A practical introduction to Bayesian Inference

June 9, 2025

Prof. Gwendolyn Eadie

UofT Statistical Sciences Research Program

Bayes' Theorem

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})}$$

- $\boldsymbol{\theta}$ → vector of model parameters
- $\boldsymbol{y}$ → data

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

Likelihood

Posterior distribution

Prior distribution

Bayesian Inference

- Parameters are not fixed
- The quantity of interest is the distribution of the parameters
- We want to calculate exactly or, if we can't, estimate the *posterior distribution*

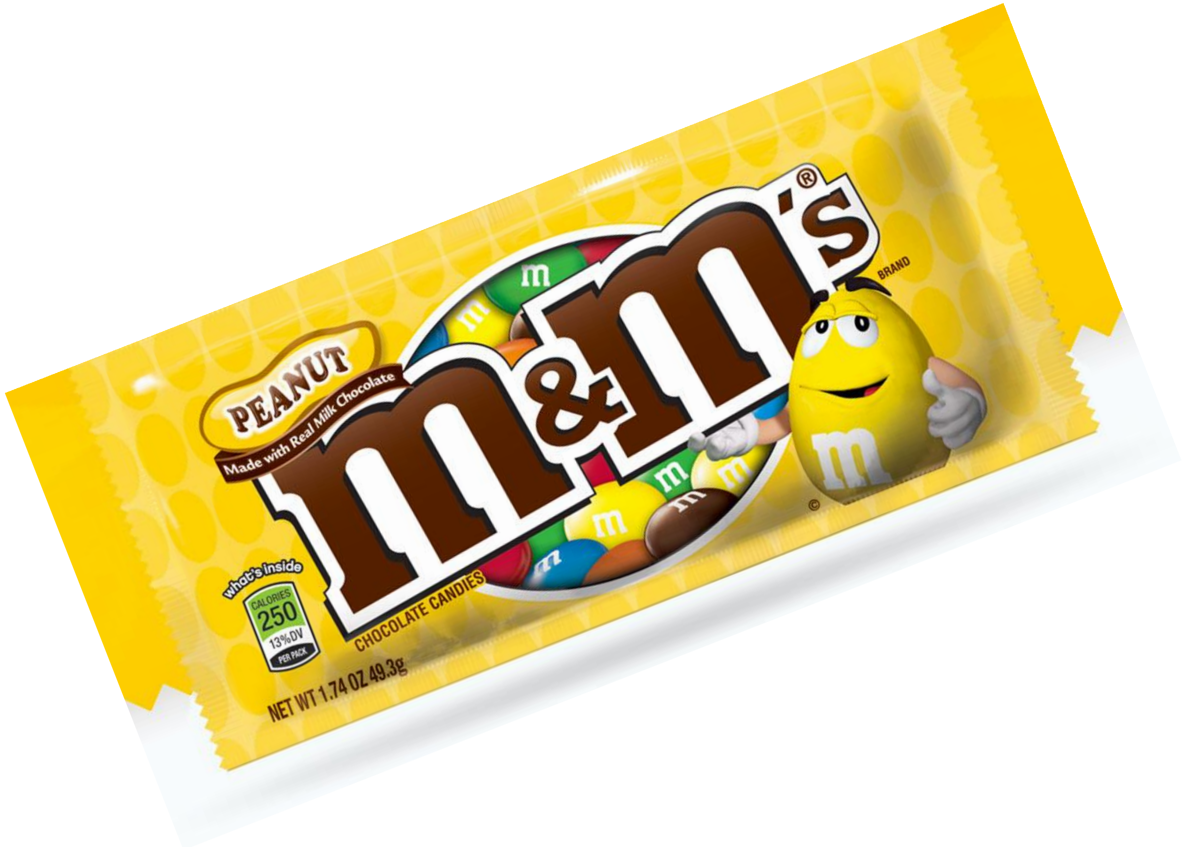
$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

- Computational methods for estimating the posterior
  - Grid of θ's → evaluate the posterior at every point, given the data ($$$)
  - Drawing samples from the posterior
    - E.g., Markov Chain Monte Carlo, Hamilton Monte Carlo, Gibbs Sampling, or some other sampling method
  - Variational Bayes
  - Other methods

Today's activity: Bayesian inference with m&m's

- Find the posterior distribution for the percentage of blue m&m's® made at the factory — *using Bayes' theorem and 15 candies from a bag of m&m's.*

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



No googling for answers!

That spoils the fun.

