

# Introduction to Common Distributions

Rebecca C. Steorts

## Reading

BMLR: Chapter 3:

<https://bookdown.org/roback/bookdown-BeyondMLR/>

I would like to thank Wenxin Guo for helping correct some typos.

# Agenda

- ▶ Bernoulli and Binomial Distribution
- ▶ Maximum Likelihood Estimation

## Traditional inference

You are given **data  $X$**  and there is an **unknown parameter** you wish to estimate  $\theta$

How would you estimate  $\theta$ ?

- ▶ Find an unbiased estimator of  $\theta$ .
- ▶ Find the maximum likelihood estimate (MLE) of  $\theta$  by looking at the likelihood of the data.
- ▶ In later classes, STA 402, you will consider how to estimate  $\theta$  when it's random

## Bernoulli distribution

The Bernoulli distribution is very common due to binary outcomes.

- ▶ Consider flipping a coin (heads or tails).
- ▶ We can represent this a binary random variable where the probability of heads is  $\theta$  and the probability of tails is  $1 - \theta$ .

Consider  $X \sim \text{Bernoulli}(\theta) \mathbb{1}(0 < \theta < 1)$

The likelihood is

$$p(x | \theta) = \theta^x (1 - \theta)^{(1-x)} \mathbb{1}(0 < \theta < 1).$$

- ▶ Exercise: what is the mean and the variance of  $X$ ?
- ▶ What is the connection with the Bernoulli and the Binomial distribution?

## Bernoulli distribution

- ▶ Suppose that  $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(\theta)$ . Then for  $x_1, \dots, x_n \in \{0, 1\}$  what is the likelihood?

## Notation

- ▶  $x_{1:n}$  denotes  $x_1, \dots, x_n$

## Bernoulli and Binomial Connection

$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(\theta)$ .<sup>1</sup>

Suppose  $Y = \sum_{i=1}^n X_i$ . Then  $Y \sim \text{Binomial}(n, \theta)$ .<sup>2</sup>

---

<sup>1</sup>This represents  $n$  coin flips with success probability  $\theta$ .

<sup>2</sup>This represents  $n$  Bernoulli trials with success probability  $\theta$ .

## Likelihood

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(\theta).$$

$$\begin{aligned} p(x_{1:n} | \theta) &= \mathbb{P}(X_1 = x_1, \dots, X_n = x_n \mid \theta) \\ &= \prod_{i=1}^n \mathbb{P}(X_i = x_i \mid \theta) \\ &= \prod_{i=1}^n p(x_i | \theta) \\ &= \prod_{i=1}^n \theta^{x_i} (1 - \theta)^{1 - x_i} \\ &= \theta^{\sum x_i} (1 - \theta)^{n - \sum x_i}. \end{aligned}$$

## Traditional inference

You are given **data**  $X$  and there is an **unknown parameter** you wish to estimate  $\theta$

How would you estimate  $\theta$ ?

- ▶ Find an unbiased estimator of  $\theta$ .
- ▶ Find the maximum likelihood estimate (MLE) of  $\theta$  by looking at the likelihood of the data.
- ▶ Suppose that  $\hat{\theta}$  estimates  $\theta$ .

Note:  $\hat{\theta}$  may depend on the data  $x_{1:n} = x_1, \dots, x_n$ .

## Unbiased Estimator

Recall that  $\hat{\theta}$  is an **unbiased estimator** of  $\theta$  if

$$E[\hat{\theta}] = \theta. \quad (1)$$

## Maximum Likelihood Estimation

For each sample point  $x_{1:n}$ , let  $\hat{\theta}$  be a parameter value at which  $p(x_{1:n} | \theta)$  attains its maximum as a function of  $\theta$ , with  $x_{1:n}$  held fixed.

A **maximum likelihood estimator** (MLE) of the parameter  $\theta$  based on a sample  $x_{1:n}$  is  $\hat{\theta}$ .

