# Beta-Binomial Inference, Part 1

STAT 251, Unit 4A

# Overview

Review

Binomial Likelihood

Beta Prior Distribution, Assuming Binomial Likelihood

Themes in next unit (Unit 4B)

# Review

# Big picture outline for process in remainder of course

- Select (build, construct) a likelihood (model)
- Select prior
- Determine posterior distribution
- Make inference with posterior distribution.

# Notation

Notation: We will use

- $\mathbf{y}$ to generically represent the *data*,
- $\boldsymbol{\theta}$ to generically represent the *parameters*,
- $f(\mathbf{y}|\boldsymbol{\theta})$ to represent the *likelihood* of the data given the parameters
- $\pi(\boldsymbol{\theta})$ to represent the *prior distribution* of the parameters
- $\pi(\boldsymbol{\theta}|\mathbf{y})$ to represent the *posterior distribution* of the parameters given the observed data.

# Recap from Last Lecture's Activity

- Parameter of interest: $\theta \equiv$ proportion of class that answered yes.
- Three guesses on possible values for $\theta$ solicited from the class: $(0.25, \quad 0.5, \quad 0.75)$
- Prior probability of each value (per class vote): $(1/3, \quad 1/3, \quad 1/3)$
- Then, a Bernoulli observation: it is observed whether or not the first sample said "yes"; the result is $y_1$. (Note: If $\theta$ were given, $Pr(Y_1 = 1)$ would be $\theta$.)
- Calculated the posterior distribution of $\theta$ given the new data
  - Multiplied the prior probability of each value by the corresponding likelihood of the observed data and standardized by dividing by the marginal likelihood
- Included additional data and updated posterior again (and again)

# Binomial Likelihood

# Towards a More Realistic Binomial Example

As before, we will work with the *Binomial*$(n, \theta)$ likelihood; the pmf should look familiar:

$$f(y|\theta) = \frac{n!}{y!(n-y)!}\theta^y(1-\theta)^{n-y}, \quad \forall y \in \{0, 1, \ldots, n\}$$

Another way to write $\frac{n!}{y!(n-y)!}$ is $\binom{n}{y}$, which is read as *n choose y*.

$$f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, \quad \forall y \in \{0, 1, \ldots, n\}$$

# When do we use the binomial distribution as the likelihood?

Here is the setting under which we would use a binomial likelihood as a data-generating model:

- The random variable $y$ is interpreted as the total number of successes in a predetermined number, n, of trials
- Each trial has an identical chance of success, $\theta$.
- Given $\theta$, each trial outcome is independent of every other trial outcome.

# Computing the Binomial Likelihood in R

Use the function *dbinom*. To learn more about this R function, you can run the following

```
?dbinom
```

Essential information to supply:

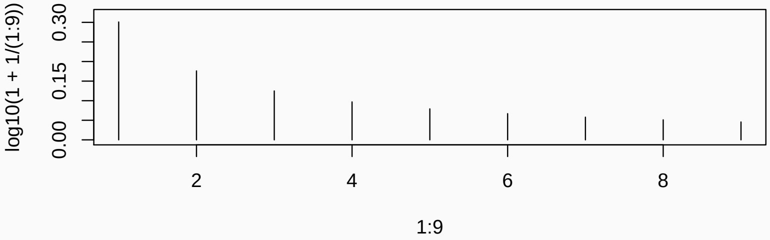
- value(s) of x (the number(s) at which to evaluate $f(\cdot|\theta)$ ),
- size (the number of trials; i.e., $n$)
- prob (what we have been calling $\theta$—the probability of success for a single trial)

```r
# Example: f(y=3 | theta=0.4) if n=7
 dbinom(3, 7, 0.4)

## [1] 0.290304

# or can be more explicit in referring to arguments:
dbinom(x=3, size=7, prob=.4, log=F)

## [1] 0.290304

##f(y|theta=0.8) if n=4, for y=0,1,2,3,4
dbinom(0:4, 4, .8)

## [1] 0.0016 0.0256 0.1536 0.4096 0.4096

dbinom(c(0,1,2,3,4), size=4, prob=0.8)
```

# Plotting a pmf

In a moment, you will be asked to plot the binomial likelihood for $n$=10 and $\theta$=0.9. As a precursor, consider the following plot

```
plot(1:9, log10(1+1/(1:9)), type="h", ylim=c(0, .32))
```



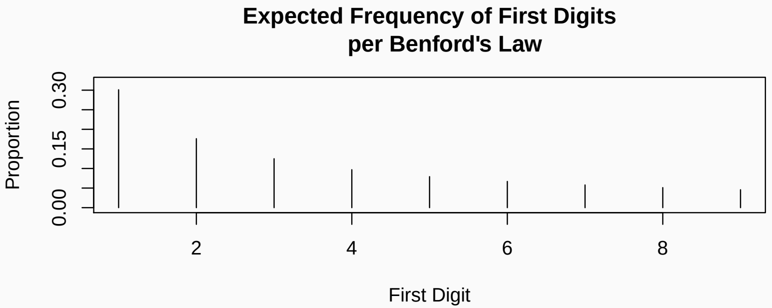
# Plotting a pmf (cont.)

$$plot(\underbrace{1:9}_{\text{x-axis values}}, \underbrace{log10(1+1/(1:9))}_{\text{y-axis values}}, \underbrace{type = "h"}_{\text{discrete}})$$

The type="h" option is used here because this will have a line segment from (x,0) to (x,y) for each (x,y) pair. For a discrete distribution, this is a preferred way to plot the distribution. See also `https://www.math.utah.edu/~treiberg/M3073BinPlot.pdf`.

