Social Data Institute

UCL

Continuing Professional Development (CPD) course

Introduction To Bayesian Inference & Modelling (June 2025)

DAY 1: INTRODUCTION TO PROBABILITY DISTRIBUTION

Dr Anwar Musah (a.musah@ucl.ac.uk)
Lecturer in Social and Geographic Data Science
UCL Geography

Additional details:
https://www.ucl.ac.uk/social-data

About the course

Description of the course

- This course will introduce the foundational and advanced topics on statistical modelling within a Bayesian framework that is highly applicable for any kinds of evidence–based research for academia or industry.

- This course broadly covers two areas:
  - How to use **Bayesian Statistics** for evidence-based research in making non-spatial and spatial predictions, as well as studying the associations between various risk factors and outcomes through different regression-based models.
  - Learning how to perform **Bayesian Statistics** through a software called **Stan (i.e., rstan package) which interfaces with RStudio**. Here, you will learn a new programming tool.

Course overview [1]

- **Day 1**: Introduction to Probability Distributions
- **Day 2**: Introduction to Bayesian Inference
- **Day 3**: Bayesian Generalised Linear Models (GLMs)
- **Day 4**: Bayesian Hierarchical Regression Models
- **Day 5:** Spatial Bayesian Risk Modelling

All the reading list, lecture notes, dataset and information for the computer practicals, as well as their associated live recordings are hosted on a dedicated web-page for this course.

Link: https://uclpg-msc-sgds.github.io/UCL-SODA-CPD-RStan/

Welcome

Timetable, Schedule & Location

My Contact Details

Reading List

1 Installation of R, RStudio & Stan

The course will cover the following topics:

Date	Downloadables	Topics
	Not Applicable	**Installation of R, RStudio & Stan**
09/06/2025	Slides;	**Introduction to Probability Distributions**
10/06/2025	Slides;	**Introduction to Bayesian Inference**
11/06/2025	Slides; [Dataset]	**Bayesian Generalised Linear Models**
12/06/2025	Slides; [Dataset]	**Bayesian Hierarchical Regression Models**
13/06/2025	Slides; [Dataset]	**Spatial Bayesian Risk Modelling**

Solutions: [Day 1] | [Day 2] | [Day 3] | [Day 4] | [Day 5]

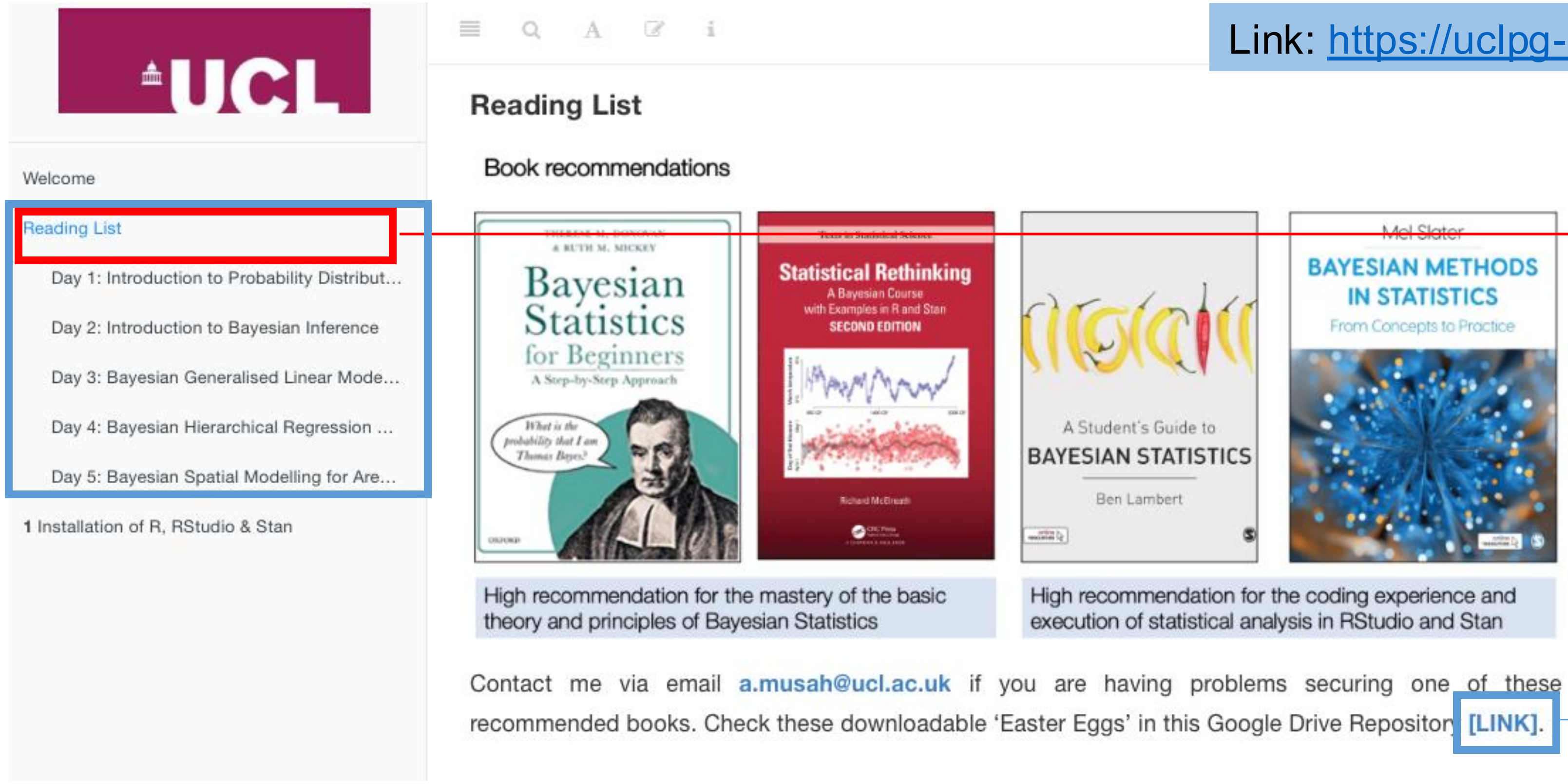
You can download the lecture and practical materials at this section from the above link

Course overview [2]

- **Day 1**: Introduction to Probability Distributions
- **Day 2**: Introduction to Bayesian Inference
- **Day 3**: Bayesian Generalised Linear Models (GLMs)
- **Day 4**: Bayesian Hierarchical Regression Models
- **Day 5:** Spatial Bayesian Risk Modelling

All the reading list, lecture notes, dataset and information for the computer practicals, as well as their associated live recordings are hosted on a dedicated web-page for this course.

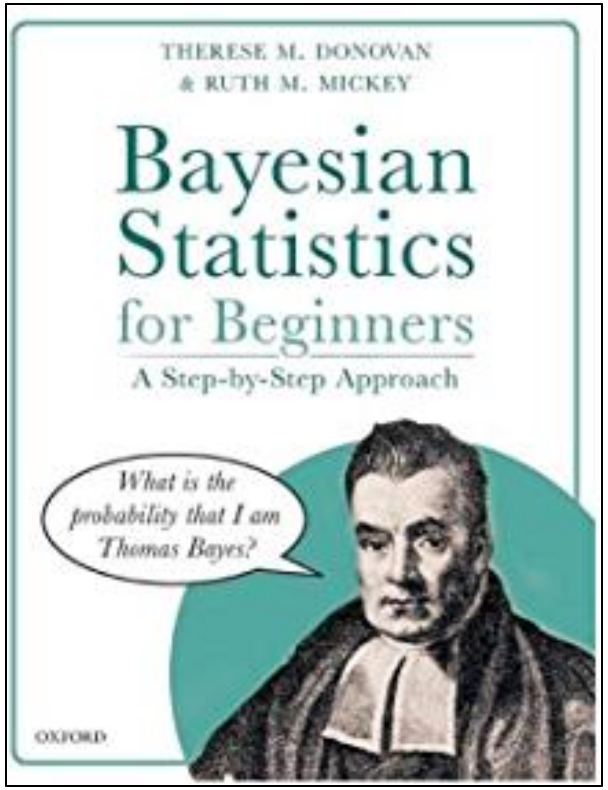
Link: https://uclpg-msc-sgds.github.io/UCL-SODA-CPD-RStan/

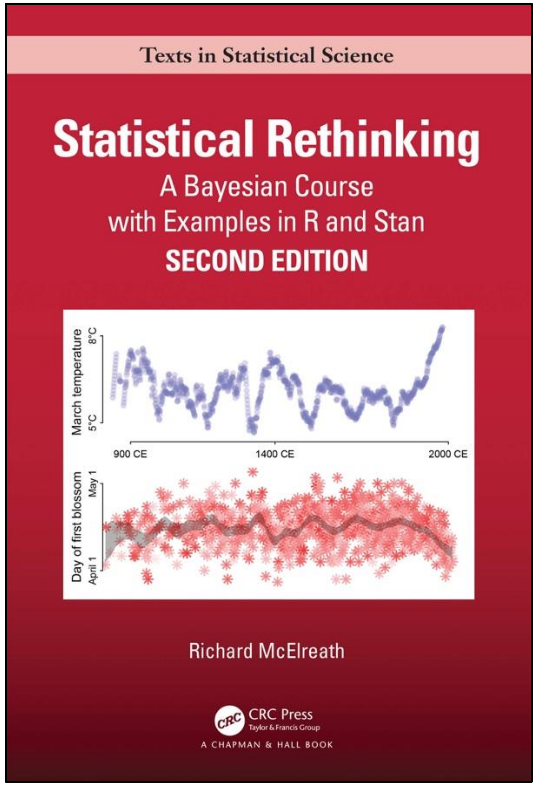


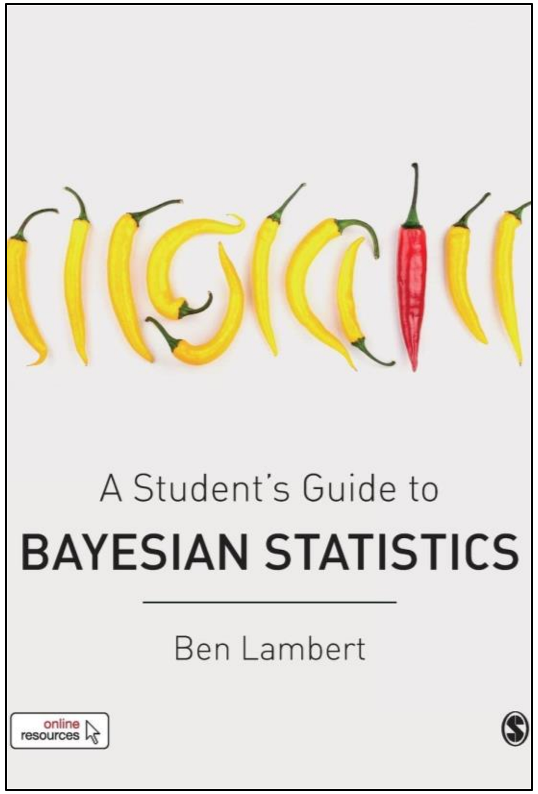
You can access to the recommended reading list for that day's session

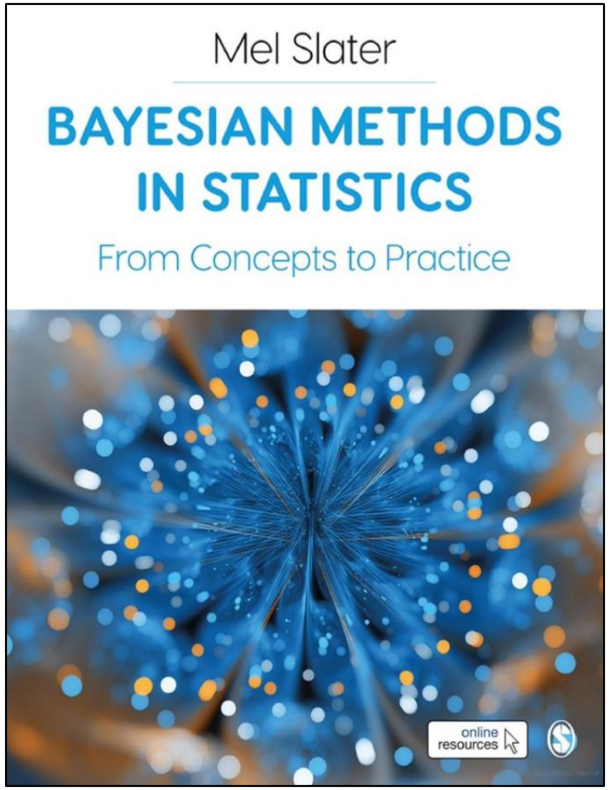
You can access the PDF books here

Book recommendations









High recommendation for the mastery of the basic theory and principles of Bayesian Statistics

High recommendation for the coding experience and execution of statistical analysis in RStudio and Stan

Course overview [3]

- **Day 1**: Introduction to Probability Distributions
- **Day 2**: Introduction to Bayesian Inference
- **Day 3**: Bayesian Generalised Linear Models (GLMs)
- **Day 4**: Bayesian Hierarchical Regression Models
- **Day 5:** Spatial Bayesian Risk Modelling

All the reading list, lecture notes, dataset and information for the computer practicals, as well as their associated live recordings are hosted on a dedicated web-page for this course.