## Finding the MLE

The solution to the MLE are the possible candidates ( $\theta$ ) that solve

$$\frac{\partial \log p(x_{1:n} | \theta)}{\partial \theta} = 0. \quad (2)$$

The solution to equation 2 are only **possible candidates** for the MLE since the first derivative being 0 is a **necessary condition** for a maximum but not a sufficient one.

Our job is to find a global maximum.

Thus, we need to ensure that we haven't found a local one. (Show that the second derivative is  $< 0$ .)

## Exercise

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(\theta). \quad (3)$$

Exercise: Derive (step by step) that the MLE for  $\theta$  is  $\frac{1}{n} \sum_i x_i = \bar{x}$ .

## Exercise Solution

Fill this in on your own.

## Approval ratings of Obama

What is the proportion of people that approve of President Obama in PA?

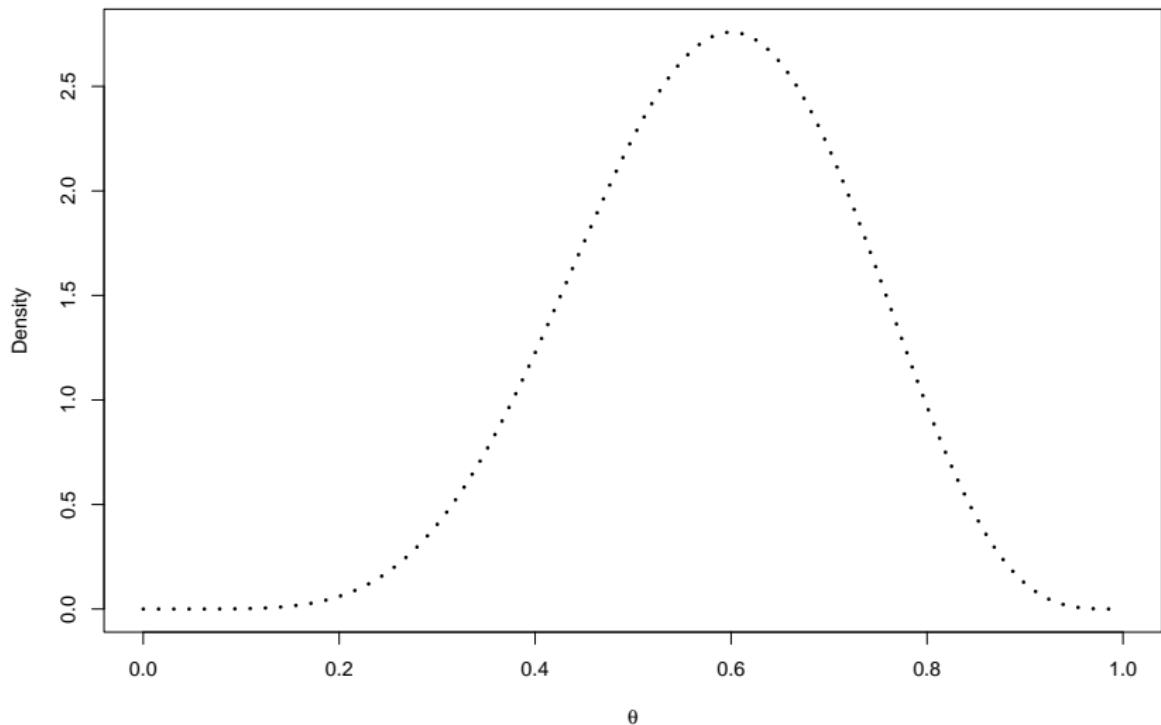
- ▶ We take a random sample of 10 people in PA and find that 6 approve of President Obama.

