Bayesian Statistics II

Matt Brems

GENERAL ASSEMBLY

Learning Objectives

- Describe questions that can be answered by Bayesian inference.
- Compare and contrast Bayesian and Frequentist inference.
- Describe the role of the prior and likelihood on the posterior distribution.
- Describe the Uniform, Normal, Binomial, Poisson, Gamma, and Beta distributions.

Based on the work you did for the Ames Housing project, what are some questions you could answer?

Bayesian inference allows us to answer questions that cannot *(easily)* be answered by Frequentist inference!

$$[\text{sale\_price}] = \beta_0 + \beta_1 [\text{lot\_area}] + \beta_2 [\text{overall\_qual}]$$

What is the probability that, as **`lot_area`** increases by 1, **`sale_price`** increases by at least $100?

What are the chances that a house with **`overall_qual`** of 5 costs more than $80,000?

What is a likely range of values for **`sale_price`** among 1,000 square foot lots?

When speaking about statistics, what is a <u>parameter</u>?

What are examples of statistical parameters?

Bayesian inference allows us to make probability statements about parameters, which is really helpful!

$$[\text{SBP}] = \beta_0 + \beta_1 [\text{drug A}]$$

What is the probability that Drug A causes a higher drop in blood pressure than Drug B?

What is the probability that Drug A causes a drop at least 15 points higher than Drug B?

What is the likeliest average difference in blood pressure between drug A and drug B?

We should not *always* follow Bayesian methods or *always* follow Frequentist methods. There are times in which one may be better than the other.

Bayesian Inference

- More intuitive answers.
- Works with low sample sizes!
- Less common.
- Harder to initially set up.
- We include our personal knowledge: interpreted as more subjective.

Frequentist Inference

- Less intuitive answers.
- Requires larger sample sizes.
- More common.
- Simpler to get set up.
- Data speak for themselves: interpreted as more objective.

How do we do Bayesian inference?

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

How do we do Bayesian inference?

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

$$[\text{posterior}] = [\text{data}] \times [\text{prior}]$$

How do we do Bayesian inference?

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

$$[\text{posterior}] = [\text{data}] \times [\text{prior}]$$

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

What does this mean?

Our goal: Generate this!

We pick these!

$$[\text{posterior}] = [\text{data}] \times [\text{prior}]$$

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

What does this mean?

$$[\text{posterior}] = [\text{data}] \times [\text{prior}]$$

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

This is our **prior distribution**: this summarizes what we think about our parameter **before we gather any data**.

What does this mean?

$$[\text{posterior}] = [\text{data}] \times [\text{prior}]$$

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

This is our **likelihood**: this summarizes how likely it is to observe the data we observed given various values of our parameter.

What does this mean?

Our goal: Generate this!

$$[\text{posterior}] = [\text{data}] \times [\text{prior}]$$

$$f(\mu|\text{data}) \propto f(\text{data}|\mu) f(\mu)$$

This is our **posterior distribution**: this summarizes what we think about our parameter **after taking our data into account**. It is a combination of our prior beliefs and our data.

What does this mean?