Link: https://uclpg-msc-sgds.github.io/UCL-SODA-CPD-RStan/

UCL

Welcome

Content

Structure

People & Contact Details

Reading List

1 Getting started

2 Bayesian Generalised Linear Models (GLMs)

3 Bayesian Generalised Additive Models (GAMs)

4 Bayesian Hierarchical Regression Models

5 Bayesian Spatial Modelling for Areal Data in ...

Content

The course will cover the following topics:

Date	Downloadables	Topics
	Not Applicable	Installation
10/06/2023	[Slides] [Data]	**Introduction to Bayesian Inference**
11/06/2023	[Slides] [Data]	**Bayesian Generalised Linear Models**
12/06/2023	[Slides] [Data]	**Bayesian Hierarchical Regression Models**
13/06/2023	[Slides] [Data]	**Spatial Bayesian Risk Modelling**

Solutions: [Day 1] | [Day 2] | [Day 3] | [Day 4]

- Access to the computer practicals for that day
- The solution to the tasks for that day will be opened to you later

General format of the course

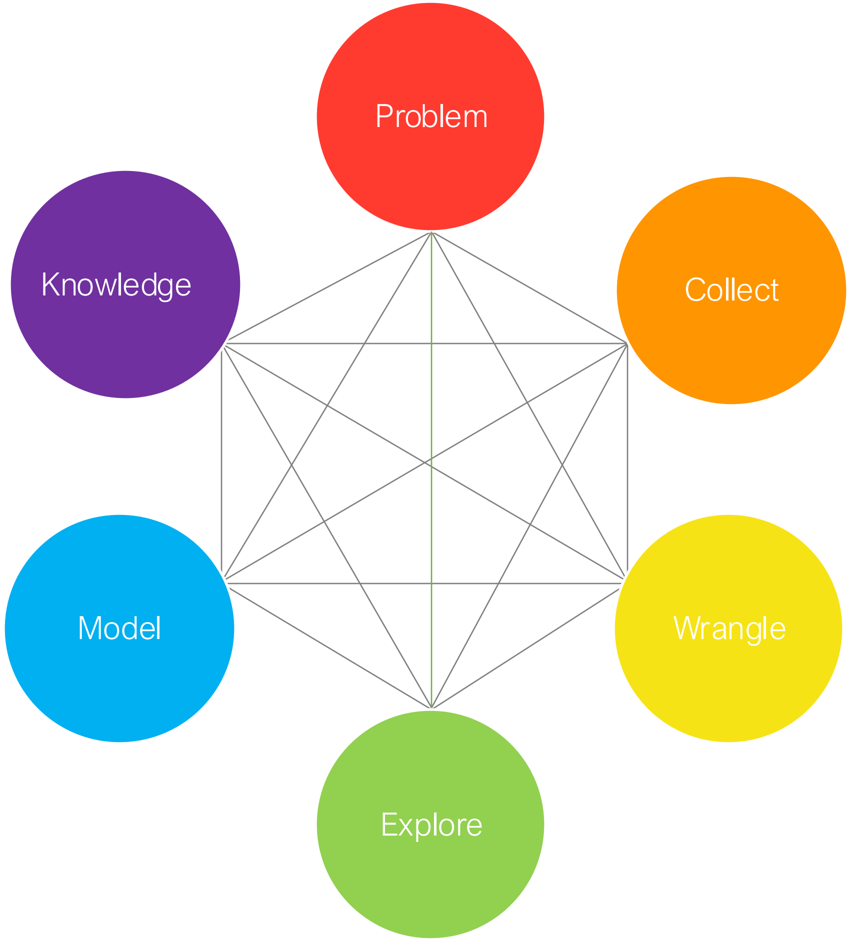
Times	Session	
10:30am to 12:00pm (1.5h)	Lecture	REC
12:00pm to 01:00pm (1h)	Lunch Break	
01:00pm to 02:00pm (1h)	Live Walkthrough & Demonstration	REC
02:00pm to 02:15pm (0h15mins)	Short Break	
02:15pm to 04:15pm (2h)	Computer Practical Session	
04:15pm to 04:30pm (0h15mins)	Wrap-up & Close	

Location(s): Charles Bell House (UCL) in Room G04 [Seminar Room 4], 43-45 Forley Street, W1W 7TY (nearest underground tube stations are Goodge Street and Oxford Circus)

NOTE: All lectures and tutorials sessions will be delivered in-person. You are expected to bring you own laptop with the required software and program installed.

Contents

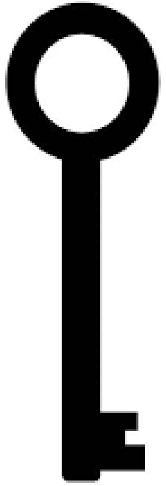
- What is Bayesian Statistics?
- Basic Probabilities – what are they & their types?
  - ❖ Unconditional, Joint, and Conditional Probabilities
  - ❖ How they work!
- What are Probability distributions?
  - ❖ Understand the features of a distribution
  - ❖ The various types of distributions
  - ❖ How they work!
- What is Bayes' Theorem?
  - ❖ How the above connects with the Bayes' rule
- Stan for Bayesian Inference



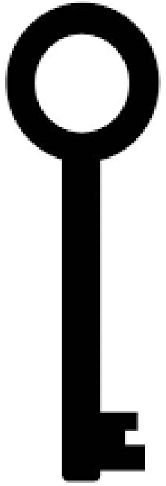
What is a Bayesian Statistics?

Definition:

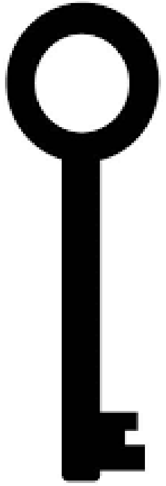
Bayesian statistics, is all about **uncertainty** (i.e., lack of complete sureness or knowledge). It is essentially the practice of expressing what you believe about something as a probability (**before observing that thing)**; and then using new evidence (**after observing that thing**) to update those beliefs thereafter.



Basic Probabilities



Probability Distribution



Bayes' Theorem

Bayesian Inference

This equation is the backbone of Bayesian Statistics

- User friendly formulation:

Posterior probability ∝ Likelihood Function x Prior Probability

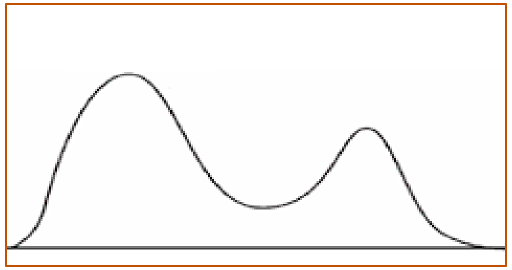
- Mathematical formulation:

Posterior: Probability of our new parameter being true **after** observing the data.

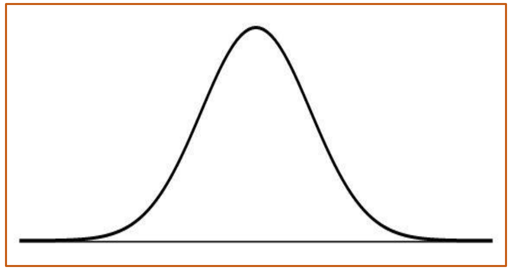
Likelihood: Plausibility of the observed data we have, given the parameter has a specific value(s)

Prior: Our belief about what the parameter is, <u>before</u> observing the data

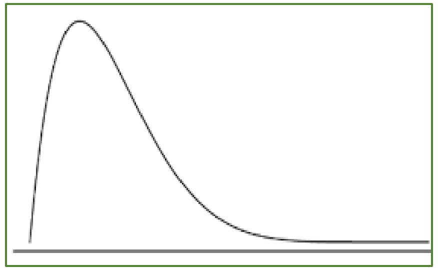
$$\Pr(\text{Parameter} \mid \text{Data}) \propto \Pr(\text{Data} \mid \text{Parameter}) \times \Pr(\text{Parameter})$$



$\propto$



$\times$



Notes:

- This equation comprise of conditionals, and marginal probability notation. So, an understanding of what **basic probabilities** are helps a lot.
- These marginal or conditional probabilities are expressed as **distributions**. In essence, you are defining, transforming and manipulating **probability distributions** to express and update uncertainty.
- This formulation is derived from the **Bayes' Rule**

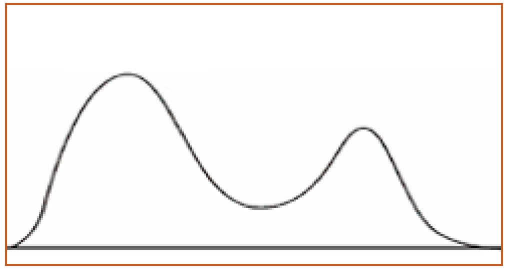
This equation is the backbone of Bayesian Statistics

Posterior: Probability of our new parameter being true **after** observing the data.

