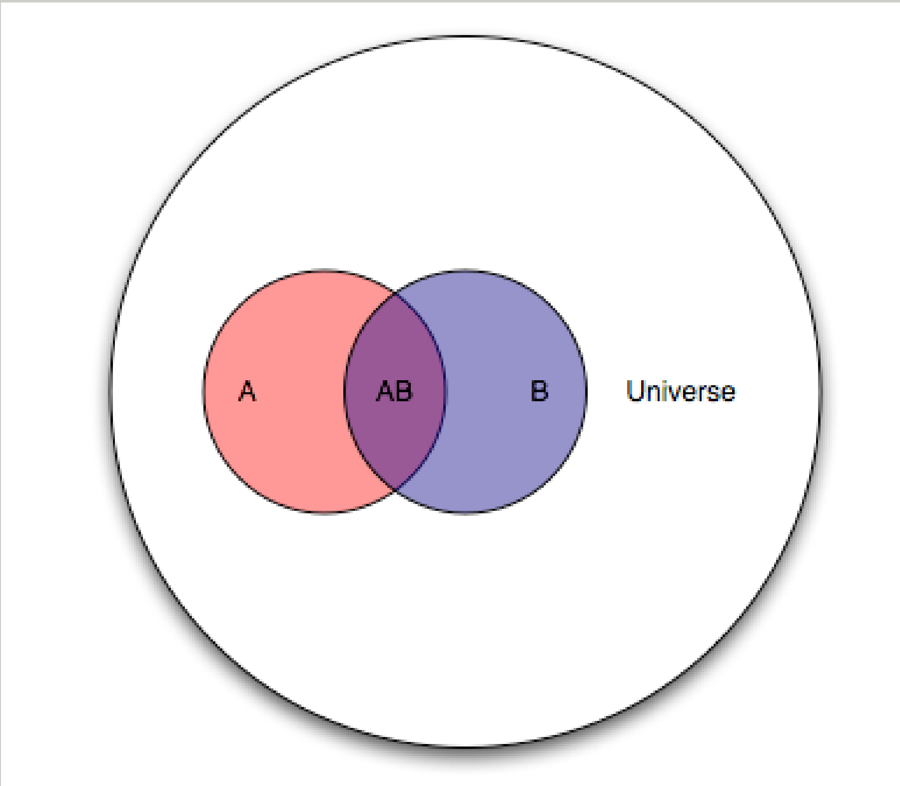
How Much of A is in B?



$$P(A|B) = \frac{P(AB)}{P(B)}$$

$$P(AB) = P(A \mid B)P(B)$$

How Much of B is in A?



$$P(B \mid A) = \frac{P(AB)}{P(A)}$$

$$P(AB) = P(B \mid A)P(A)$$

Putting the Two Together

We can put the two identities together and solve for $A \mid B$:

$$P(A \mid B)P(B) = P(B \mid A)P(A)$$

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$$

And that's it. That's Bayes' Rule.

Bayes' Rule as an Analytical Tool

We just used the laws of probability to derive Bayes' Rule:

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$$

Now let's use this in the context of a data analysis

$$P(model \mid data) = \frac{P(data|model)P(model)}{P(data)}$$

Can also think in terms of hypothesis and evidence

$$P(hypothesis \mid evidence) = \frac{P(evidence|hypothesis)P(hypothesis)}{P(evidence)}$$

Using Bayes' Rule

$$P(model \mid data) = \frac{P(data \mid model)P(model)}{P(data)}$$

$P(model \mid data)$: probability of a hypothesis given data

$P(data \mid model)$: the likelihood of our data for each hypothesis

$P(model)$: the prior probability of the model, before data

$P\left(data\right)$: a normalizing constant

Bayesian Inference

- 40 subjects, half randomly assigned a treatment expected to reduce the probability of an event
- By random assignment, we would expect the event to fall evenly across treatment and control groups[1]
- What is the probability p that an observed event occurred within the treatment group?

