Intro to Bayesian Analysis

1st IAA-CSIC Severo Ochoa School on Statistics, Data Mining, and Machine Learning
Nov 4, 2019

Gwendolyn Eadie
Assistant Professor
Department of Astronomy & Astrophysics, Department of Statistical Sciences
University of Toronto

This exercise is from Eadie et al (2019), Journal of Statistics Education

Introducing Bayesian Analysis

with a tasty example!

What is the probability of drawing a blue m&m's® from an individual bag of m&m's®?

What is the percentage of blue m&m's® made at the factory?

What is the probability of drawing a blue m&m's® from an individual bag of m&m's®?

What is the percentage of blue m&m's® made at the factory?

No online searches, that spoils the fun!

What do you think is the simplest way to estimate the number of blue m&m's made at the factory?

What do you think is the simplest way to estimate the number of blue m&m's made at the factory?

What are the pitfalls to this approach?

What if I gave you only one bag of m&m's?

What if I only let you use the first 20 m&m's from that bag as your data?

Quick Review: Bayes' Theorem

Let's go to the blackboard.

Bayes' Theorem

$$p\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

Posterior distribution

$\boldsymbol{\theta} \longrightarrow$ model parameters

$\boldsymbol{y} \longrightarrow$ data

Bayes' Theorem

$$p\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

Likelihood

Posterior distribution

Bayes' Theorem

$$p\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

Likelihood

Posterior distribution

Prior distribution

m&m's® Activity

$$p\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

m&m's® Activity

$$p\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

Find the posterior distribution for the percentage of blue m&m's® made at the factory ***— using Bayes' theorem and one bag of m&m's.***

m&m's® Activity

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

$\boldsymbol{\theta} \rightarrow$ percentage of **blue** **m&m's**® made at the factory

$\boldsymbol{y} \rightarrow$ data (**m&m's**®)

Initial Questions (3 minutes)

1. What kind of data are m&m's?
    a. Numerical
    b. Categorical
    c. Continuous

2. How will you record the data?
3. Will you sample with replacement or without?

Initial Questions

1. What kind of data are m&m's?
   a. Numerical
   b. Categorical
   c. Continuous

2. How will you record the data?
3. Will you sample with replacement or without?

Initial Questions

1. What kind of data are m&m’s?
   a. Numerical
   b. Categorical
   c. Continuous

2. How will you record the data? Table (all colours)
3. Will you sample with replacement or without?

Initial Questions

1. What kind of data are m&m's?
   a. Numerical
   b. Categorical
   c. Continuous

2. How will you record the data? Table (all colours)
3. Will you sample with replacement or without?

What's the sampling distribution? (3 min)

i.e. *How might we model the probability of drawing a blue m&m?*

Likelihood of drawing ***y*** blue **m&m's**® given **n** trials:

y $\longrightarrow$ # of successes (blue **m&m**®)

n-y $\longrightarrow$ # of failures (not a blue **m&m**®)

$$p\left(\boldsymbol{\theta}\middle|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}\middle|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

Likelihood of drawing ***y*** blue **m&m's®** given **n** trials:

$$p\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto p\left(\boldsymbol{y}|\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)$$

y $\longrightarrow$ # of successes (blue **m&m®**)

n-y $\longrightarrow$ # of failures (not a blue **m&m®**)

Binomial Distribution:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{y} (1-\boldsymbol{\theta})^{n-y}$$

Likelihood of drawing ***y*** blue **m&m's®** given **n** trials:

y → # of successes (blue **m&m®**)

n-y → # of failures (not a blue **m&m®**)

Binomial Distribution:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{y}\,(1-\boldsymbol{\theta})^{n-y}$$