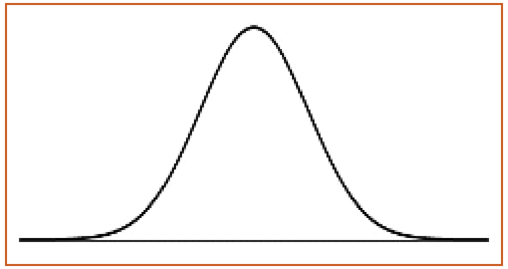
Likelihood: Plausibility of the observed data we have, given the parameter has a specific value(s)

Prior: Our belief about what the parameter is, **before** observing the data

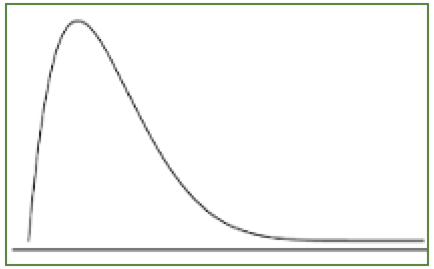
$$\Pr(\text{Parameter} \mid \text{Data}) \propto \Pr(\text{Data} \mid \text{Parameter}) \times \Pr(\text{Parameter})$$

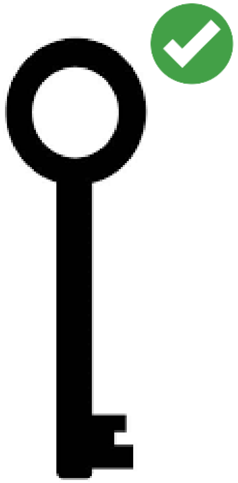


$\propto$

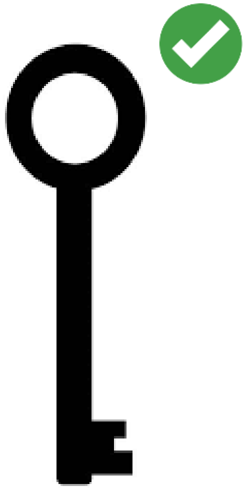


$\times$

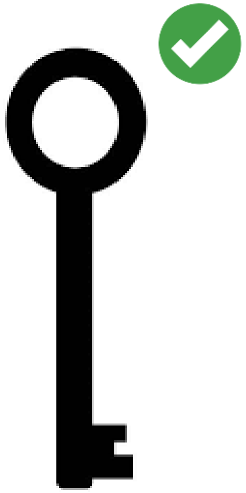




Basic Probabilities



Probability Distribution



Bayes' Rule

Bayesian Inference

Basic Probability – what is it?

Definition:

Probability is the language of uncertainty. It quantifies the uncertainty of a defined event, taking a value between 0 and 1. An event with a probability near zero implies that it is very unlikely to occur, while a probability near one implies that it is very likely.

- The usual notation for a probability is: **Pr( )** or **P( )** or **Prob( )**. We will use the notation for **Pr( )** throughout this lecture.
- A probability, **Pr( )** is always defined in relation to an event 'E' which has a set number of outcomes. The outcomes are calculated within a sample space (or population size), which is the total number of all possible outcomes that event 'E'
- We can write this and say **Pr(E)**, which is read as "**the probability that event E will occur**"

Note 1: When we say uncertainty, we are referring to lack of complete sureness (or knowledge) about an event.

Note 2: There are three key terms here, which will be defined in the next slide – An event, an outcome, and a sample space.

Basic terminology: event, outcome & sample space

In probability, the following means:

- An **outcome** simply refers to a single result that can emerge from a study.
- An **event** is referred to as the set of outcomes that share a common characteristic in a study.
- A **sample space** is the set of <u>all possible outcomes</u> observed in a study.

- Example 1: We are conducting a mosquito infestation survey in 7 households in a village, where a home is either **infested** or **not infested**. The dataset collected was as follows = **Infested, Not Infested, Not Infested, Not Infested, Infested, Not Infested, Infested**

  Outcome: Each home "Infested" or "Not infested" is an individual outcome, so in our dataset, the first entry "Infested" is one outcome and so on.

  Event: The event "<u>Infested</u>" consists of all 3 homes (i.e., 1st, 5th and 7th) that are infested i.e., the number of outcomes is 3 infested homes in this event
  Suppose the event is "<u>Not infested</u>", then it will consist of all 4 homes (i.e., 2nd, 3rd, 4th and 6th) that not infested. In this instances, its 4 outcomes in this event.

  Sample space: The possible outcomes that can emerge: {Infested, Not infested}
  All observed sample space that include 7 outcomes = {Infested, Not Infested, Not Infested, Not Infested, Infested, Not Infested, Infested}

Remember, **Pr(E), the probability that an event 'E' will occur.** This is what the notation for the above example will look like:

- The probability that a house is infested with mosquitoes = **Pr(Infested)**
- The probability that a house is not infested with mosquitoes = **Pr(Not infested)**

Basic terminology: event, outcome & sample space

In probability, the following means:

- An **outcome** simply refers to a single result that can emerge from a study.
- An **event** is referred to as the set of outcomes that share a common characteristic in a study.
- A **sample space** is the set of all possible outcomes observed in a study.

- Example 2: We are performing a survey to know the burden of Obesity by measuring BMI status of 10 people whose BMI can range anywhere between 15.0kg/m^2 to 66.0kg/m^2. The data collected was as follows = **18.9, 24.7, 32.4, 40.1, 21.4, 29.2, 24.5, 16.3, 19.7, 30.3**

Outcome: Each BMI value, such as 18.9, 24.7, or 32.4 and so on, is an individual outcome. The 8th entry 16.3 is one outcome.

Event: We are interested Obesity (BMI $\geq$ 30) therefore the observed values meet this criteria {32.4, 40.1, 30.3}. This event contains 3 outcomes.

Sample space: The possible outcome that can emerge can be any continuous BMI measurements within this range 15.0kg/m^2 to 66.0kg/m^2
All observed sample space that includes 10 BMI outcomes = {18.9, 24.7, **32.4, 40.1**, 21.4, 29.2, 24.5, 16.3, 19.7, **30.3**}

Remember, **Pr(E), the probability that an event 'E' will occur.** This is what the notation for the above example will look like:

- Probability that a person from this cohort is obese = **Pr(Obesity)**

Example 1: We are conducting a mosquito infestation survey in 7 households in a village. We want to know the probability that a house is infested with mosquitoes:

Event: The event "Infested" consists of all 3 outcomes (i.e., 1st, 5th and 7th home) that are infested
Sample space: All observed 7 outcomes = {**Infested**, Not Infested, Not Infested, Not Infested, **Infested**, Not Infested, **Infested**}

Probability that a house is infested with mosquitoes = Pr(Infested) = (set number of households infested/sample size) = 3/7 = 0.4285 = 42.85%

Example 2: We are performing a survey to know the burden of Obesity by measuring BMI status of 10 people. We want to know the probability that a person's obese:

Event: This event contains 3 outcomes that meet the criteria of obesity status (BMI $\geq$ 30) i.e., {32.4, 40.1, 30.3}.
Sample space: All observed 10 BMI outcomes = {18.9, 24.7, **32.4, 40.1**, 21.4, 29.2, 24.5, 16.3, 19.7, **30.3**}

Probability that a person from this cohort is obese = Pr(Obesity) = (set number of students BMI$\geq$30/sample size) = 3/10 = 0.30 = 30.00%

The above instance dealing with 'single events' are typical examples of an **Unconditional Probability.**

There are three major types of probabilities: **Unconditional**, **Joint** and **Conditional** probability.

Unconditional Probabilities

An unconditional probability is the chance (or likelihood) that a particular event will occur, without regards to external events that occurred in the past, or present or in the future.

- Usually, the set outcome of a single event can be affected by any number of factors; however, with unconditional probabilities, the likelihood of an event ending with a specific result does not account for other conditions that may affect it.
- An **Unconditional Probability** is also called a **'Marginal Probability'**
- When there's data - unconditional (or marginal) probabilities are calculated accordingly as follows:

Probability that E will occur

$$\Pr(\mathrm{E}) = \frac{\text{Number of observed outcomes (n)}}{\text{Total sample space (N)}} = \frac{\mathrm{n}}{\mathrm{N}}$$

Probability that E will NOT occur

$$\Pr(\mathrm{E}') = 1 - \left(\frac{\text{Number of trials (n)}}{\text{Total sample size (N)}}\right) = 1 - \frac{\mathrm{n}}{\mathrm{N}}$$

We call this **complement** of E (i.e., E')

Note 1:
P(E) + P(E') = 1

Note 2:
It is basic a prevalence or proportion value

Note 3:
Union (U) in set theory is also the total sample size

Joint Probabilities

A joint probability refers to the likelihood of more than one event occurring at the same time. For example, if there are two events i.e., E_1 and E_2, the joint probability is the chance that both events will occur at the same time.

- To simply put this – it is basically the probability of E_1 and E_2 when they happen at the same time. Best seen specifically in contingency tables.

- There are several notation for representing joint probabilities:

  $\Pr(E_1 \,\&\, E_2)$ is joint probability of events $\mathrm{E}_1 \,\&\, \mathrm{E}_2$; where E_1 **and** E_2 are two different type events that intersect or occur together

- <u>When there's data particularly from a contingency table</u> – joint probabilities are calculated as follows:

$$\Pr(\mathrm{E}_1 \,\&\, \mathrm{E}_2) = \frac{\text{Number of observed outcomes in both events}}{\text{Grand total in that sample space}}$$

Conditional Probabilities

A **conditional probability** of an event E_1 is the probability that E_1 will happen given that the event E_2 has already occurred.

- Simply put – it's the likelihood of an event E_1 occurring, based on the occurrence of a previous event E_2.
- We say: "Probability of 'this' given 'that'" – where the probability of event E_1 depends on the occurrence of event E_2
- The notation for representing conditional probabilities using this symbol "|" to represent 'given'. It is written as: $\Pr(E_1 \mid E_2)$ which means the probability of E_1 given E_2
- Conditional probabilities are computed accordingly as follows:

Conditional probabilities:

$$\Pr(\mathrm{E}_1 \mid \mathrm{E}_2) = \frac{\Pr(\mathrm{E}_1 \,\&\, \mathrm{E}_2)}{\Pr(\mathrm{E}_2)}$$

Notes: Calculating the conditional probabilities are quite involved. You will need to calculated the joint probabilities of E1 and E2, and unconditional probability of E2 and divide them together.

Example: Study on measuring abundance of Adult mosquitoes in Location A [1]

	Breeding sites: Aedes	Breeding sites: No Aedes	Sum (Row)
Urban	182	1008	1190
Rural	132	861	993
Sum (Column)	314	1869	Grand total: 2183

- We are interested detected presence of adult mosquitoes from breeding sites within an urban or rural setting

Let the B represent the event Breeding sites: Aedes
Let the B’ represent the event Breeding sites: No Aedes

Let the U represent the event urban area (i.e., total number of breeding sites found in an urban setting)
Let the U’ represent the event rural area (i.e., total number of breeding sites found in a rural setting)

Unconditional Probabilities

Joint Probabilities

Conditional Probabilities

Example: Study on measuring abundance of Adult mosquitoes in Location A [2]

	Breeding sites: Aedes	Breeding sites: No Aedes	Sum (Row)
Urban	182	1008	1190 (/2183)
Rural	132	861	993 (/2183)
Sum (Column)	314 (/2183)	1869 (/2183)	Grand total: 2183

Here, we can compute the probabilities by simply dividing the number of events observed by the overall total sample space which is 2183

The second table, we have simply converted the raw values to probabilities. The light blue shaded cells are the unconditional (or marginal) probabilities

	Breeding sites: Aedes (B)	Breeding sites: No Aedes (B')	Sum (Row)
Urban (U)			0.55 P(U)
Rural (U')			0.45 P(U')
Sum (Column)	0.14 P(B)	0.86 P(B')	Grand total: 1

Probability of finding breeding site in Urban location = 0.55

Probability of finding breeding site in Rural location = 0.45

Probability of finding breeding site with Aedes mosquito = 0.14

Probability of finding breeding site with no Aedes mosquito = 0.86

Example: Study on measuring abundance of Adult mosquitoes in Location A [2]

	Breeding sites: Aedes	Breeding sites: No Aedes	Sum (Row)
Urban	182	1008	1190 (/2183)
Rural	132	861	993 (/2183)
Sum (Column)	314 (/2183)	1869 (/2183)	Grand total: 2183

	Breeding sites: Aedes (B)	Breeding sites: No Aedes (B')	Sum (Row)
Urban (U)	B & U	B' & U	0.55 P(U)
Rural (U')	B & U'	B' & U'	0.45 P(U')
Sum (Column)	0.14 P(B)	0.86 P(B')	Grand total: 1

Example: Study on measuring abundance of Adult mosquitoes in Location A [3]

	Breeding sites: Aedes	Breeding sites: No Aedes	Sum (Row)
Urban	182 (/2183)	1008 (/2183)	1190
Rural	132 (/2183)	861(/2183)	993
Sum (Column)	314	1869	Grand total: 2183

In the second table, we can convert these raw values to joint probabilities. The dark blue shaded cells are the joints probabilities computed by using the formula shown in slide #20

	Breeding sites: Aedes (B)	Breeding sites: No Aedes (B')	Sum (Row)
Urban (U)	0.0834 Pr(B & U)	0.4617 Pr(B' & U)	0.55 Pr(U)
Rural (U')	0.0605 Pr(B & U')	0.3944 Pr(B' & U')	0.45 Pr(U')
Sum (Column)	0.14 Pr(B)	0.86 Pr(B')	Grand total: 1

Probability of finding a breeding site with Aedes mosquitoes in rural areas = 0.0605 (6.05%)

Probability of finding a breeding site with no Aedes mosquitoes in rural areas = 0.3944 (39.44%)

Example: Study on measuring abundance of Adult mosquitoes in Location A [3]

	Breeding sites: Aedes	Breeding sites: No Aedes	Sum (Row)
Urban	182	1008	1190
Rural	132	861	993
Sum (Column)	314	1869	Grand total: 2183

Suppose, we want to know what the probability that that Aedes mosquito are present in breeding sites given the setting is urban i.e., Pr(B|U), a conditional probability.