Setting Up Our Hypotheses

$H_0 : p = 50\%$ No treatment effect

$H_1 : p < 50\%$ Treatment effect

- 20 events - 4 events in the treatment group and 16 events in the control group
- How likely are these four events to have occurred within the treatment group?

Frequentist Hypothesis Test

$$P\left(data \geq data_{observed} \mid H_0\right)$$

estimate	statistic	p.value	parameter	conf.low	conf.high	method	alternative
0.2	4	0.0118	20	0.0573	0.437	Exact binomial test	two.sided

$$\frac{4}{20} = 20\%$$

$$p = .012$$

Setting Up the Bayesian Engine

1. Set a range of plausible values (the model space)
2. Calculate the likelihood of the data for each plausible value
3. Set the prior probability of each plausible value
4. Multiply the likelihood by the prior (numerator)
5. Divide by the denominator to get the posterior probability

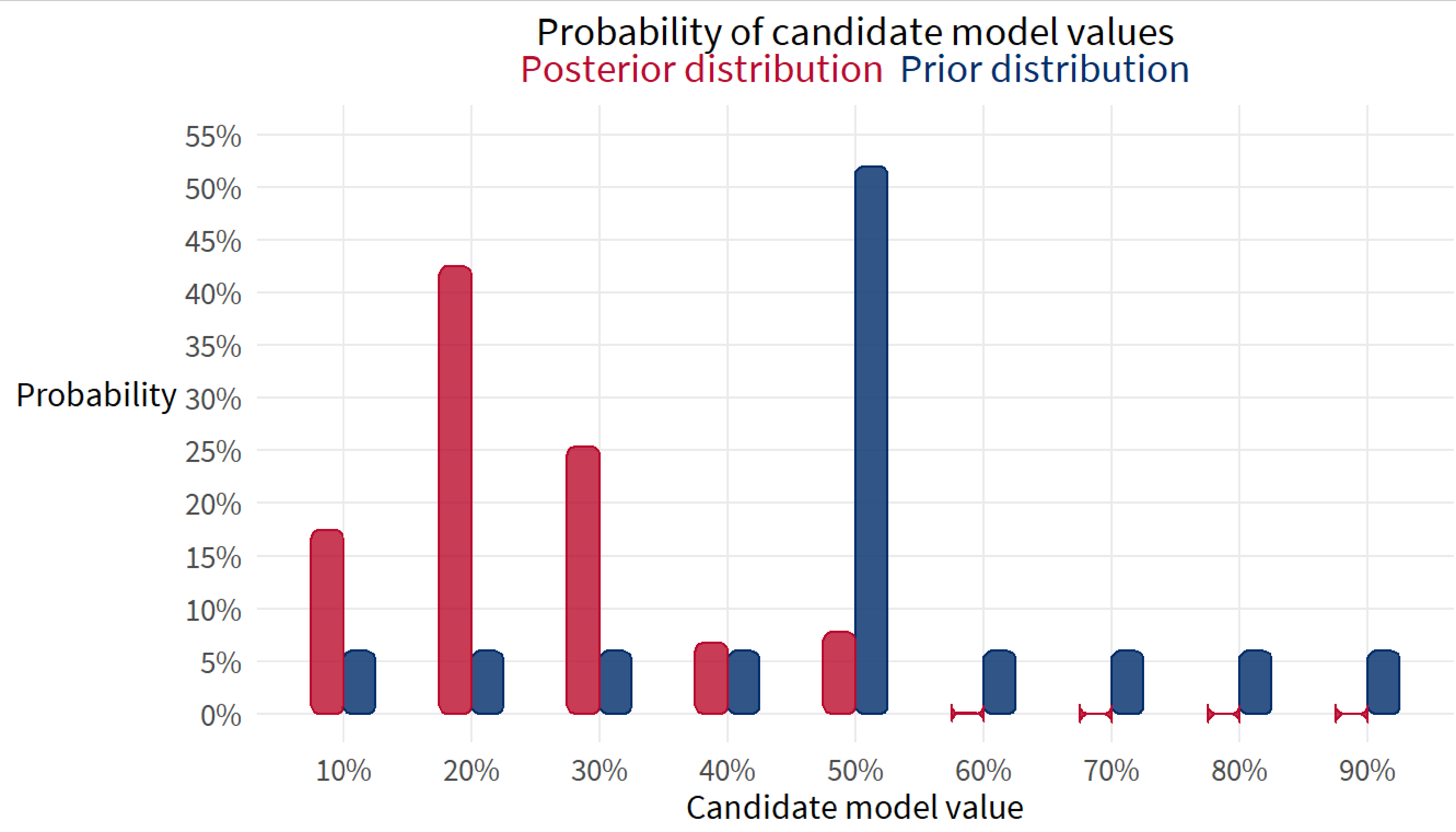
Results of the Bayesian Engine

hypothesis	likelihood	prior	numerator	posterior
10%	0.090	6%	0.005	17.5%
20%	**0.218**	**6%**	**0.013**	**42.5%**
30%	0.130	6%	0.008	25.4%
40%	0.035	6%	0.002	6.8%
50%	0.005	52%	0.002	7.8%
60%	0.000	6%	0.000	0.1%
70%	0.000	6%	0.000	0.0%
80%	0.000	6%	0.000	0.0%
90%	0.000	6%	0.000	0.0%

A treatment effect of 20 percent is most likely

But notice that we get back an entire distribution, not just a point estimate

From Prior to Posterior



Bayesian Analysis as Science

- Recall what we learned as kids about the scientific method:
  - Observe a state of the world
  - Develop a hypothesis about how the world works
  - Test your hypothesis with new data
  - Update your beliefs and repeat

Think Like a Bayesian

- Using Bayes' Rule to conduct inference follows the scientific method!
- Let's think like a Bayesian
- Stay tuned for Bayes' Rule used in machine learning