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

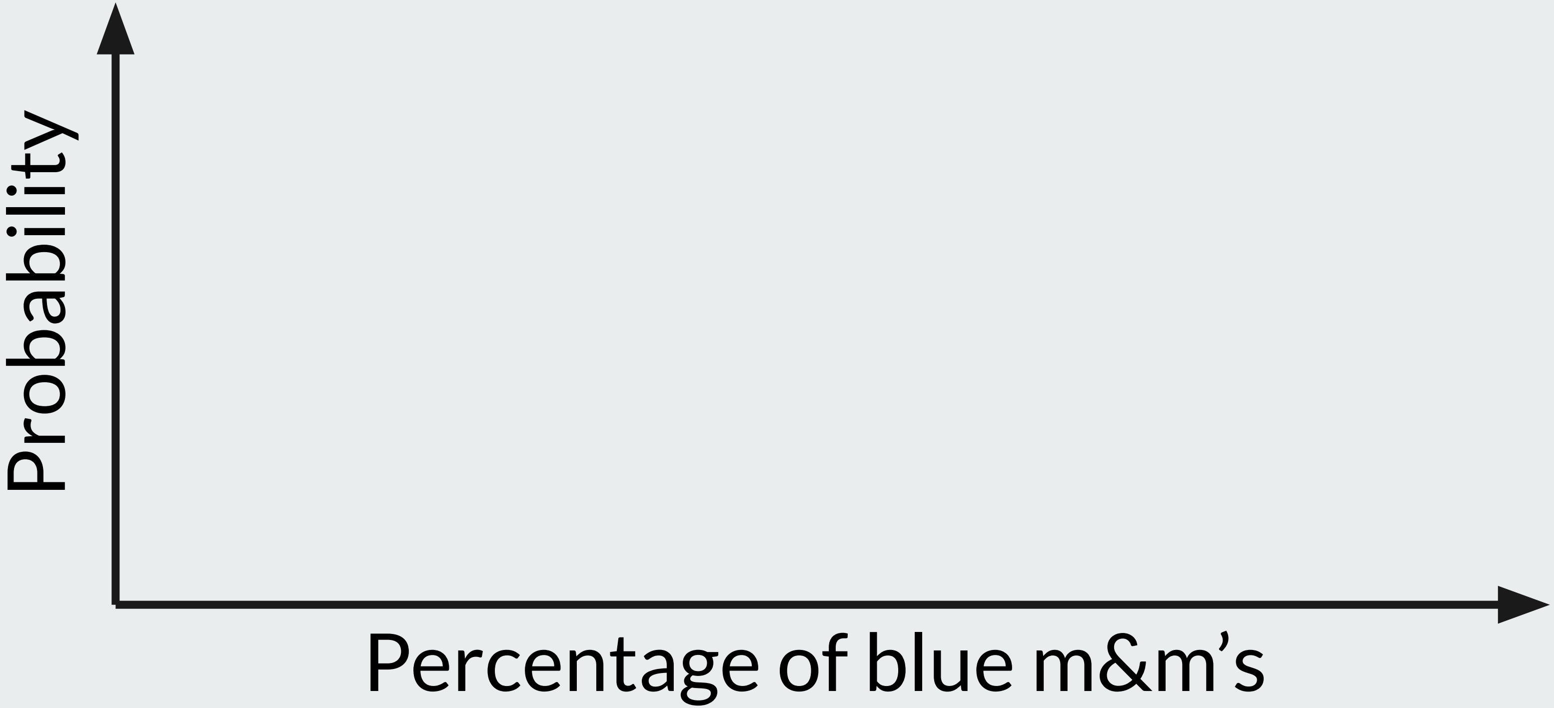
Prior Distribution:

use prior knowledge about percentage of blue **m&m's®**

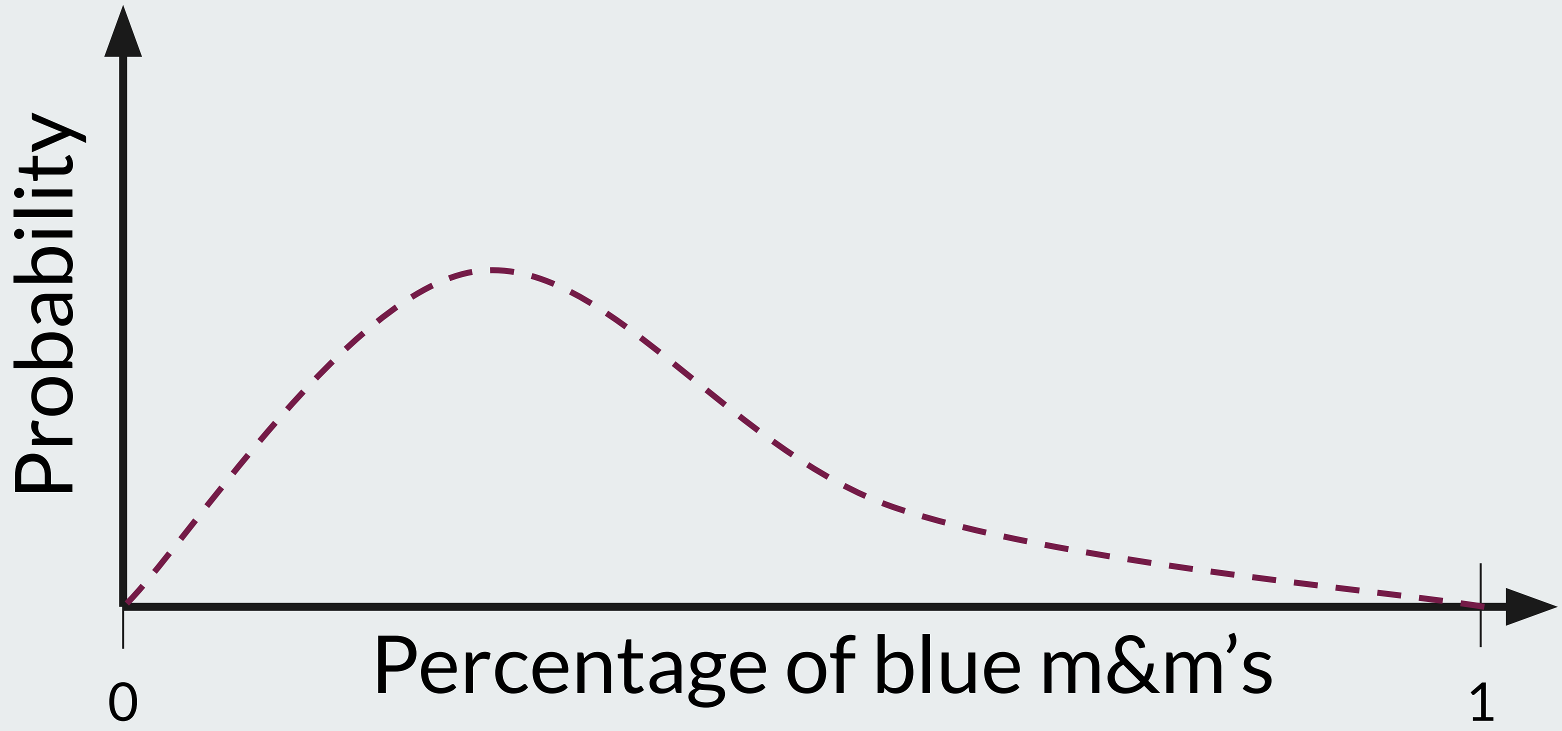
What’s our prior information?

- How many different colours of m&m’s are there?
- Do you think the m&m’s are well-mixed before they go into a bag at the factory?
- What percentage of blue m&m’s do you think are made at the factory?
- Do you think every bag will have the same percentage of blue m&m’s?

Sketch prior knowledge



Sketch prior knowledge



Likelihood of drawing ***y*** blue **m&m's®** given **n** trials:

y → # of successes (blue **m&m®**)

n-y → # of failures (not a blue **m&m®**)

Binomial Distribution:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{y}(1-\boldsymbol{\theta})^{n-y}$$

$$\boxed{p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}$$

Prior Distribution:
use prior knowledge about percentage of blue **m&m's®**

→ **Conjugate prior**

Beta Distribution:

$$p(\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$$

Prior Distribution
hyperparameters

$\alpha = ?$

$\beta = ?$

$$\boxed{p(\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}}$$

Prior Distribution
hyperparameters

$\alpha = ?$
$\beta = ?$

$$\boxed{p(\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}}$$

- Find values for the α and β hyperparameters that best match your sketched prior distribution