$$[\text{posterior}] = [\text{data}] \times [\text{prior}]$$

https://rpsychologist.com/d3/bayes/

Questions to Answer

1. Play around with the observed effect. How would you describe the posterior distribution?

2. All else held equal, what happens when you increase the sample size? Decrease the sample size?

3. All else held equal, what happens when you change the uncertainty in your prior beliefs?

Steps to Bayesian Inference

1. Select a prior distribution.

2. Select a likelihood.

3. Generate the posterior distribution.

4. Do whatever inference we want!

Probability Distributions

Bayesian Inference

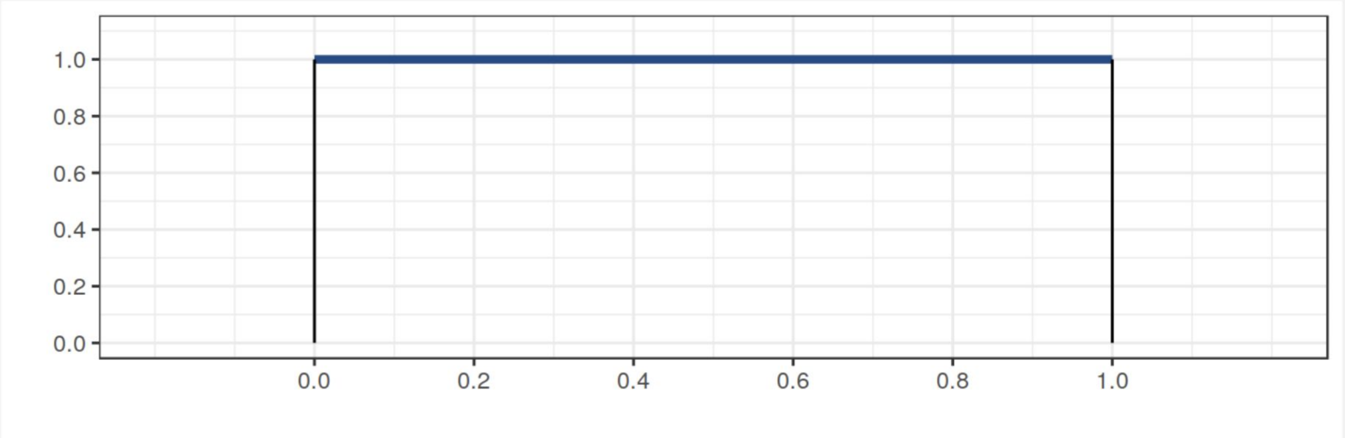
In order to do Bayesian inference, it's helpful to have probability distributions at our fingertips.

1. Uniform
2. Normal
3. Binomial
4. Poisson
5. Gamma
6. Beta

Uniform Distribution

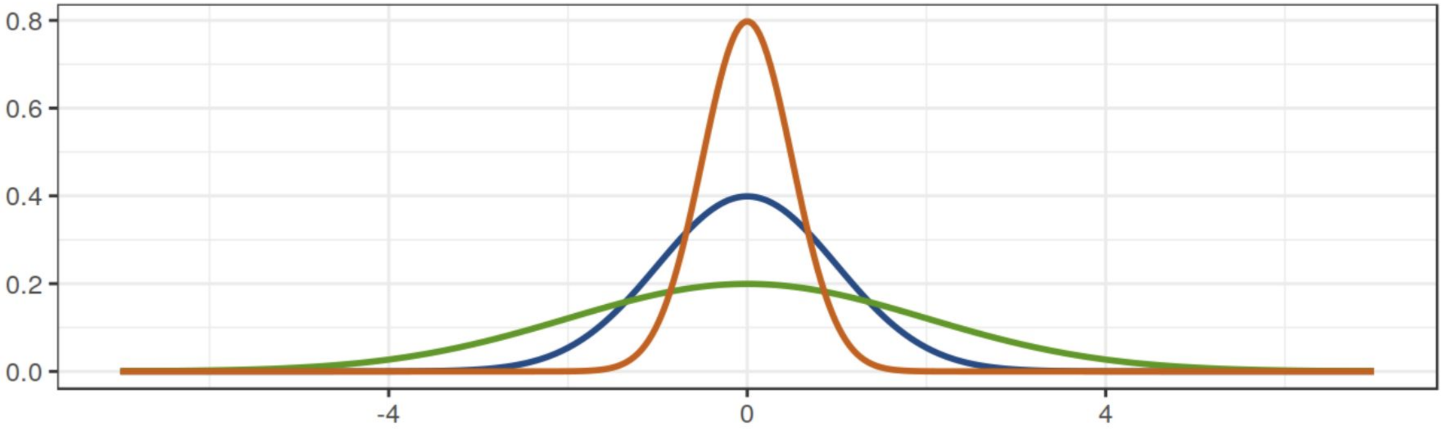
A Uniform Distribution is used to assign **equal probability** to all outcomes.

Can be continuous or discrete.