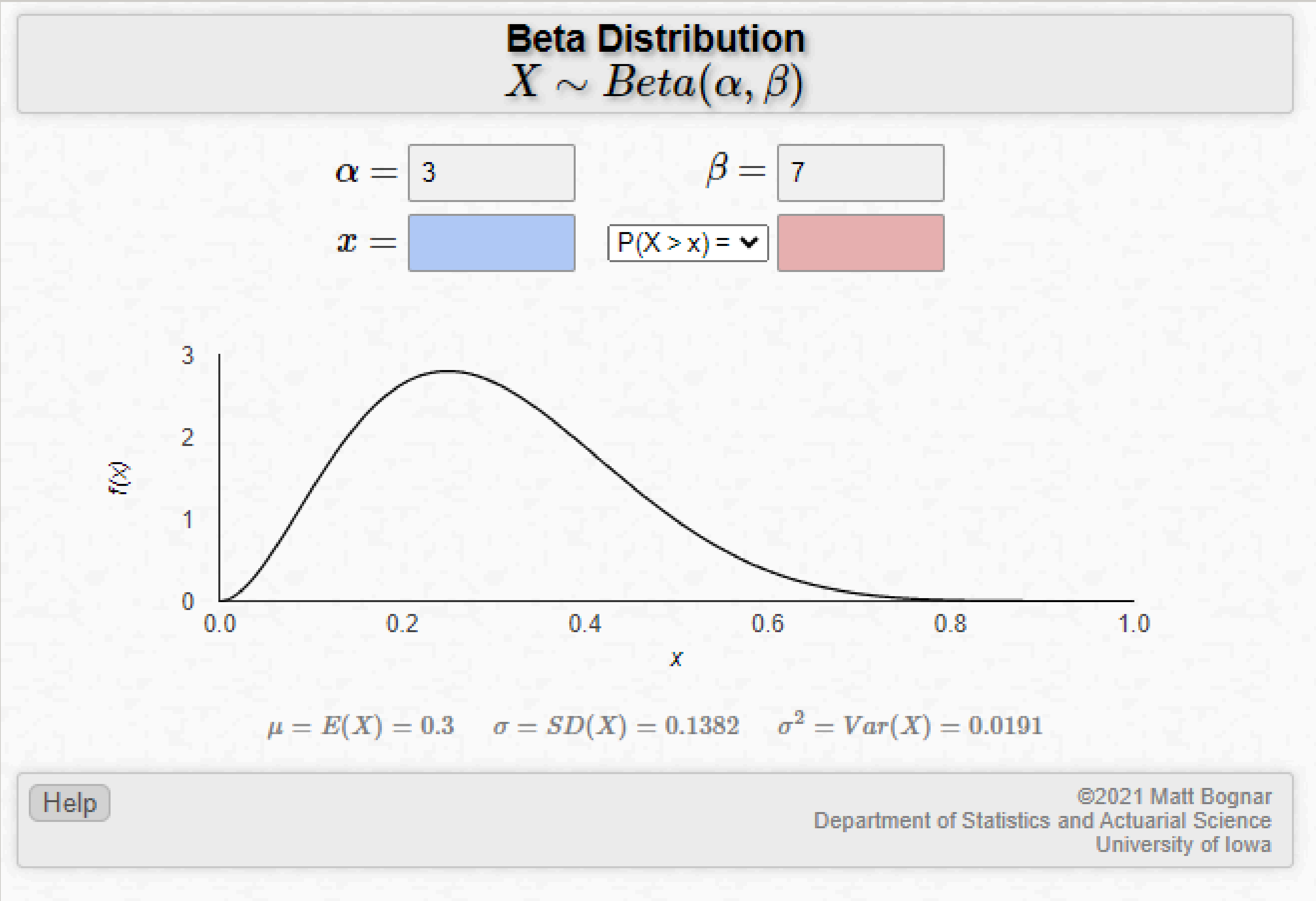
Bonus Content

- Setting Bayesian priors as qualitative research
- Naive Bayes Classifier
- Expectation Maximization (EM) algorithm