- Some values to try:
  - when $\alpha = \beta$
  - when $\alpha = \beta = 1$

In RStudio… (15 minutes)

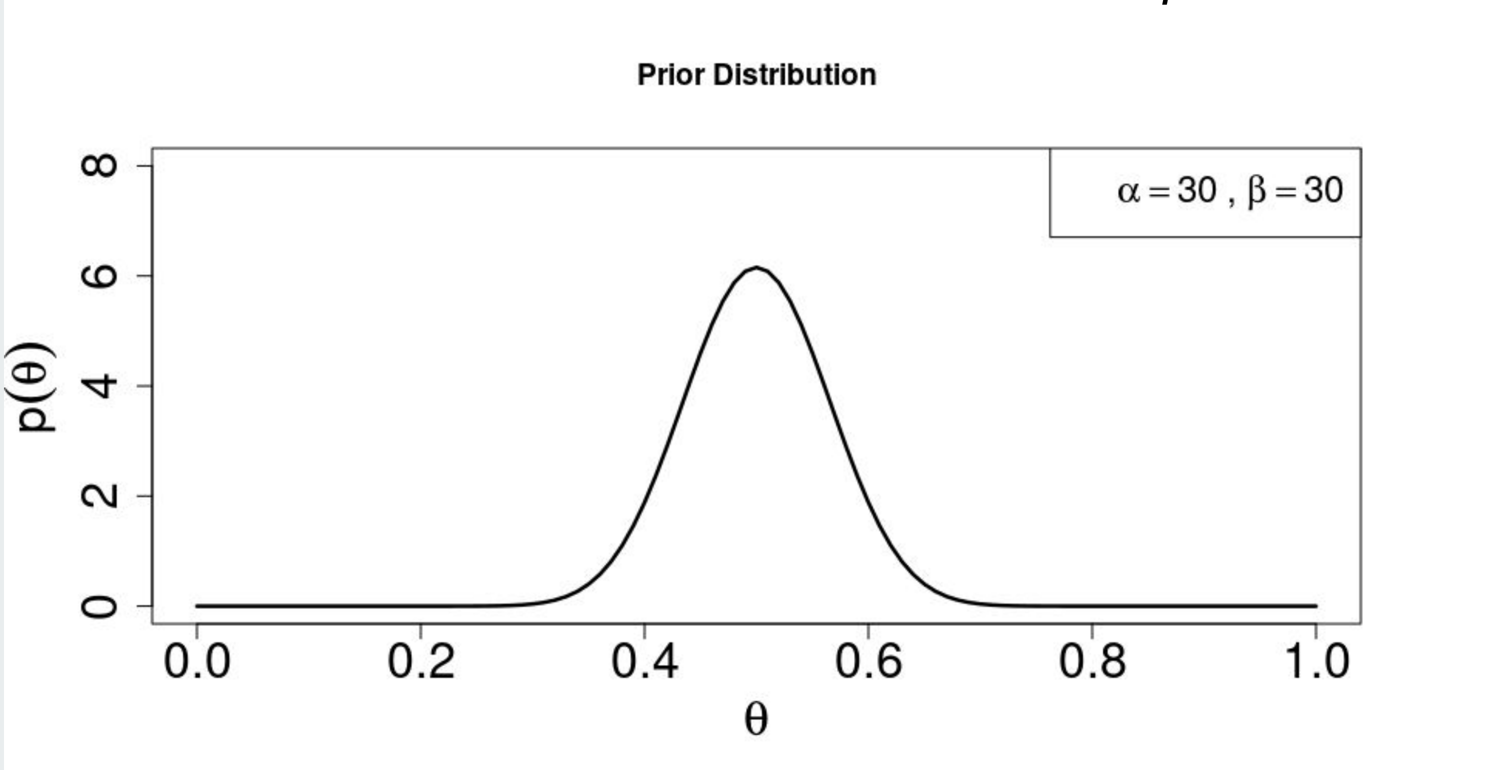
1. Write a script to plot the beta distribution given the alpha and beta parameter values (*Hint:* check out ?Distributions)