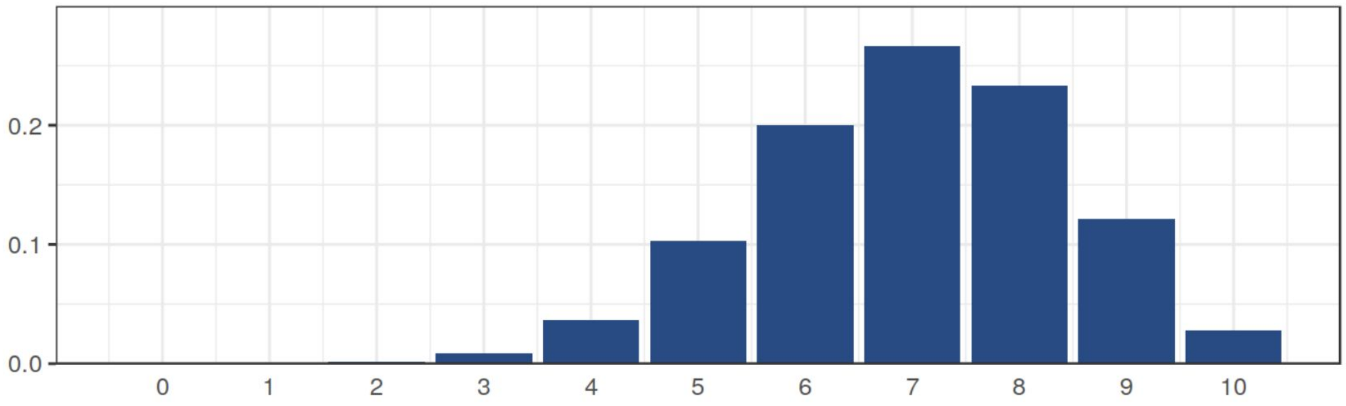
Normal Distribution

A Normal Distribution is symmetric and bell-shaped. It is defined by its mean and standard deviation. It ranges from -∞ to +∞.



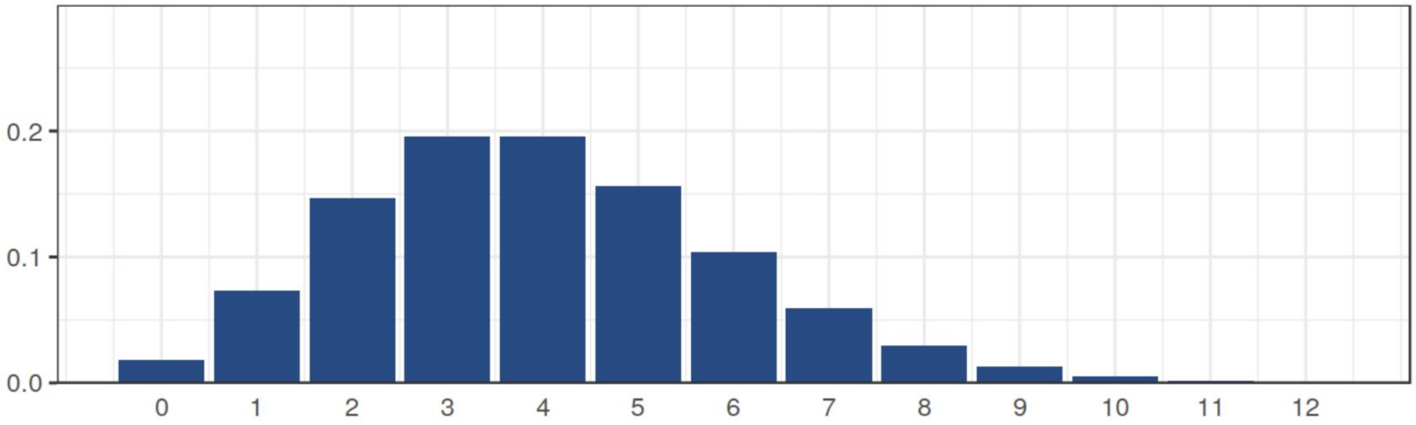
Binomial Distribution

A Binomial Distribution is used to count the number of successes out of a pre-defined number of trials, including all integers from 0 to *n*. It is defined by the number of trials, *n*, and the probability of success, *p*.