Initial exploratory questions

- What kind of data are the colours of m&m's?
  - Numerical
  - Categorical
  - Continuous
- How will you record the data?
- Will you sample with replacement or without?
- How might we model the probability of drawing a blue m&m?

What’s (y)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 (or at least think about) what your prior distribution for θ looks like

Sketch of a prior distribution for the percentage of blue m&m's made at the factory

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

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

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

Binomial Distribution:

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

Prior knowledge

- Outlined/sketch previously
- For convenience, use the conjugate prior
- find parameter values that make the conjugate prior look like our prior distribution sketch.

Beta Distribution:

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

Binomial Distribution:

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

Beta distribution:

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

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

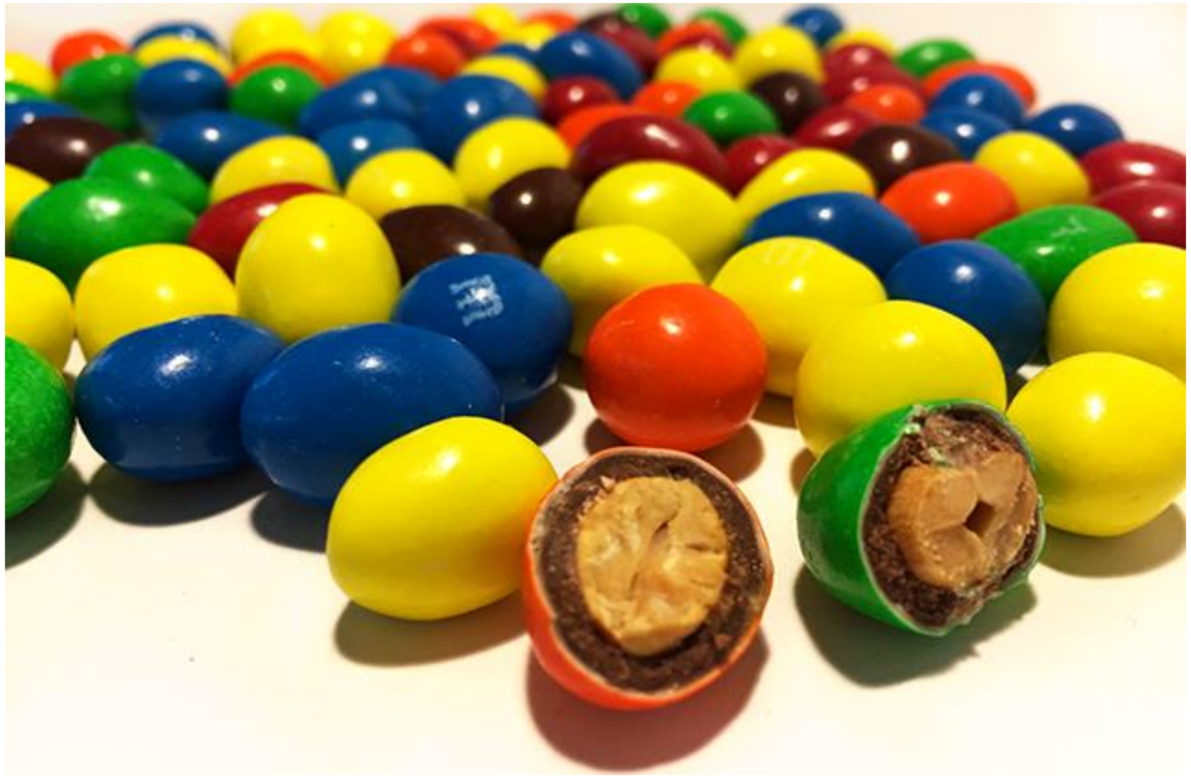
In-Class Activity

- Go to the R markdown or Jupyter notebook on Github

1. write a function for the prior that takes α and β as inputs
2. Write code to plot the prior distribution
3. Choose parameters to define the prior distribution
   a) adjust the α and β hyperparameters that best match your sketched prior distribution
4. Calculate the mean and variance of the prior you chose

Exercise

- Gather some data!
- Open your bag of m&m's
- Take out 15 , and record how many colours you have of each
- Perform Bayesian inference to predict the percentage of blue m&m's made at the factory



Please help us collect data!
bit.ly/mmcounts

Think-Pair-Share

- 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?

Pool the data from the entire class

- How would you expect the posterior to change given more data?
- **Let's pool the data and find out!**
  - (What assumptions are we making here?)
- Plot the posterior again
- Compare to prior