We will need the joint probability Pr(B & U)
We will need the unconditional probability for Pr(U)

	Breeding sites: Aedes (B)	Breeding sites: No Aedes (B')	Sum (Row)
Urban (U)	0.0834 Pr(B & U)	0.4617 Pr(B' & U)	0.55 Pr(U)
Rural (U')	0.0605 Pr(B & U')	0.3944 Pr(B' & U')	0.45 Pr(U')
Sum (Column)	0.14 Pr(B)	0.86 Pr(B')	Grand total: 1

Conditional probabilities:

$$\Pr(\mathrm{E}_1 \mid \mathrm{E}_2) = \frac{\Pr(\mathrm{E}_1 \;\&\; \mathrm{E}_2)}{\Pr(\mathrm{E}_2)}$$

$$\Pr(B|U) = \frac{\Pr(B \;\&\; U)}{\Pr(U)} = \frac{0.0834}{0.55} = 0.1529 = 15.29\%$$

Basic Probabilities | Probability Distribution | Bayes' Rule | Bayesian Inference

Probability Distributions

Definition:

Probability distribution is a mathematical function that estimates the **plausibility (or frequency)** of observing a particular value (or range of values) for a variable. It can be used to estimate **cumulative probability up to a certain value.**

There are broadly two groups for probability distributions with several subtypes:

Probability Mass Function (PMFs) (Discrete)

- Discrete uniform distribution
- Bernoulli distribution
- Binomial distribution*
- Poisson distribution

Note: These distributions can handle variables that counts, or a discrete (or distinct) in nature. <u>Directly estimates the probability</u>.

Probability Density Function (PDFs) (Continuous)

- Uniform distribution
- Gaussian (or Normal) distribution (flexible)*
- Gamma distribution (flexible)*
- Beta distribution (flexible)*
- Cauchy distribution (flexible)

Note: These distributions can handle variables that are continuous in nature. First estimates the plausibility (density).

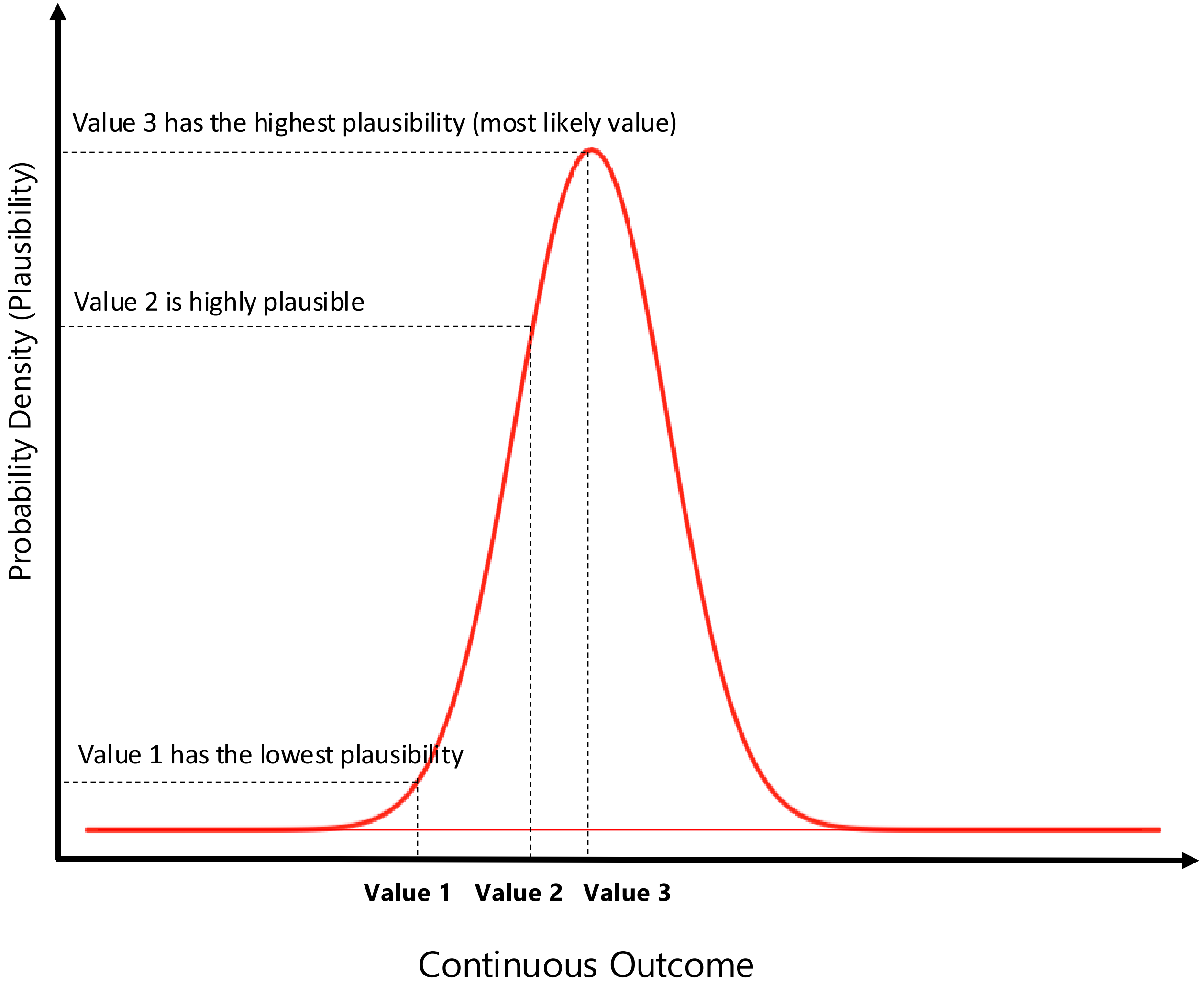
Important note:

In Bayesian inference, we will have to apply probability distributions to the various **conditional** and **marginal probability** components (i.e., **likelihood function** and **prior probability**, respectively) of the Bayes' rule formula to derive the **posterior probability**.

We impose a probability distribution to model an outcome (likelihood), while we subjectively choose another probability distribution (according to our believes aka priors) that best describes the shape of the parameters before observing the data

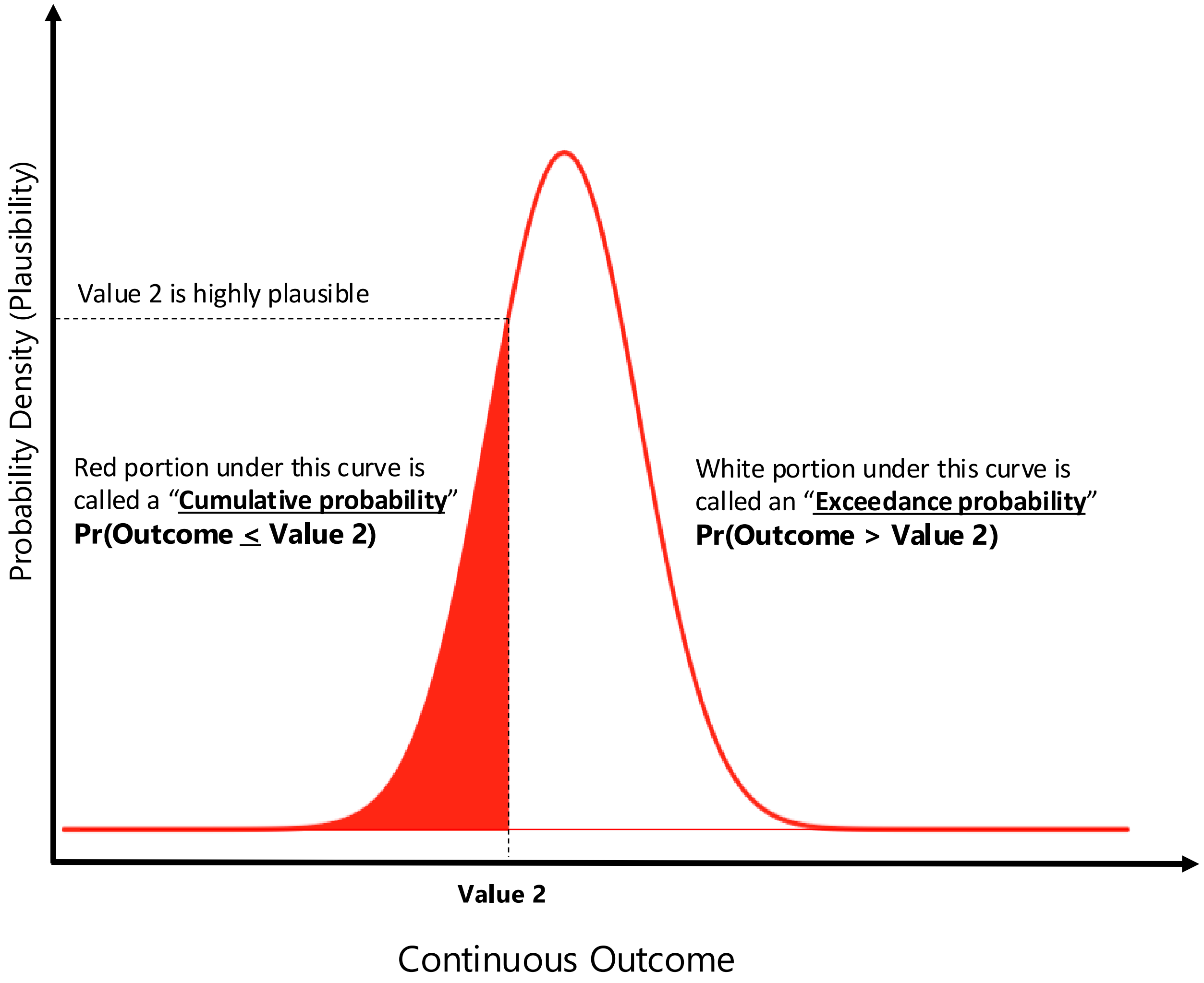
Understanding the features of probability distribution [1]

[1] Point values



This plot tell us the probability density (or plausibility) associated with the most (or least) likely or frequent