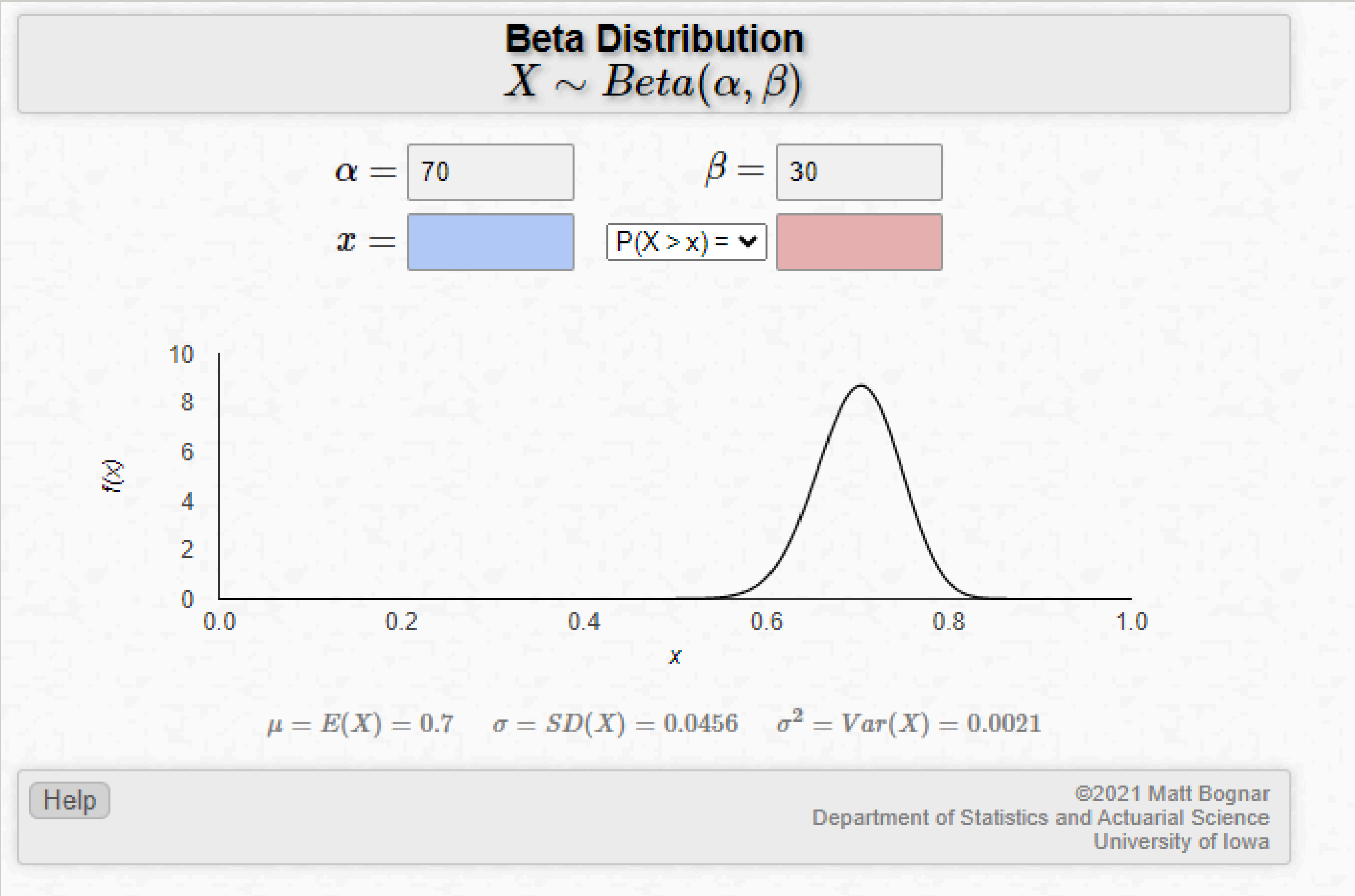
Bayesian Priors

- The prior probability $P(model)$ reflects our current state of understanding about our hypothesis
- Stakeholders have a prior probability of the hypothesis, even if they don’t think of it in terms of a Bayesian analysis
- What if we used the elicitation of prior probability as a qualitative research method?

Uncertain About Hypothesis



More Confident About Hypothesis



Bayes Priors as Qualitative Inquiry

1 What percent of respondents do you expect to report being aware of / familiar with USAID (respond for both West Bank and Gaza)?

	Percentage (%)	Margin of Error (%)
West Bank	Enter answer	Enter answer
Gaza	Enter answer	Enter answer