Many options are available in the plot function, such as axis labels (*xlab=* , *ylab=*), chart title (*main=*), and lower/upper limits for the plot regions (xlim= , *ylim=*). Consider the following:

```
plot(1:9,log10(1+1/(1:9)),type="h",
     main="Expected Frequency of First Digits
     per Benford's Law",xlab="First Digit",
     ylab="Proportion",ylim=c(0,0.32))
```



# Your turn

- Plot the binomial likelihood for $n$=10 and $\theta$=0.9. Make sure that it is properly formatted to be a self-explanatory graphic.
- Plot the binomial likelihood for $n$=10 and $\theta$=0.85.
- If n=10 and y= 7, which value of $\theta$, 0.9 or 0.85, leads to a higher likelihood (i.e., a higher value of $f(y|\theta)$)?

# Properties of the Binomial Distribution

If $Y|\theta$ has the *binomial*$(n, \theta)$ distribution, then

- $f(y|\theta) = Prob(Y = y) = \binom{n}{y}\theta^y(1-\theta)^{n-y}\mathbb{1}_{(y\in\{0,1,\ldots,n\})}$
- $E(Y) =$ mean of Y $= n\theta$
- $Var(Y) =$ variance of Y $= n\theta(1-\theta)$
- If $n\theta \geq 10$ and $n(1-\theta) \geq 10$, then the pmf of y will look much like a discrete version of the normal distribution.

Recall our activity from last lecture. Realistically, there are more than the three possible values of $\theta$ that we considered.

Let's move to a different possibility: that $\theta$ can be **any** number between 0 and 1, inclusive.

What value of $\theta$ would then maximize the likelihood? (This is known as the Maximum Likelihood Estimate, or MLE).

$$\begin{aligned}\hat{\theta} &\equiv \arg\max_{\theta} \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \mathbb{1}_{(0 \leq \theta \leq 1)} \\ &= y/n\end{aligned}$$

Note: The expression $\mathbb{1}_{0 \leq \theta \leq 1}$ is what is known as an indicator function: it has the value of 1 if the condition $0 \leq \theta \leq 1$ is met, and a 0 otherwise.

Binomial likelihood takes $\theta$ as given, and thus doesn't convey any information about whether $\theta$ is discrete or continuous.

We want to consider that $\theta$ could be anything in $[0, 1]$.

Beta distribution to the rescue!

# Beta Prior Distribution, Assuming Binomial Likelihood

# Beta Distribution

Beta distribution is characterized by knowing two values: $a > 0$ and $b > 0$. It has all of its probability on the interval [0,1]. If $\theta \sim \text{Beta}(a, b)$ then the pdf of $\theta$ is:

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} \mathbb{1}_{0 \le \theta \le 1}$$

This uses the Gamma function, $\Gamma(\cdot)$, which is defined as

$$\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx$$

In R, use *dbeta* to evaluate the pdf of the beta distribution; see *?dbeta* for further information.

If you ever want to evaluate $\Gamma(u)$ in R, use *gamma(u)*.

# Summaries of the beta distribution

If $\theta \sim Beta(a, b)$ (that is, if $\theta$ has the Beta(a,b) distribution),

- Mean ($E(\theta)$, or $\mu$): $\frac{a}{a+b}$
- Var($\theta$): $\frac{ab}{(a+b)^2(a+b+1)}$
- Mode: $\frac{(a-1)}{(a+b-2)}$ (so long as $a > 1$ and $a + b > 2$ )

The variance formula can also be expressed as $\frac{\mu(1-\mu)}{(a+b+1)}$. Here it is easy to see that for a given mean, the variance will decrease as $a + b$ increases.

# Your Turn

For each of the following distributions, plot the pdf and report the mean, variance, and (if it exists) mode of the distribution.

- Beta(1,1)
- Beta(3,7)
- Beta(100, 74)
- Beta(0.2, 5)
- Beta(0.5, 0.5)

Also, consider what each of the distributions imply about our beliefs regarding $\theta$.

# Themes in next unit (Unit 4B)

# Beta-Binomial Result

- If we have
  - a binomial likelihood (that is, the model for $y|\theta$ is $f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$ for $y \in \{0, 1, \ldots, n\}$

  AND
  - a beta distribution for the prior distribution of $\theta$ (i.e., $\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\mathbb{1}_{(0\leq\theta\leq 1)}$),

  THEN
  - the posterior distribution, $\pi(\theta|y)$, will be another beta distribution (which can be written as $Beta(a^\star, b^\star)$, where $a^\star = a + y$ and $b^\star = b + (n-y)$).

  What do we call this property of a prior? (We say a prior distribution is **conjugate** for a given likelihood if the posterior distribution must be in the same family of distributions as the prior distribution if the given likelihood is assumed.)

# In Beta-Binomial, part 2

We'll derive the result from the previous slide. In doing so, we'll discuss how this choice of prior (the beta distribution) leads to a recognizable distribution for the posterior when we have binomial data.

We'll discuss how to choose a and b to match prior beliefs about a success probability $\theta$.

We'll practice making inference using the Bayesian paradigm.