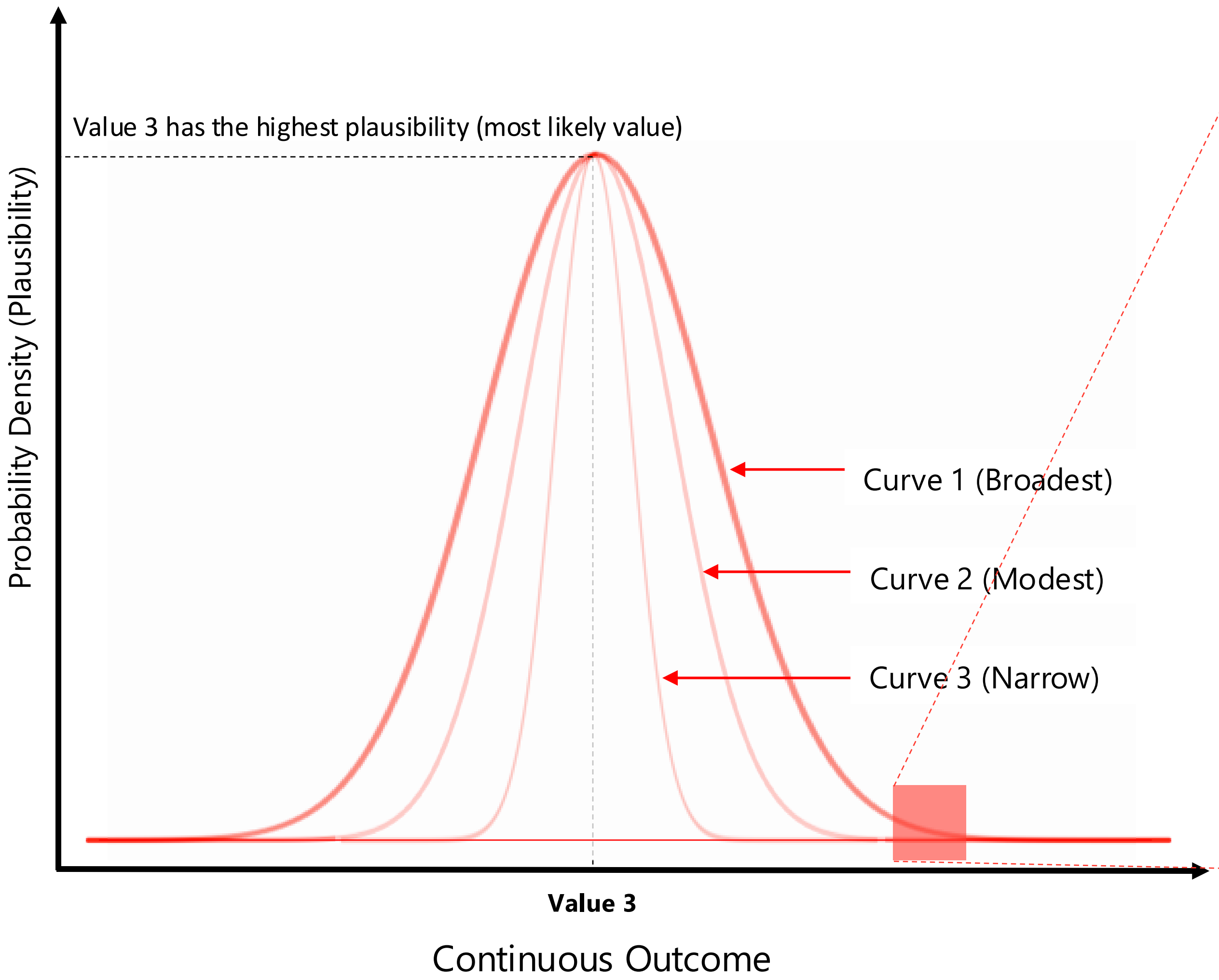
[2] Areas under a curve



The shaded area under the curve in this plot tells us the actual probability for observing a particular value i.e., **Pr(Outcome ≤ Value 2)**. From a distribution with a continuous outcome, if you want to calculate a probability – some shaded portion under this curve must be calculated.

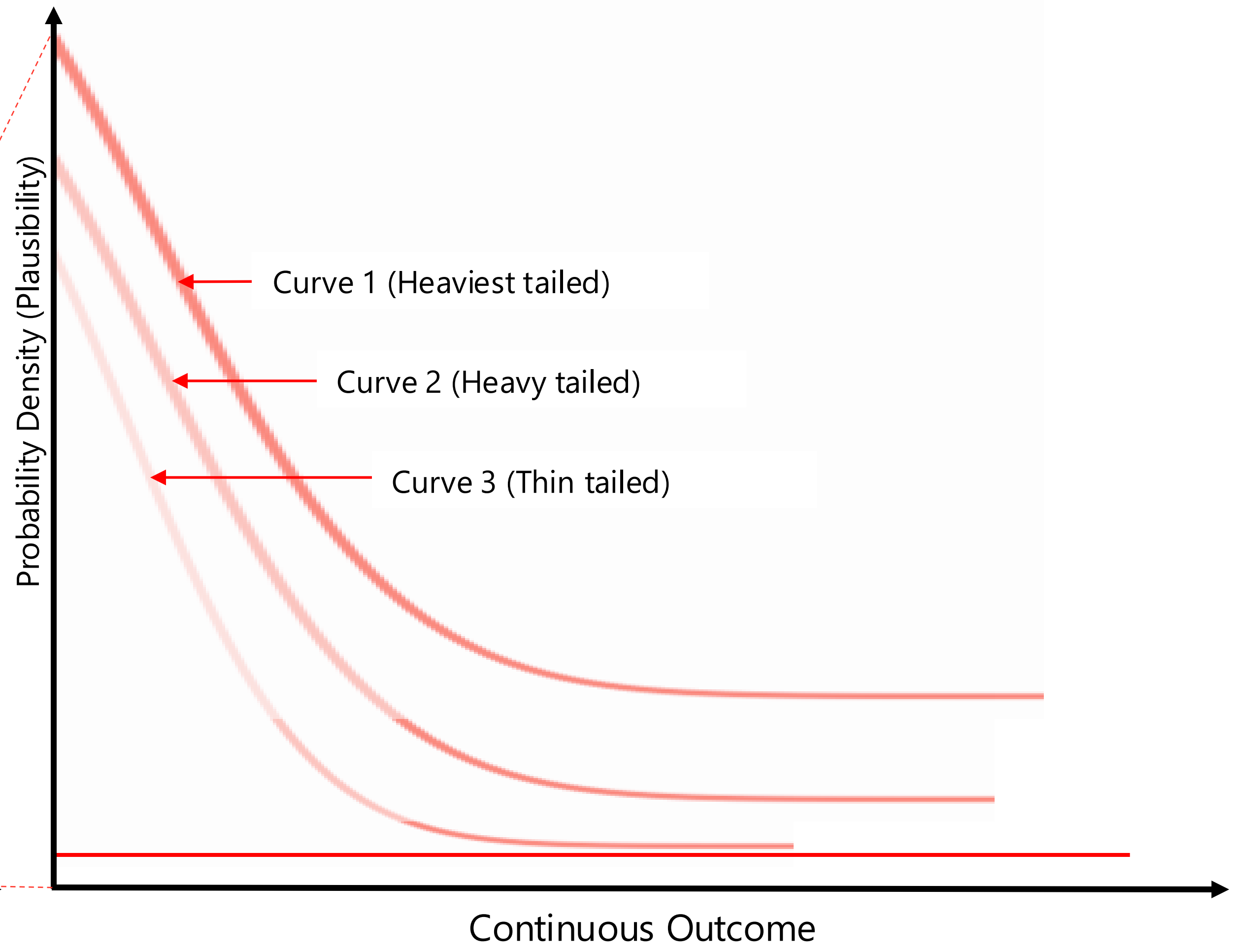
Understanding the features of probability distribution [2]

[3] Spread (or variation)



This plot tell us how widely the values in a distribution are dispersed, and so broader curves indicate more variation (e.g., curve 1), while narrower curves show less (e.g., curve 3). Or one can say the following: 'the wider the curve, the greater the variation, the narrower the curve, the more tightly values cluster around the centre'.

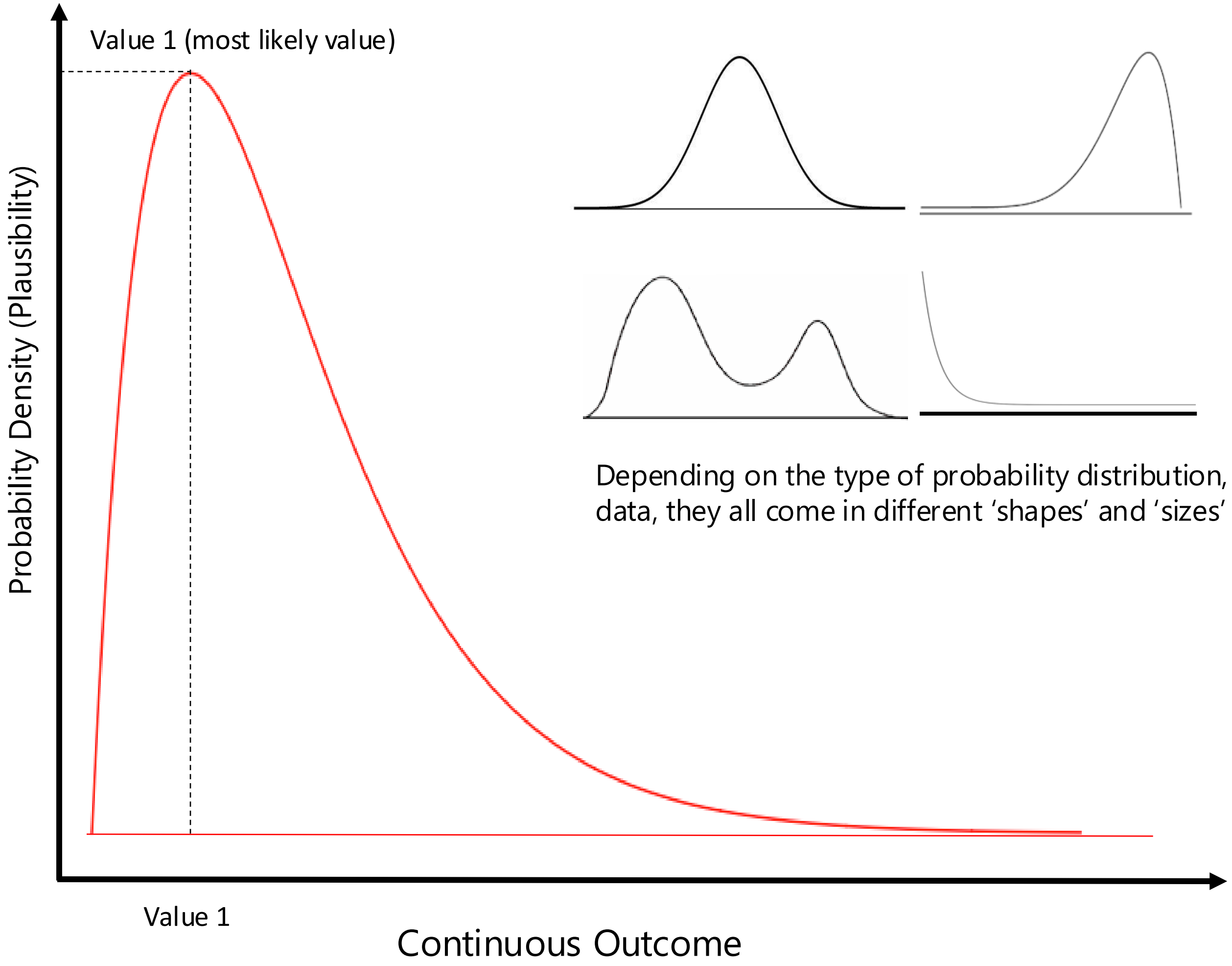
[4] Heavy-tailed



This plot tell us where a heavy tail exist in a distribution. A heavy-tailed distribution is one that gives more plausibility to extreme values. For instance, curve 3 has extreme values that highly unlikely because it's a thinned tail, while curve 1 – there's some plausibility (though small) in its extreme values