## Obama Example

```
n <- 10
th <- seq(0, 1, length = 500)
x <- 6
like <- dbeta(th, x + 1 , n - x + 1)
```

## Likelihood

```
plot(th, like, type = "l", ylab = "Density",
      lty = 3, lwd = 3, xlab = expression(theta))
```



## Supplemental Material

- ▶ Continuous Random Variables
- ▶ Discrete Random Variables

## Continuous Random Variables

A continuous random variable (r.v.) can take on an uncountably infinite number of values.

Given a probability density function (pdf),  $f(y)$ , allows us to compute

$$P(a \leq Y \leq b) = \int_a^b f(y) dy.$$

Properties of continuous random r.v.'s:

- ▶  $\int f(y) dy = 1$ .
- ▶ For any value  $y$ ,

$$P(Y = y) = \int_y^y f(y) dy = 0 \implies$$

$$P(y < Y) = P(y \leq Y).$$

## Discrete Random Variables

A discrete random variable has a countable number of possible values, where the associated probabilities are calculated for each possible value using a probability mass function (pmf).

A pmf is a function that calculates  $P(Y = y)$ , given each variable's parameters.

## Common Discrete distributions

- ▶ Bernoulli/Binomial (already covered)
- ▶ Poisson
- ▶ Geometric
- ▶ Negative Binomial
- ▶ Hypergeometric

## Common Continuous distributions

- ▶ Exponential
- ▶ Beta
- ▶ Gamma
- ▶ Normal (Gaussian)

## Beta distribution

The Beta distribution is frequently used in situations where the data are constrained to the interval  $[0, 1]$ . It often used to model proportions, rates, and probabilities.

Examples:

- ▶ In manufacturing, the proportion of defective items in a batch is a common quantity that can be modeled using the Beta distribution.
- ▶ In finance, the proportion of a portfolio invested in risky assets (such as stocks) is typically between 0 and 1.
- ▶ The Beta distribution is often used as a prior for the probability of success in Bernoulli or Binomial experiments in Bayesian statistics (STA 402).

## Beta distribution

Given  $a, b > 0$ , we write  $\theta \sim \text{Beta}(a, b)$  to mean that  $\theta$  has pdf

$$p(y) = \text{Beta}(y|a, b) = \frac{1}{B(a, b)} y^{a-1} (1-y)^{b-1} \mathbb{1}(0 < y < 1),$$

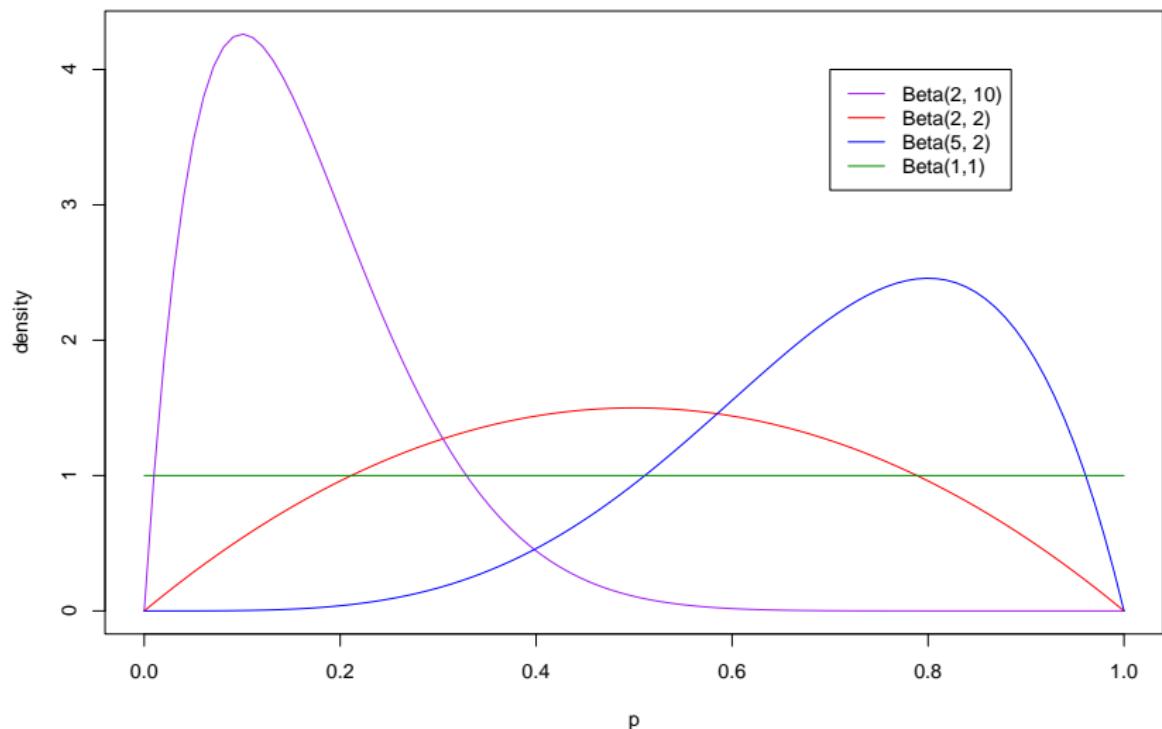
i.e.,  $p(y) \propto y^{a-1} (1-y)^{b-1}$  on the interval from 0 to 1.