2. What happens when: $\alpha = \beta$? $\alpha = \beta = 1$?

3. Find values for the α and β hyperparameters that best match your sketched prior distribution

Now you have a prior:

$p(\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$



$$p(\theta|y) \propto p(y|\theta)\, p(\theta)$$

Binomial Distribution:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{y} (1-\boldsymbol{\theta})^{n-y}$$

$$p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

Binomial Distribution:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{y} (1-\boldsymbol{\theta})^{n-y}$$

Beta Distribution:

$$p(\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$$

$$\boldsymbol{p(\theta|y)} \propto \boldsymbol{p(y|\theta)}\,\boldsymbol{p(\theta)}$$

Binomial Distribution:

$$p(\boldsymbol{y}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{y}(1-\boldsymbol{\theta})^{n-y}$$

Beta Distribution:

$$p(\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$$

$$\boldsymbol{p(\theta|y) \propto p(y|\theta)\, p(\theta)}$$

Simplify the expression

Binomial Distribution:

$$p(\mathbf{y}|\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{y}\,(1-\boldsymbol{\theta})^{n-y}$$

Beta Distribution:

$$p(\boldsymbol{\theta}) \propto \boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$$

$$\mathbf{p(\theta|y) \propto p(y|\theta)\; p(\theta)}$$

$$p(\theta|y) \propto \theta^{\,y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

What kind of distribution is this? (5 minutes)

$$p(\theta|y) \propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

What kind of distribution is this?

$$p(\theta|y) \propto \theta^{y + \alpha - 1} (1 - \theta)^{n-y+\beta-1}$$

Write a script that will plot $p(\theta|y)$ given y, n, α, and β

Gather some data!

Gather some data! (10 minutes)

- Take out the m&m’s (n=?)
- Record number of each colour
- Plot the posterior

Think-Pair-Share (10 minutes)

- Is the posterior distribution what you expected?
- Compare the posterior distribution to the prior distribution
- Is this the result you expected, given six different colours? Does this result tell you about the percentages of the other colours?
- How sensitive is the posterior to the prior distribution?

Think-Pair-Share

- How would you expect the posterior to change given more data?

Think-Pair-Share

- How would you expect the posterior to change given more data?

- Let's pool all the data to find out!
    - What assumptions are we making here?