Poisson Distribution

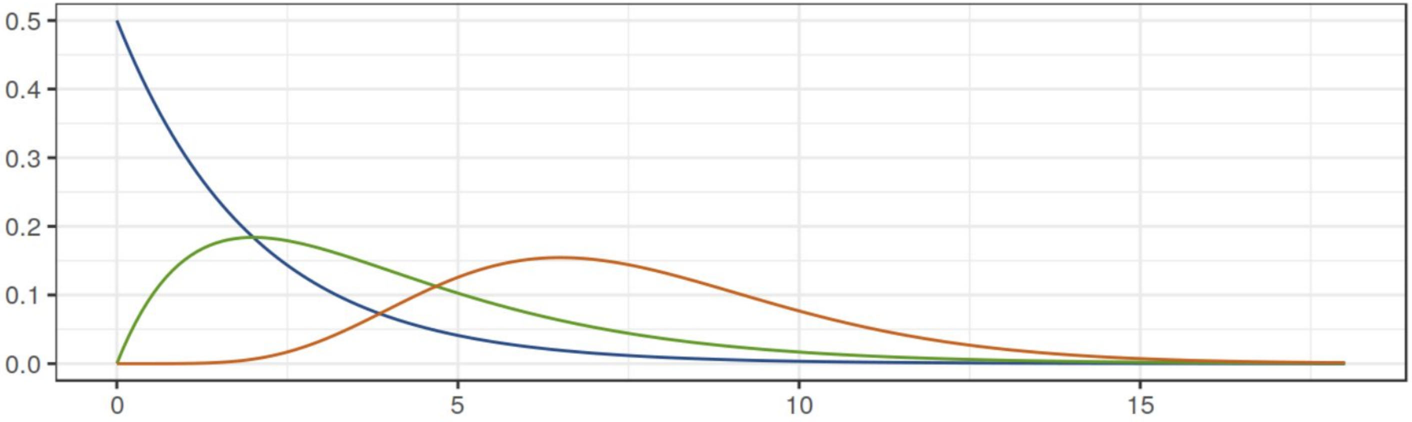
A Poisson Distribution is used to count the number of successes in a fixed interval of time, including all integers from 0 to +∞. It is defined by the average number of successes, λ, in the window of time.



Gamma Distribution

A Gamma Distribution is used to measure the amount of time until a number of successes happen and ranges from 0 to +∞.

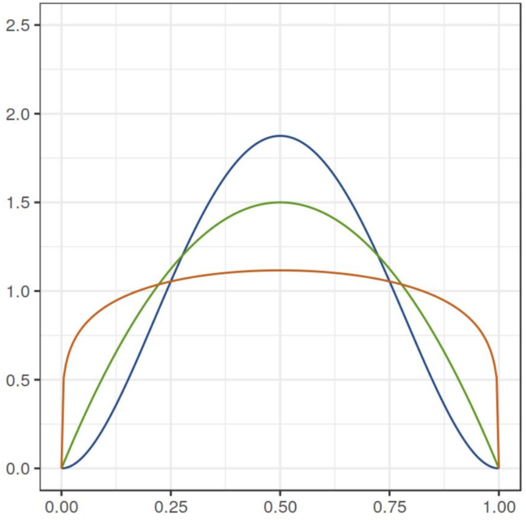
The parameters of a Gamma distribution correspond to the shape and scale of the distribution.



Beta Distribution

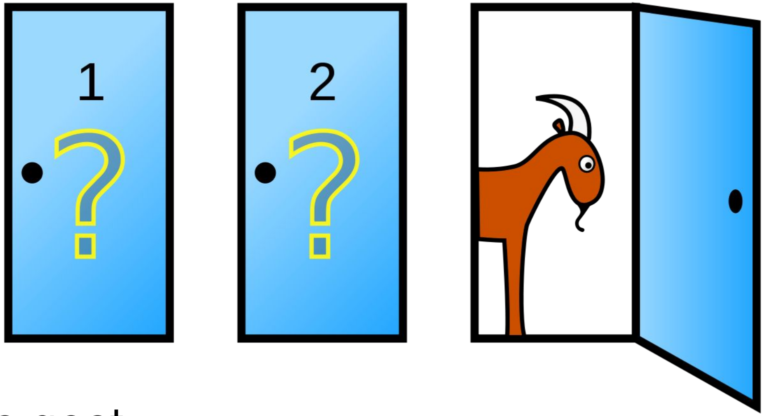
A Beta Distribution ranges from 0 to 1 and is a very common choice to describe probabilities.