- ▶ Here,

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

- ▶ Parameters  $a, b$  control the shape of the distribution.
- ▶ This distribution models random behavior of percentages/proportions.

## Beta distribution



## Beta distribution example

Suppose that a college models probabilities of student accepting admission via the  $\text{Beta}(a, b)$  distribution, where  $a, b > 0$  are fixed and known.

What is the probability that a randomly selected student has prob of accepting an offer larger than 80 percent, where  $a=4/3$  and  $b=2$ .

```
pbeta(0.8, shape1 = 4/3, shape2 = 2, lower.tail = FALSE)
```

```
## [1] 0.05930466
```

## Exponential distribution

Data that follows an Exponential distribution typically represents the time between events in a Poisson process, where events happen at a constant average rate and are independent of each other.

The Exponential distribution is widely used in various fields to model waiting times, lifetimes, and inter-arrival times.

Examples: time until a device fails, time between arrivals in a line, time between arrivals/departures, among others.

## Exponential distribution

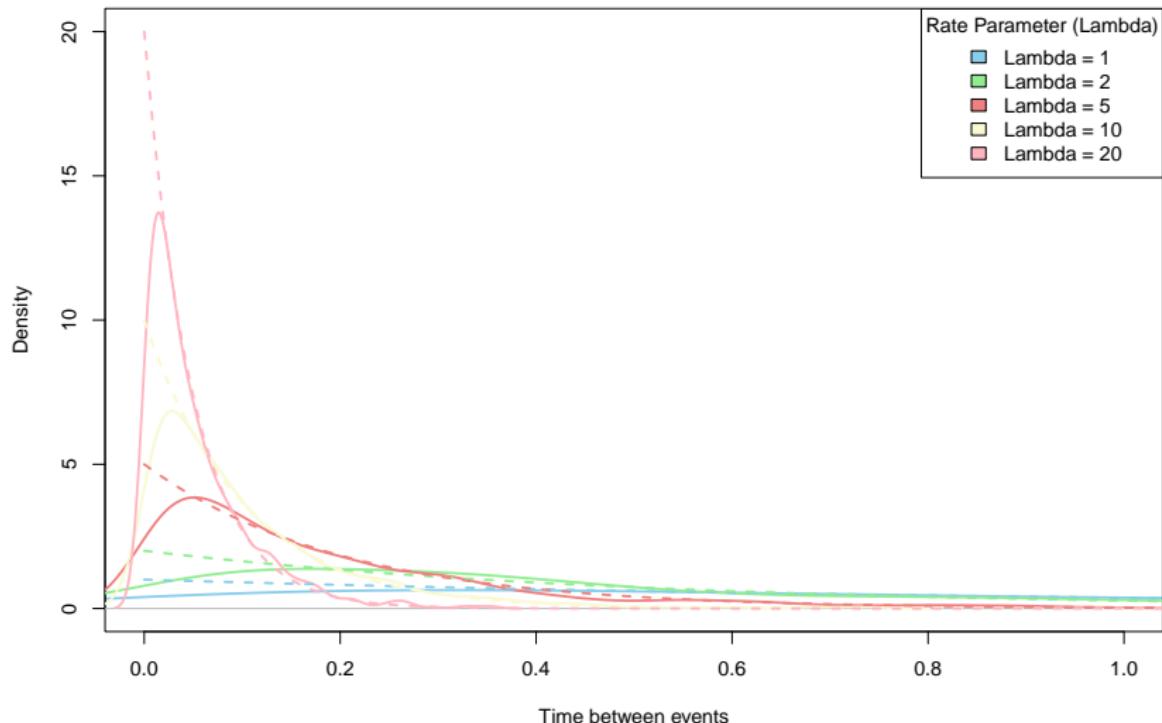
Assume  $\lambda > 0$ , which is the called rate parameter (the rate at which some event occurs).