Understanding the features of probability distribution [3]

[5] Shape

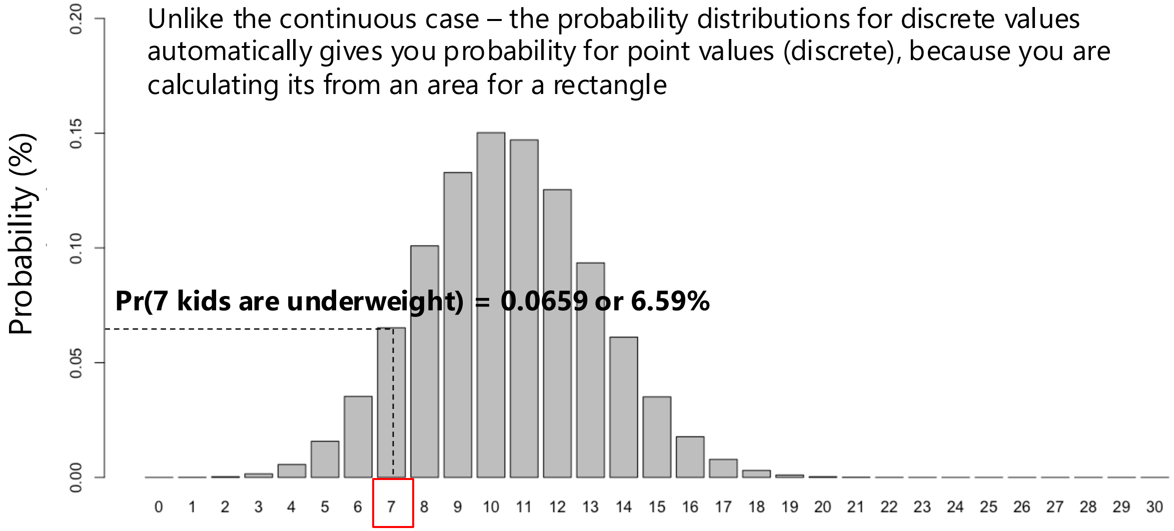


Depending on the type of probability distribution, or data, they all come in different 'shapes' and 'sizes'

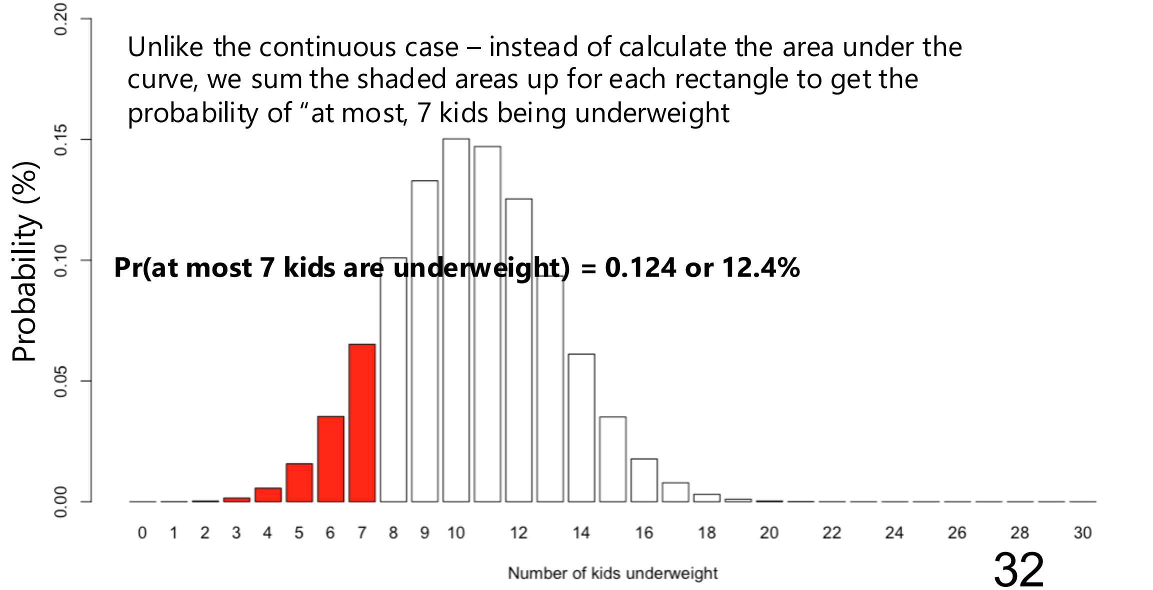
This plot tell us that probability distributions can take on many different shapes. For instance, some can be symmetric or asymmetric, some are skewed, and some have more than one peak. The shape, spread, and type of data offer important clues about which distribution a dataset may follow - main image is an example of a Gamma or Beta distribution.

[6] Discrete case (Unique)

Unlike the continuous case – the probability distributions for discrete values automatically gives you probability for point values (discrete), because you are calculating its from an area for a rectangle



Unlike the continuous case – instead of calculate the area under the curve, we sum the shaded areas up for each rectangle to get the probability of "at most, 7 kids being underweight



How do probability distributions work?

Gaussian (or Normal) Distribution [1]

What is it?
It characterised by a bell-shaped symmetrical curve for **continuous data** where most values around the mean (or average), and the farther you move from that average (either higher or lower on the spectrum of values) the fewer values you'll find.

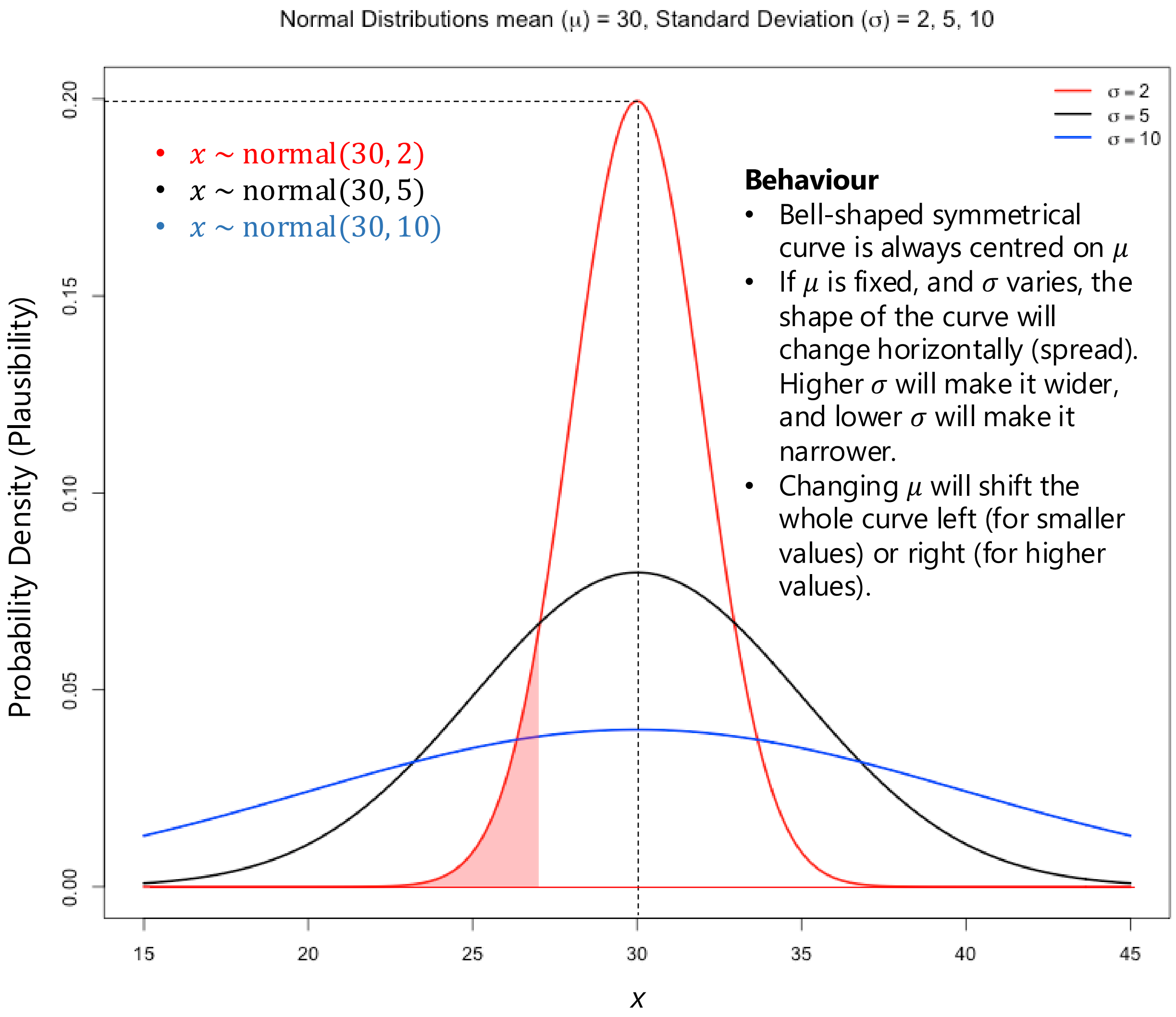
What do we need?
- ❖ Assumed mean (average), represented as μ (mu)
- ❖ Assumed standard deviation, represented as σ (sigma)

What is the formulation?

$$\Pr(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Where:
- ❖ Mean is μ (mu)
- ❖ Standard deviation is σ (sigma)
- ❖ π (pi) is 3.141593
- ❖ x is a data input that is a **continuous measure**
- ❖ Another way to write the above formula: $x \sim \mathrm{normal}(\mu, \sigma)$

What is it used for?
- ❖ To get a plausibility estimate for a measured value in that distribution (i.e. a probability density)
- ❖ To get a cumulative probability.
- ❖ Generating random data points where most values generated are closer to mean i.e., μ (mu) and spread according around it dictated by standard deviation σ (sigma)



Behaviour
- Bell-shaped symmetrical curve is always centred on μ
- If μ is fixed, and σ varies, the shape of the curve will change horizontally (spread). Higher σ will make it wider, and lower σ will make it narrower.
- Changing μ will shift the whole curve left (for smaller values) or right (for higher values).

Important note: Understanding behaviour of a probability distribution is a vital when doing Bayesian inference, particularly in the specification of priors (i.e., inputting our own subjective belief or knowledge before observing the data).

When use a normal distribution as a prior, and state the mean, you are saying that "*you believe that values are concentrated around the value of 30, where that 30 is the most likely*". Where you state the standard deviation (e.g., 2) in this distribution, you are saying that "*you believe that the range around 30 is between 28 and 32 (1-SD) (or alternatively approximately between 24 and 36 (3-SD))*".

How do probability distributions work?