The parameters of a Beta distribution correspond to the shape and scale of the distribution.



Examples

Monty Hall Revisited



There are three doors, with a car behind one.

We pick a door we think the car is behind.

Monty Hall, our host, opens one door to reveal a goat.

We decide whether we want to **keep our original choice** or **switch to the other door**.

https://commons.wikimedia.org/wiki/File:Monty_open_door.svg

Monty Hall Revisited: What prior did we use here?

$$P(\text{car in A}|\text{goat in B}) = \frac{P(\text{goat in B}|\text{car in A})P(\text{car in A})}{P(\text{goat in B})}$$

$$P(\text{car in B}|\text{goat in B}) = \frac{P(\text{goat in B}|\text{car in B})P(\text{car in B})}{P(\text{goat in B})}$$

$$P(\text{car in C}|\text{goat in B}) = \frac{P(\text{goat in B}|\text{car in C})P(\text{car in C})}{P(\text{goat in B})}$$

Scenario: Placenta Previa

We are doing a study to estimate the probability of a female birth given that the mother has a condition called placenta previa.

What is the parameter we're trying to study?

Example adapted from Bayesian Data Analysis, 3rd Edition.

Scenario: Placenta Previa

We are doing a study to estimate the probability of a female birth given that the mother has a condition called placenta previa.

What is the parameter we're trying to study?

What distribution might be good for this parameter? Why?

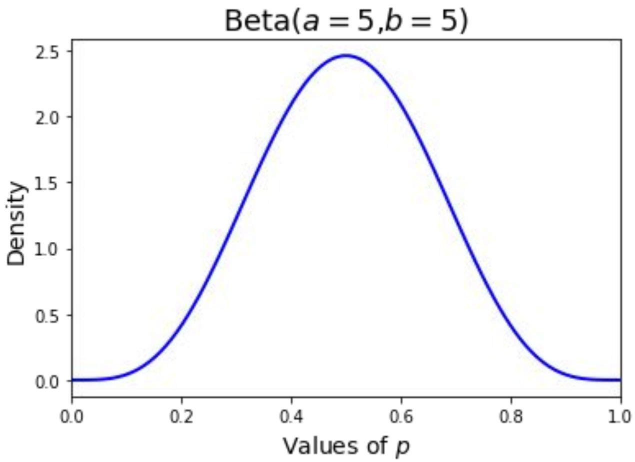
Example adapted from Bayesian Data Analysis, 3rd Edition.

Scenario: Placenta Previa

We are doing a study to estimate the probability of a female birth given that the mother has a condition called placenta previa.

Let's use a Beta(5, 5) distribution for our prior.

A study among mothers with placenta previa was conducted in Germany, with 437 female births out of 980 total births. What would be a good distribution to use for our likelihood?



Example adapted from Bayesian Data Analysis, 3rd Edition.

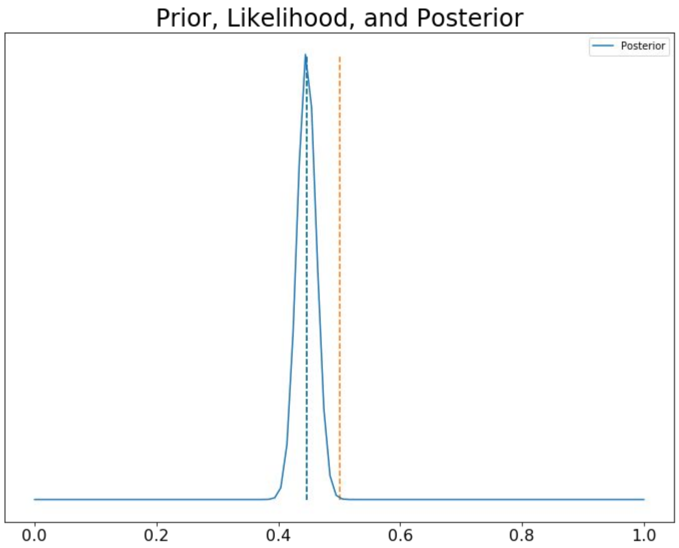
Scenario: Placenta Previa

We are doing a study to estimate the probability of a female birth given that the mother has a condition called placenta previa.