The density function is given by

$$p(y) = \text{Exp}(y \mid \lambda) = \lambda \exp^{-\lambda y} I(y > 0).$$

# Exponential distribution

Density Curves of Five Exponential Distributions



## Gamma distribution

The Gamma distribution is a continuous probability distribution that is often used to model waiting times, lifetimes, and other phenomena where the events are continuous and the process involves a sum of exponentially distributed random variables.

The Gamma distribution is commonly used in reliability theory, queueing theory, Bayesian statistics, and life data analysis.

## Rainfall example

The Gamma distribution is used to model the accumulated rainfall over a given period, particularly in areas where precipitation events occur at a constant rate.

The total accumulated rainfall over a month could be modeled as a Gamma distribution, where the shape parameter  $k$  reflects the number of significant rainfall events, and the rate  $\lambda$  represents the intensity of the rainfall.

For example, the accumulated rainfall in a region that experiences 10 or more rainstorms per year, with an average rainfall rate of 0.5 inches per storm, could be modeled as a  $\text{Gamma}(10, 0.5)$  distribution.

## Queueing Systems (Time Until $k$ Customers Arrive)

The Gamma distribution is used to model the waiting time for the occurrence of  $k$  events, such as the arrival of  $k$  customers at a service station.

In a service system where customers arrive at an average rate of  $\lambda$  per minute, the time it takes for the system to serve  $k$  customers is modeled as a Gamma distribution with shape parameter  $k$  and rate parameter  $\lambda$ .

For example, the time to serve 4 customers in a queue, where customers arrive at a rate of 2 per minute, could be modeled with a  $\text{Gamma}(4, 2)$  distribution.

## Gamma distribution (shape, rate)

Assuming shape parameter  $k$  and rate parameter  $\lambda$ , the density function is

$$f(y \mid k, \lambda) = \text{Gamma}(y \mid k = \text{shape}, \lambda = \text{rate}) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{\Gamma(k)}, \quad y \geq 0$$

This parameterization tends to be more common in Bayesian statistics and other applied fields. However, there exists another parameterization for other contexts.

## Gamma distribution (shape, scale)

Assuming shape parameter  $k$  and scale parameter  $\theta = 1/\lambda$ , the density function is

$$f(y \mid k, \theta) = \text{Gamma}(y \mid k = \text{shape}, \theta = \text{scale}) = \frac{y^{k-1} e^{-y/\theta}}{\Gamma(k)\theta^k}, \quad y \geq 0$$

Summary of the Gamma distribution:

[https://en.wikipedia.org/wiki/Gamma\\_distribution](https://en.wikipedia.org/wiki/Gamma_distribution)