Gaussian (or Normal) Distribution [2]

Example using the Normal Distribution for computing **plausibility** of an observed value, and **cumulative or exceedance probability**

Monitoring the environmental concentrations of Bti larvicide dunks in standing water to prevent mosquito breeding hotspots in Brazil (Source: CDC)

FIXED PARAMETER INPUTS

- μ_{BTI} = 25.0 mg/L (i.e., assumed mean concentration of Bti larvicide in the environment (most plausible value)
- σ_{BTI} = 5.0 mg/L (i.e., assumed standard deviation for concentrations in the environment)
- π = 3.141593

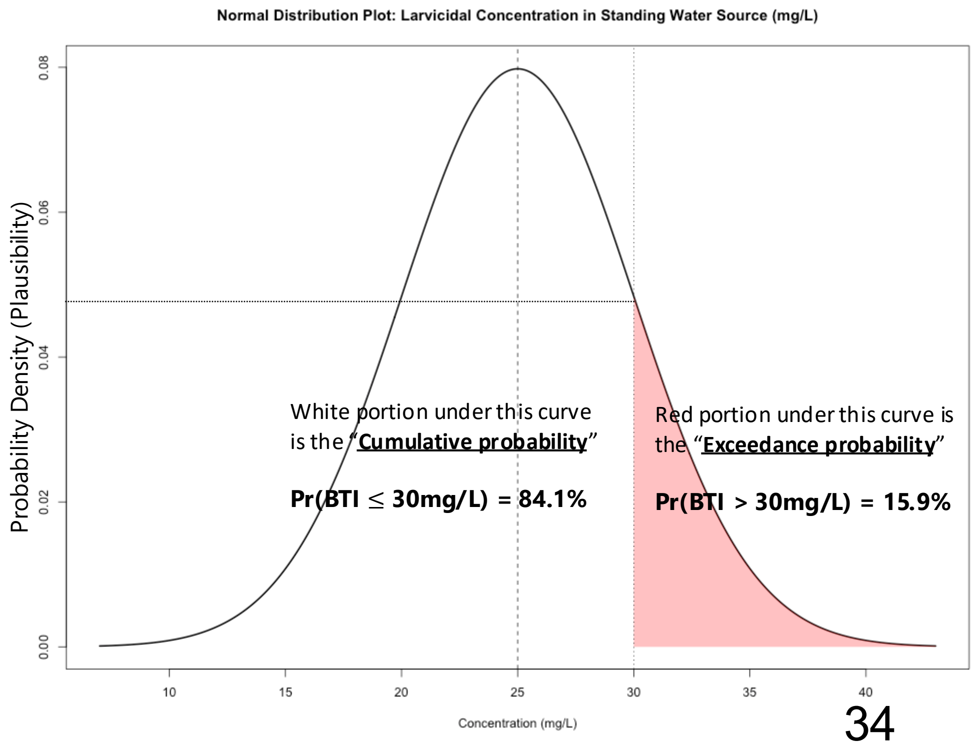
DATA INPUTS

BTI = concentrations ranging from 7.0 to 43.0 (at increments of 0.1, so 7.0, 7.1, 7.2… 42.8, 42.9 and 43.0)

Each value in this range is inserted into this formula to get a plausibility value.

$$\Pr(\mathrm{BTI}) = \frac{1}{5.0\sqrt{2(3.1415193)}} e^{-\frac{1}{2}\left(\frac{\mathrm{BTI}-25.0}{5.0}\right)^2}$$

It generates this curve



Interpretation: We can estimate how plausible it is for BTI levels to be exactly 30 mg/L. This value of 0.0483 represents the likelihood of BTI being 30 mg/L relative to all other BTI values in the distribution. Again, it NOT a probability nor percentage – it's a measure for plausibility.

The probability of BTI levels being up to 30 mg/L in the environment is 0.841 (84.1%). The probability that BTI levels are over 30 mg/L in the environment is 0.159 (15.9%).

How do probability distributions work?

Gamma Distribution [1]

What is it?
It characterised by a smooth curve for **continuous data are only positive values.** It is a distribution best for describing **positively right skewed** information. **You can control its shape and scale of the gamma distribution**.

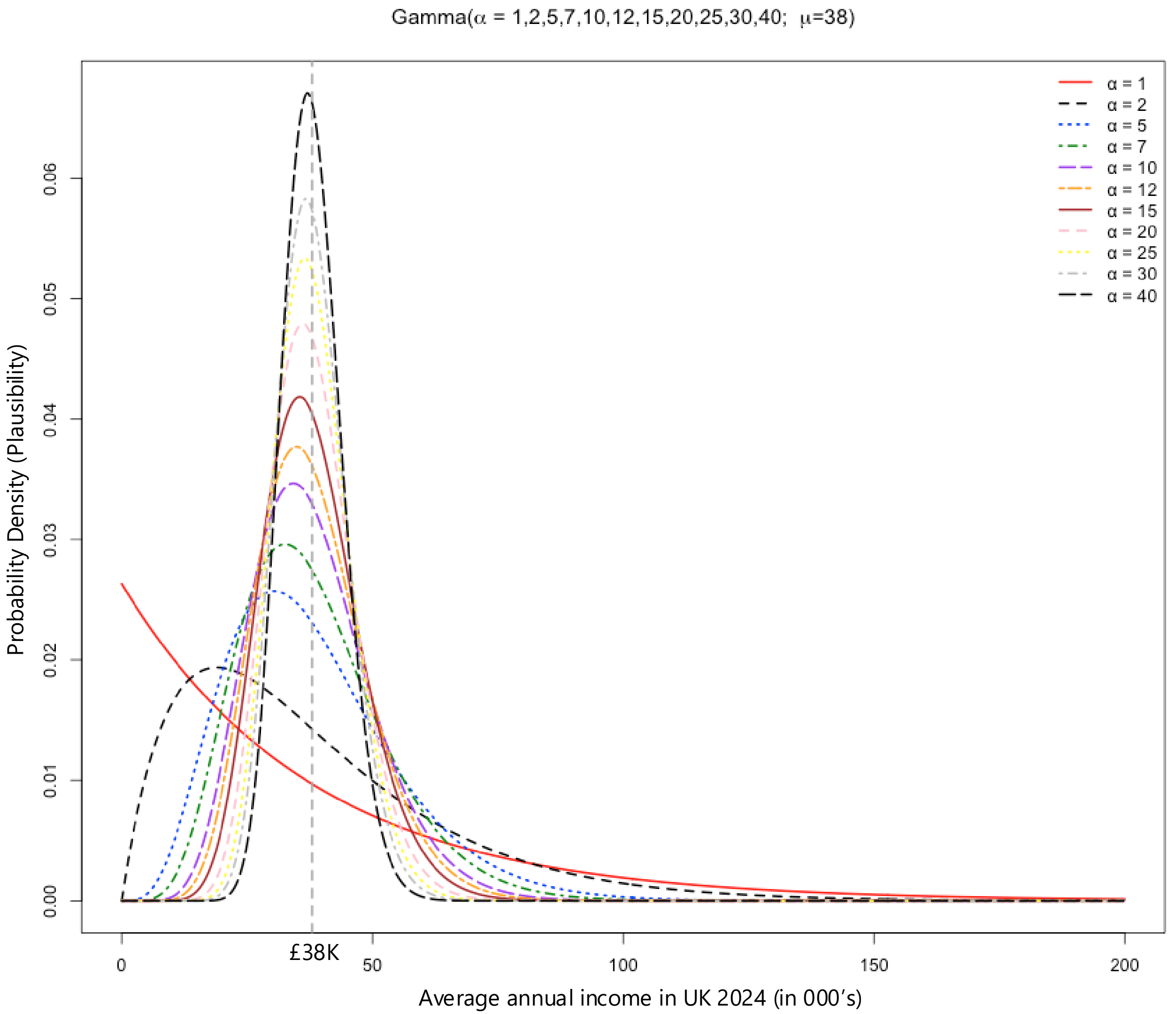
What do we need?
- ❖ Mean represented as μ (mu) in which the skewness is based on
- ❖ **The mean is used to get it centred closely to where the peak or hump is (mode). An appropriate shape value α must be provided. This controls the hump of the distribution in terms of 'where' it sits and its height.**
- ❖ The scale β controls the tail to show slower or faster decay. You can define the scale how you want, but if you have a mean – then it's calculated as: ${}^{\mu}/_{\alpha}$

What is the formulation?

$$\Pr(x) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}$$

Where:
- ❖ Shape is α
- ❖ Scale is β
- ❖ Denominator is a constant value – $\Gamma(\alpha) = (\alpha - 1)!$
- ❖ x is a data input that is a **continuous measure that is positive value**
- ❖ Another way to write the above formula: $x \sim \text{gamma}(\alpha, \beta)$

What is it used for?
- ❖ To get a plausibility estimate for a measured value in that distribution (i.e. a probability density) and to find where the most plausible value lies in that distribution which is represented as the mode (estimated as $(\alpha - 1)! \times \beta$).
- ❖ To get a cumulative probability.
- ❖ Generating random data points from a positively skewed distribution



Important note: Understanding behaviour of a probability distribution is a vital when doing Bayesian inference, particularly in the specification of priors (i.e., inputting our own subjective belief or knowledge before observing the data).

When use a gamma distribution as a prior – you will want to use knowledge of the process to define the shape parameter correctly. In Bayesian inference, we mainly use the Gamma distribution for this purpose. In this example, we want to generate a skewed distribution for average income (£38,000) as the mean. The appropriate shape value will be α was experimented for anything from 1to 40.

How do probability distributions work?

Gamma Distribution [2]

Example using the Gamma Distribution for computing **plausibility** of an observed value, and **cumulative or exceedance probability**

FIXED PARAMETER INPUTS

- $\mu_{DEET} = 20\mu g/m^3$ (i.e., assumed mean concentration of DEET in households). It's used for deriving the scale.
- The concentrations of DEET in households are positively skewed as a lot of households tend to have low-levels of DEET, where fewer households will have excess amounts of DEET. To capture this pattern, we set $\alpha = 2$ (shape) and use $\beta = 10$ (scale i.e., $^{\mu_{DEET}}/_{\alpha} = ^{20}/_{2}$)

DATA INPUTS

DEET = concentrations ranging from 0.0 to 200.0$\mu g/m^3$ (at increments of 0.1, so 0.0, 0.1, 0.2… 199.8, 199.9 and 200.0)

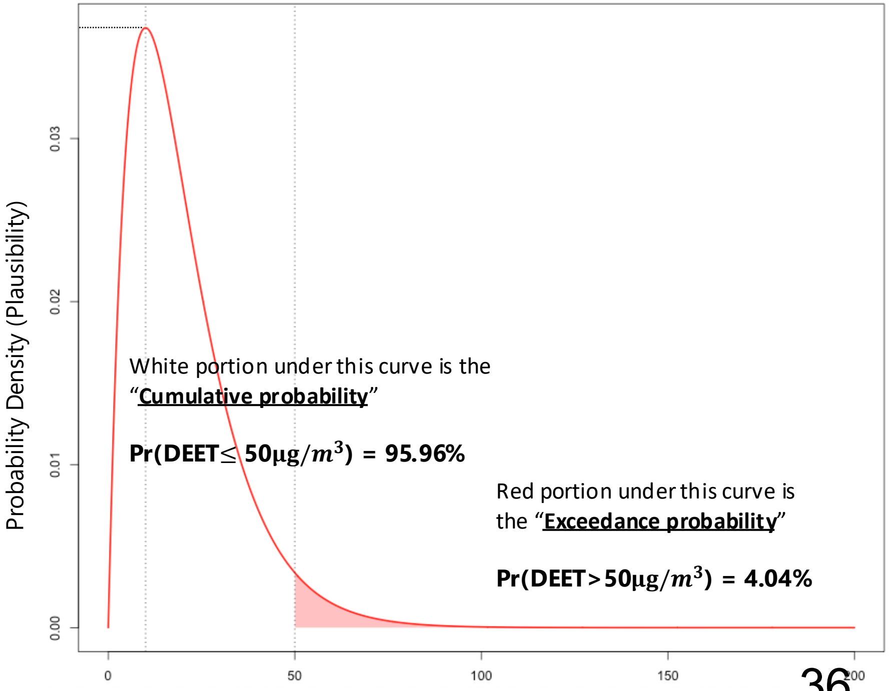
Each value in this range is inserted into this formula to get a plausibility value.

$$\Pr(\text{DEET}) = \frac{10^2 \times \text{DEET}^{2-1} \times e^{-10(\text{DEET})}}{(2-1)!}$$

It generates this curve

Monitoring the ambient chemical concentrations of anti-mosquito repellents (DEET) in households after application to outdoor airspace in Sao Paula, Brazil (Source: Wall Street Journal)

Gamma Distribution Plot: Ambient Anti-mosquito (DEET) Concentration in Households (mg/L)



Interpretation: We can estimate how plausible it is for DEET levels from the average of 20 $\mu g/m^3$. This corresponds to modal value of 10 $\mu g/m^3$ (i.e., plausibility is 0.03678) which represents the likelihood of DEET being 10 $\mu g/m^3$ relative to all other values in the distribution.