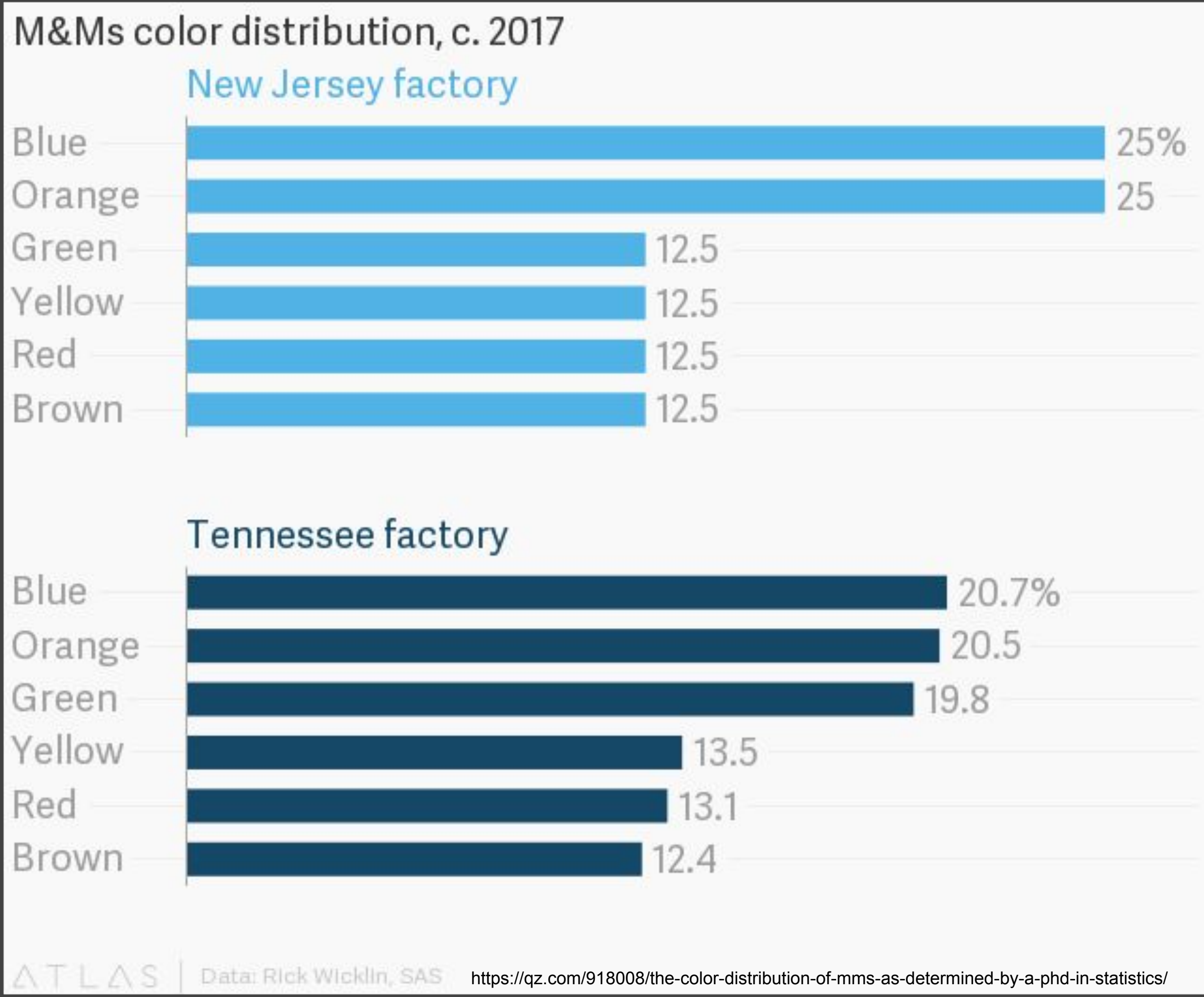
Let’s try looking at the red ones

- How many were red in the bag?
- Calculate the posterior for the red ones

Surprise twist!