Let’s use a Beta(5, 5) distribution for our prior.

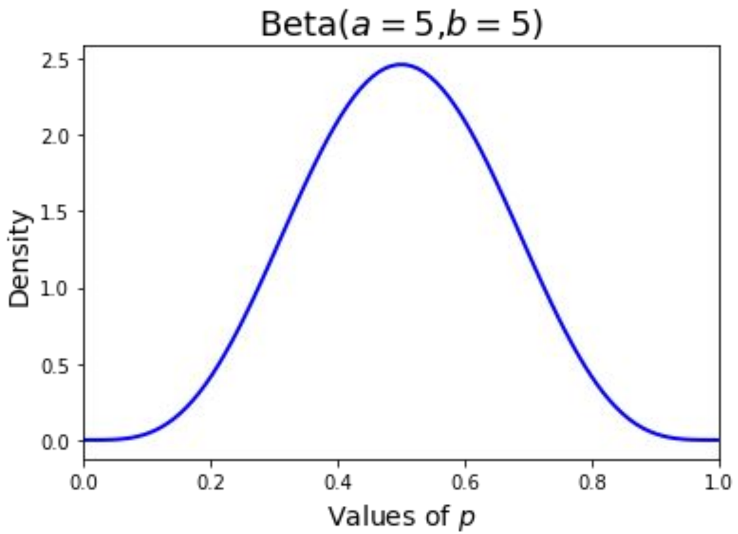
Let’s use a Binomial likelihood.

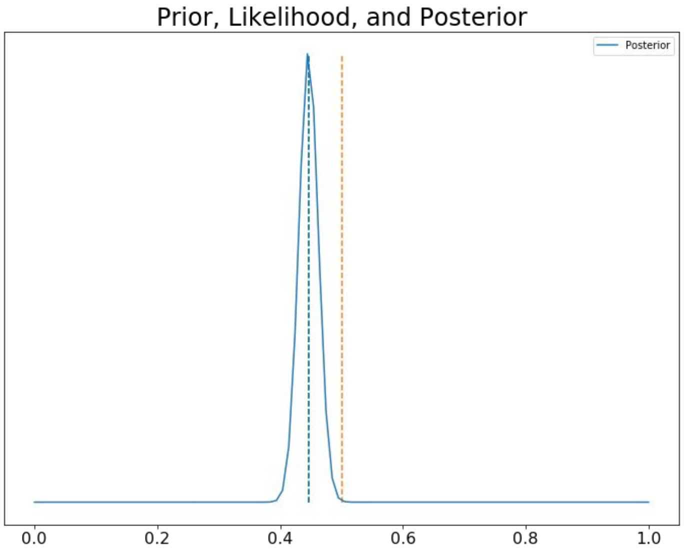
We can show that the posterior distribution is then a Beta(442, 548).



Example adapted from Bayesian Data Analysis, 3rd Edition.

Scenario: Placenta Previa





Example adapted from Bayesian Data Analysis, 3rd Edition.

Appendix

Bayes' Rule

How does Bayes’ Theorem connect to this?

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

How does Bayes' Theorem connect to this? (Step 1)

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

$$P(\mu = 0|\text{data}) = \frac{P(\text{data}|\mu{=}0)P(\mu{=}0)}{P(\text{data})}$$

How does Bayes' Theorem connect to this? (Step 2)

$$P(\mu = 0|\text{data}) = \frac{P(\text{data}|\mu{=}0)P(\mu{=}0)}{P(\text{data})}$$

$$f(\mu|\text{data}) = \frac{f(\text{data}|\mu)f(\mu)}{f(\text{data})}$$

How does Bayes' Theorem connect to this? (Step 3)

$$f(\mu|\text{data}) = \frac{f(\text{data}|\mu)f(\mu)}{f(\text{data})}$$

$$f(\mu|\text{data}) \propto f(\text{data}|\mu)f(\mu)$$

How does Bayes' Theorem connect to this? (Cheat Sheet)

1. Replace event **A** with parameter μ and event **B** with our data.

2. Instead of manually calculating this for all values of μ, let's use probability distributions to summarize what μ might look like.

3. Disregard the denominator.