2 Can you share a bit more about why you entered those estimates?

Enter answer

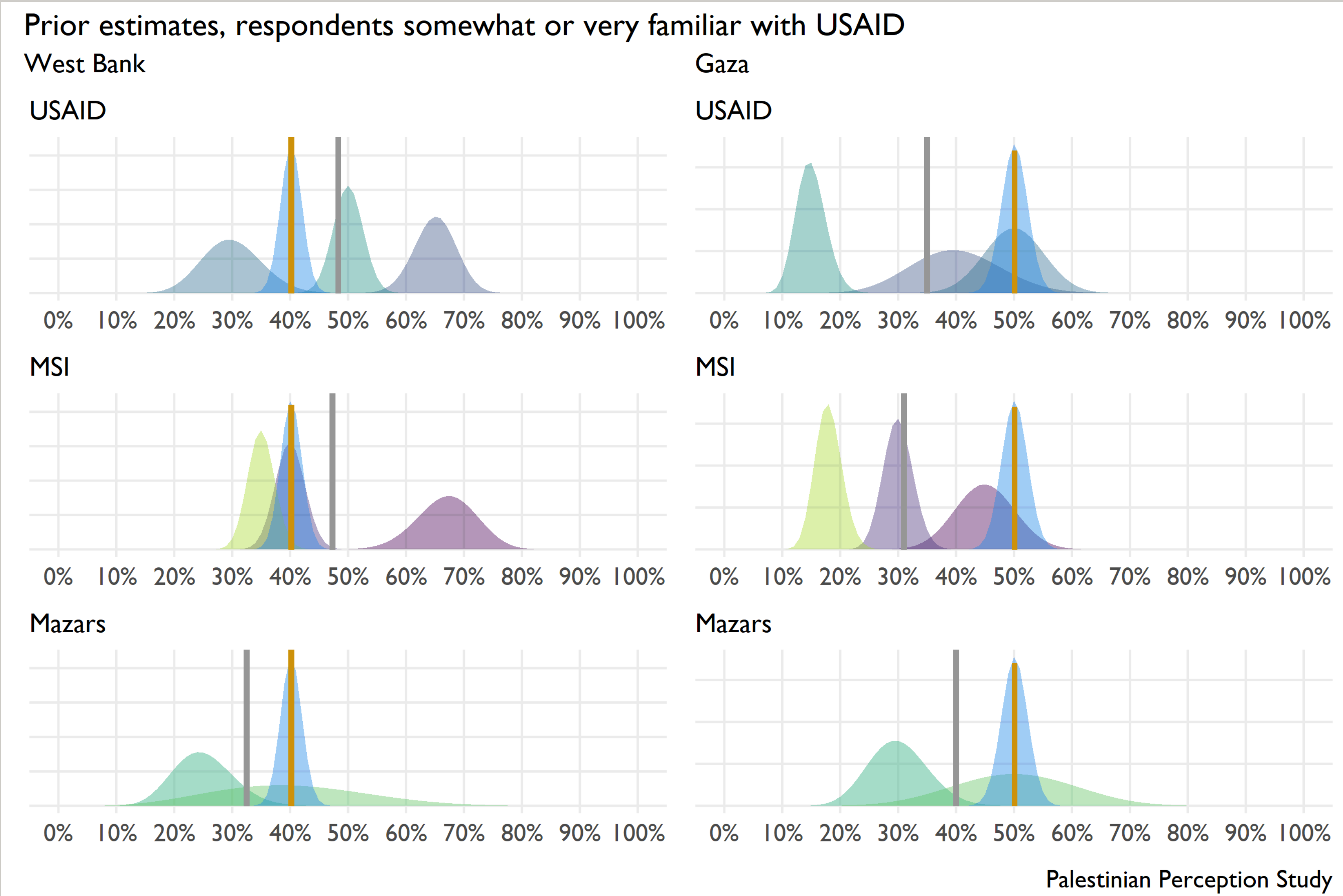
3 What percent of respondents do you expect to report a somewhat or very positive perception of USAID (respond for both West Bank and Gaza)?

	Percentage (%)	Margin of Error (%)
West Bank	Enter answer	Enter answer
Gaza	Enter answer	Enter answer