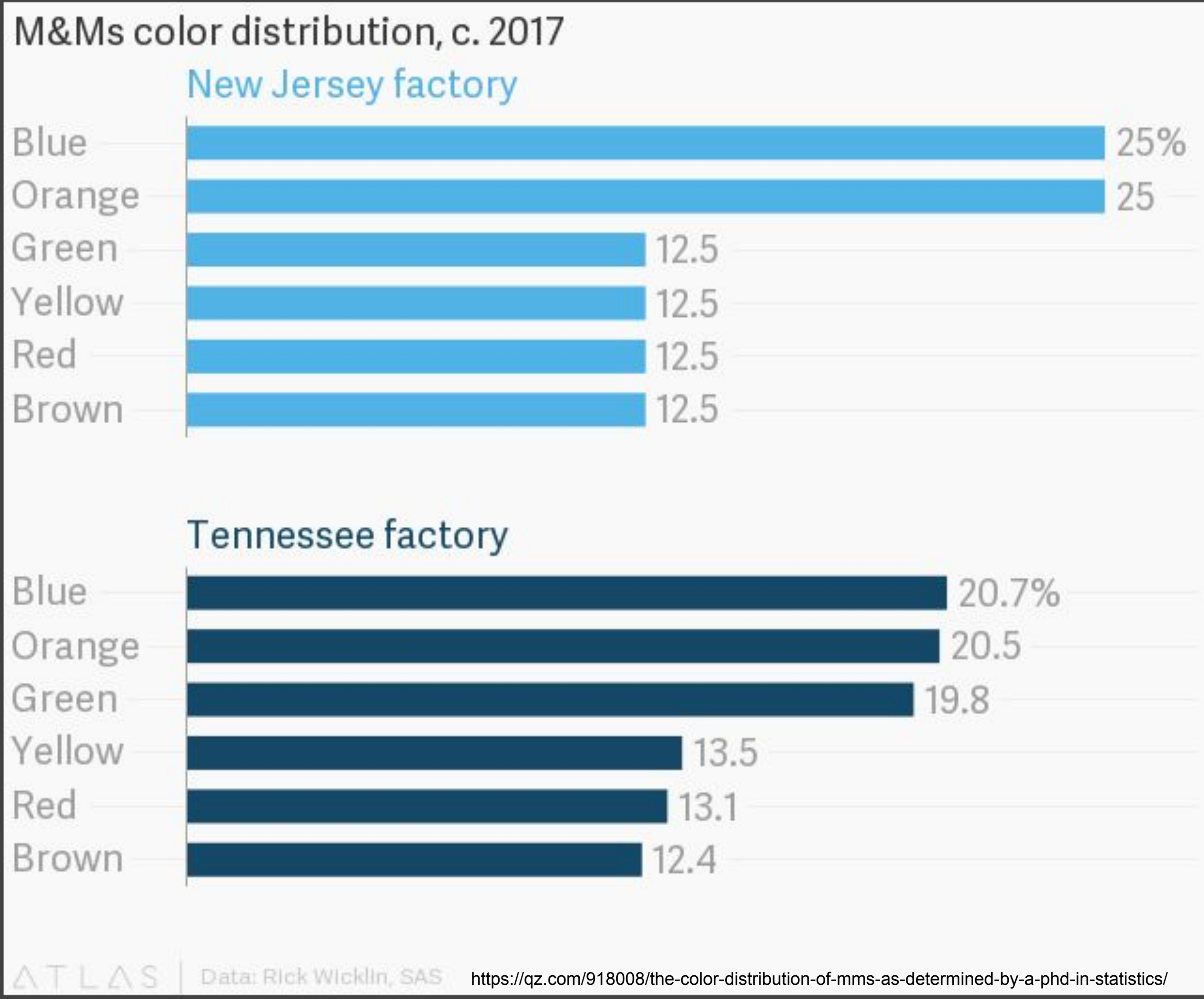
Surprise twist!

Different factories make different colour distributions of m&m’s !



https://qz.com/918008/the-color-distribution-of-mms-as-determined-by-a-phd-in-statistics/

Which factory did our m&m's come from?



https://qz.com/918008/the-color-distribution-of-mms-as-determined-by-a-phd-in-statistics/

New Jersey = HKP

Tennessee = CLV

Thomas Bayes
(1701-1761)

Image: Evan-Amos, Vanamo Media

What if we wanted to model all the colours simultaneously?

What distribution might we use for the likelihood?

Image: Evan-Amos, Vanamo Media

What if we wanted to model all the colours simultaneously?

What distribution might we use for the likelihood?

Multinomial Distribution

Wikipedia (different notation than we've been using)

$$f(x_1, \ldots, x_k; p_1, \ldots, p_k) = \frac{\Gamma(\sum_i x_i + 1)}{\prod_i \Gamma(x_i + 1)} \prod_{i=1}^{k} p_i^{x_i}.$$

Conjugate prior to the multinomial distribution is the

Dirichlet distribution

Wikipedia (different notation than we've been using)

$$\frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{i=1}^{K} x_i^{\alpha_i - 1}$$

$$\mathrm{B}(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}$$

$$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$$

Toy problem

I hand out bags of m&m's, but the lot codes are hidden on the back.

Toy problem

I hand out bags of m&m's, but the lot codes are hidden on the back.

I tell you that some people have m&m's from New Jersey and others from Tennessee.

Toy problem

I hand out bags of m&m's, but the lot codes are hidden on the back.

I tell you that some people have m&m's from New Jersey and others from Tennessee.

But I don't tell you what proportion are from NJ!

Develop a model that will tell you

1. Which factory your m&m's bag came from

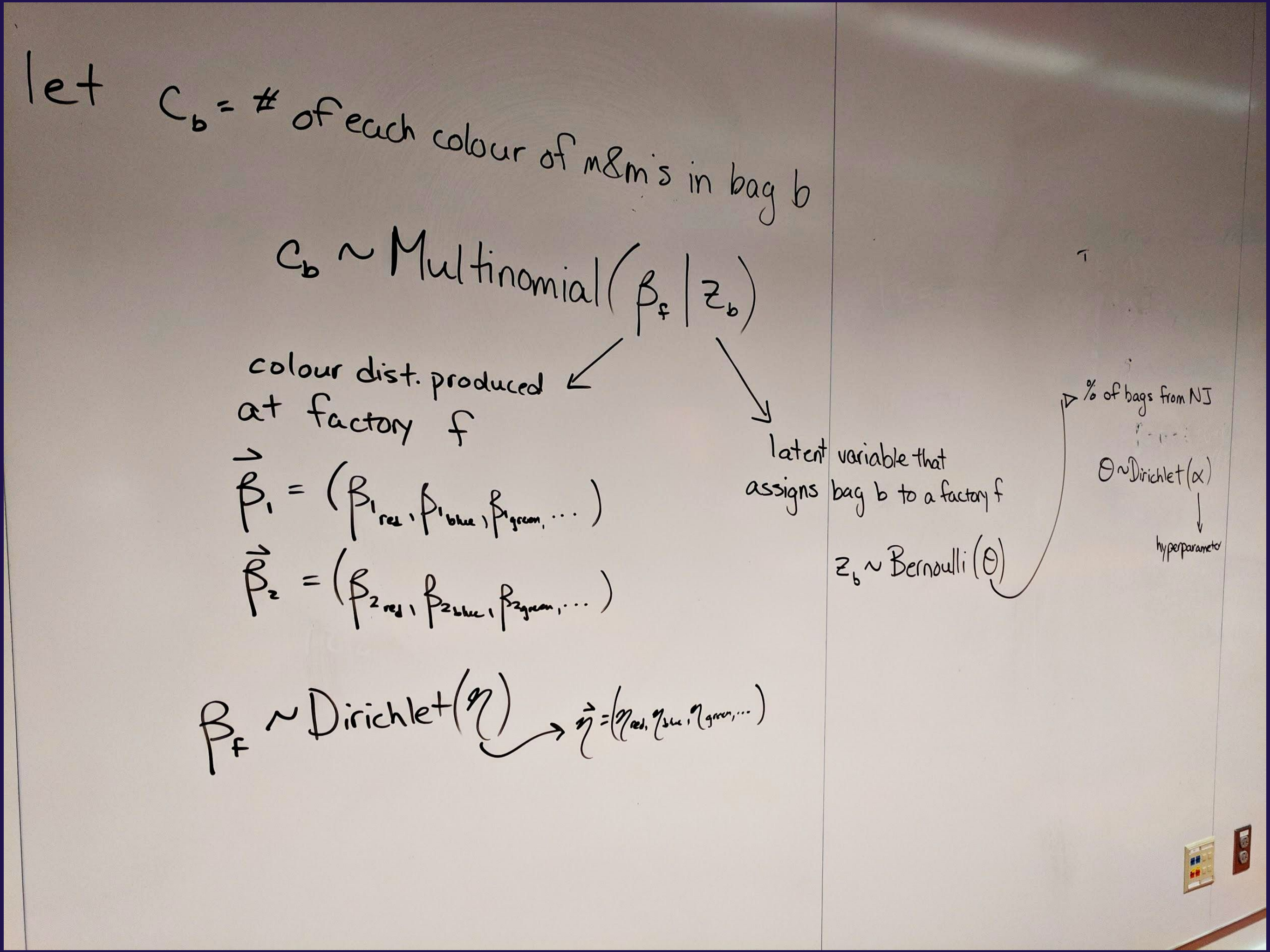
2. The proportion of bags from NJ

3. The colour distributions produced at each factory

Hierarchical Bayesian Model

You may visit the link below to input data from an <u>entire</u> bag of plain m&m's.
This will help us keep track of the colour distributions of m&m's for future classes
(please make sure to note the factory code!)

http://bit.ly/2Pc8ljx



(Picture taken of my notes, after class… need to make nicer slides of this!)