4 Can you share a bit more about why you entered those estimates?

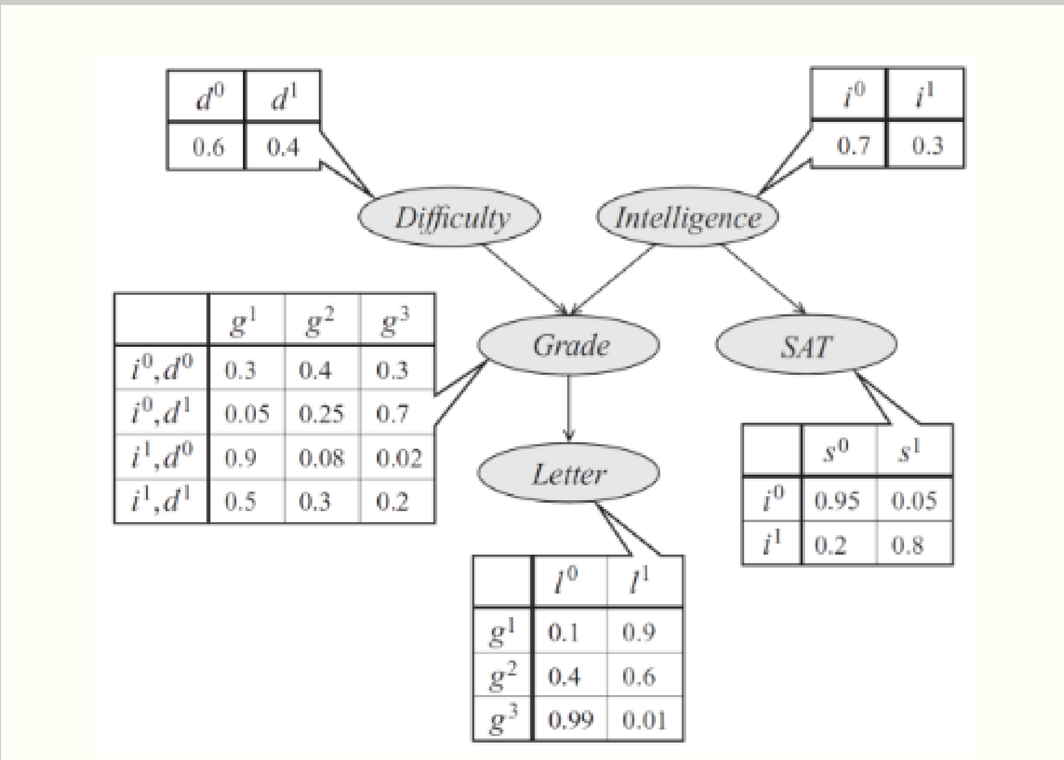
Enter answer

Prior Probability Elicitation



Naive Bayes

- Consider a logic model where we can assign actual probabilities



Putting the Naive in Naive Bayes

- Assume each variable is independent of the others, even if we know that is not true ("naive")
- Given the assumption of independence, all we have to do is multiply probabilities together to estimate an outcome

$$posterior \propto prior \times likelihood$$

- Automagically, naive Bayes gives us helpful answers!

Expectation Maximization (EM)

- Expectation maximization is an algorithm that iteratively updates probabilities using Bayes' Rule
- EM is used to impute missing data, or detect latent variables
- Starting with a reasonable best-guess of parameter values, the model learns from the data and updates the probabilities
- Automagically, the model converges to the best parameter values

E to the M to the E

Expectation step

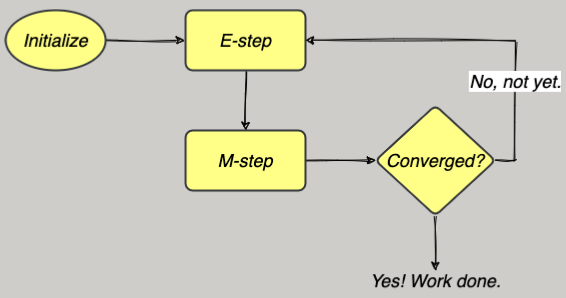
- For our best-guess of parameter values and observed data, what is the posterior distribution?

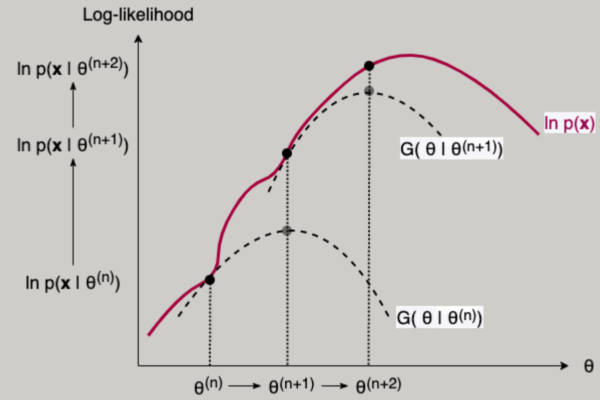
Maximization step

- Use the most likely values from the expectation step to make our next best-guess of parameter values

Expectation step.. Maximization step..

EM Climbs the Hill of Likelihood





https://yangxiaozhou.github.io/data/2020/10/20/EM-algorithm-explained.html

Recap of bonus content

- Bayesian analysis starts simple with Bayes' Rule
- Because Bayes' Rule is based on the laws of probability, we can build very complex algorithms on top of it
- It is important to *practically* understand Bayes' Rule, and use it in our every day thinking about the world
- It is important to *conceptually* understand how Bayes' Rule can be applied in more complex ways to help us learn

Looking forward

Stay tuned for sessions on

- Machine Learning
- Large Language Models
- Artificial General Intelligence

Thank you!