The probability of DEET levels not exceeding 50 $\mu g/m^3$ in households is 0.9596 (95.96%). The probability that DEET levels exceeds this threshold in households is 0.0404 (4.04%).

How do probability distributions work?

Beta Distribution [1]

What is it?
It characterised by a smooth curve for **continuous data that are positive values between 0 and 1, inclusively.** It is a distribution best for describing **proportions with unknown numerator and/or denominator**. **You can control its shape and location of the Beta distribution**.

What do we need?

- **Scenario 1**: You need to specify both α and β. When α is greater than β, the resulting distribution is that of a **negatively skewed distribution**. If you assume higher proportion values closer to 1 (i.e., 100%) to be more plausible – then let α be greater than β.
- **Scenario 2**: You need to specify both α and β. When β is greater than α, the resulting distribution is that of a positively skewed distribution. If you assume lower proportion values closer to 0 (i.e., 0%) to be more plausible – let β be greater than α.
- **Scenario 3:** You need to specify both α and β. When α is equal to β, and increasing both at the same time, the resulting distribution starts of as a uniform distribution and then converges to that which is akin to a normal distribution. Uniform – every value for the proportion is equally plausible; while as both parameters increase in the same way – we assume that the proportion of 50% became more and more plausible.
- **Scenario 4:** When you have a specific value for the proportion which you want the Beta distribution to be centred on – you will have to solve for both α and β! Meaning that you will need to specify an assumed mean value (μ) for the proportion and a small value for the standard deviation (σ).

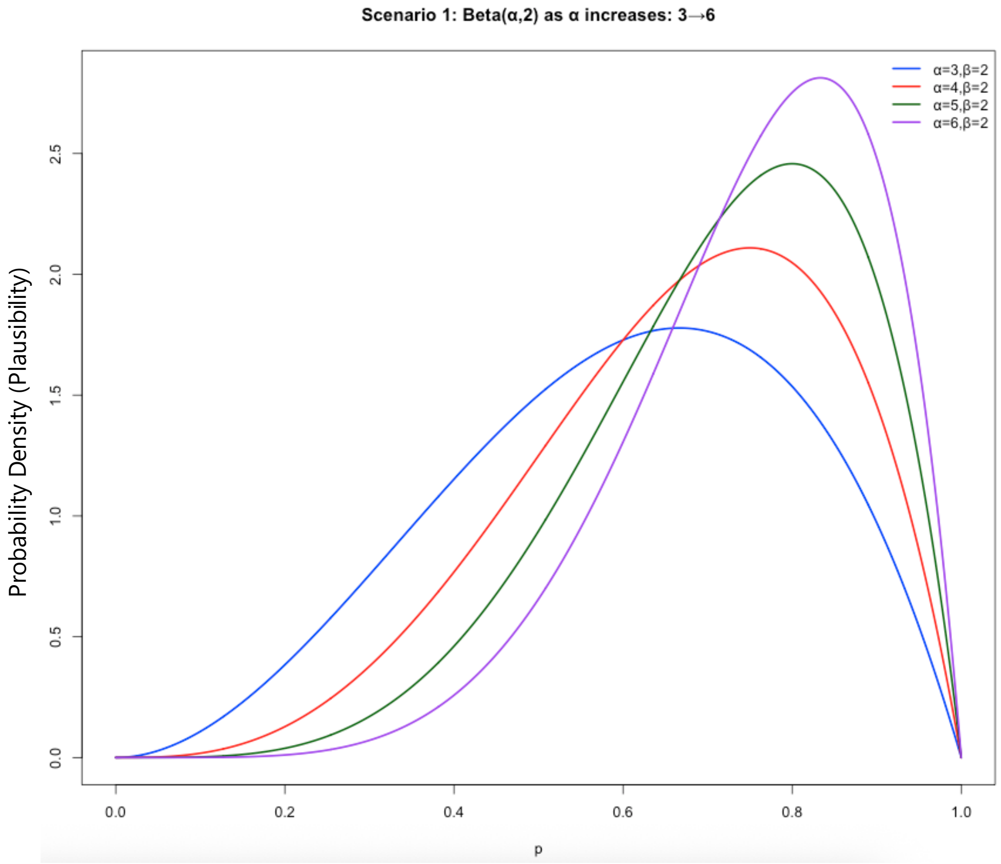
$$\beta = \mu - 1 + \left[\frac{\mu \times (1-\mu)^2}{\sigma^2}\right]$$

$$\alpha = \frac{\beta \times \mu}{(1-\mu)}$$

What is the formulation?

$$\Pr(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

Where:
- Shape is α
- Shape is β
- Denominator is a constant value – $\mathrm{B}(\alpha,\beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$
- p is a data input that is a **continuous measure that lies between 0 and 1, inclusively – it is a proportion (percentage).**
- Another way to write the above formula: $p \sim \mathrm{beta}(\alpha,\beta)$



How do probability distributions work?

Beta Distribution [2]

What is it?

It characterised by a smooth curve for **continuous data that are positive values between 0 and 1, inclusively.** It is a distribution best for describing **proportions with unknown numerator and/or denominator**. **You can control its shape and location of the Beta distribution**.

What do we need?

- **Scenario 1**: You need to specify both α and β. When α is greater than β, the resulting distribution is that of a negatively skewed distribution. If you assume higher proportion values closer to 1 (i.e., 100%) to be more plausible – let α be greater than β.
- **Scenario 2**: You need to specify both α and β. When β is greater than α, the resulting distribution is that of a **positively skewed distribution**. If you assume lower proportion values closer to 0 (i.e., 0%) to be more plausible – then let β be greater than α.
- **Scenario 3:** You need to specify both α and β. When α is equal to β, and increasing both at the same time, the resulting distribution starts of as a uniform distribution and then converges to that which is akin to a normal distribution. Uniform – every value for the proportion is equally plausible, while as both parameters increase in the same way, we assume that the proportion of 50% became more and more plausible.
- **Scenario 4:** When you have a specific value for the proportion which you want the Beta distribution to be centred on – you will have to solve for both α and β! Meaning that you will need to specify an assumed mean value (μ) for the proportion and a small value for the standard deviation (σ).

$$\beta = \mu - 1 + \left[\frac{\mu \times (1-\mu)^2}{\sigma^2}\right]$$

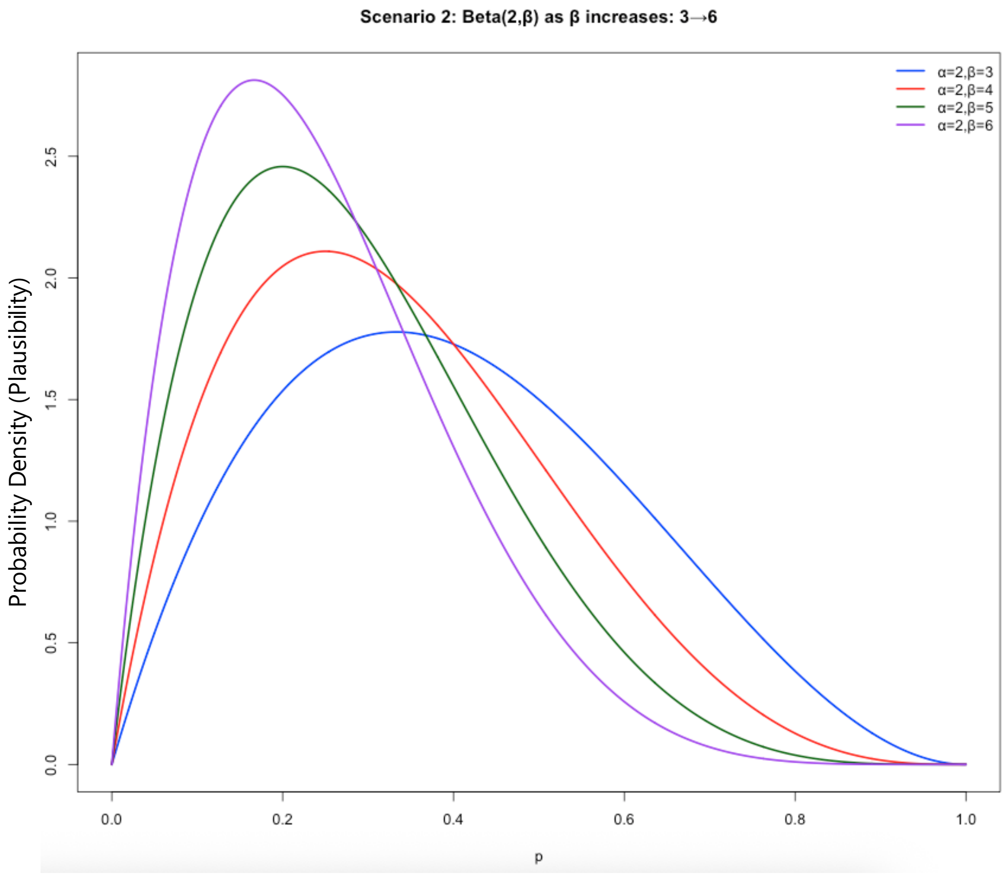
$$\alpha = \frac{\beta \times \mu}{(1-\mu)}$$

What is the formulation?

$$\Pr(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

Where:

- Shape is α
- Shape is β
- Denominator is a constant value – $\mathrm{B}(\alpha,\beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$
- p is a data input that is a **continuous measure that lies between 0 and 1, inclusively – it is a proportion (percentage).**
- Another way to write the above formula: $p \sim \mathrm{beta}(\alpha,\beta)$



How do probability distributions work?

Beta Distribution [3]

What is it?

It characterised by a smooth curve for **continuous data that are positive values between 0 and 1, inclusively.** It is a distribution best for describing **proportions with unknown numerator and/or denominator**. **You can control its shape and location of the Beta distribution**.

What do we need?

- **Scenario 1**: You need to specify both α and β. When α is greater than β, the resulting distribution is that of a negatively skewed distribution. If you assume higher proportion values closer to 1 (i.e., 100%) to be more plausible – let α be greater than β.
- **Scenario 2**: You need to specify both α and β. When β is greater than α, the resulting distribution is that of a positively skewed distribution. If you assume lower proportion values closer to 0 (i.e., 0%) to be more plausible – then let β be greater than α.
- **Scenario 3:** You need to specify both α and β. When α is equal to β, and increasing both at the same time, **the resulting distribution starts of as a uniform distribution and then converges to that which is akin to a normal distribution**. Uniform – every value for the proportion is equally plausible, while as both parameters increase in the same way, we assume that the proportion of 50% became more and more plausible.
- **Scenario 4:** When you have a specific value for the proportion which you want the Beta distribution to be centred on – you will have to solve for both α and β! Meaning that you will need to specify an assumed mean value (μ) for the proportion and a small value for the standard deviation (σ).

$$\beta = \mu - 1 + \left[\frac{\mu \times (1-\mu)^2}{\sigma^2}\right]$$

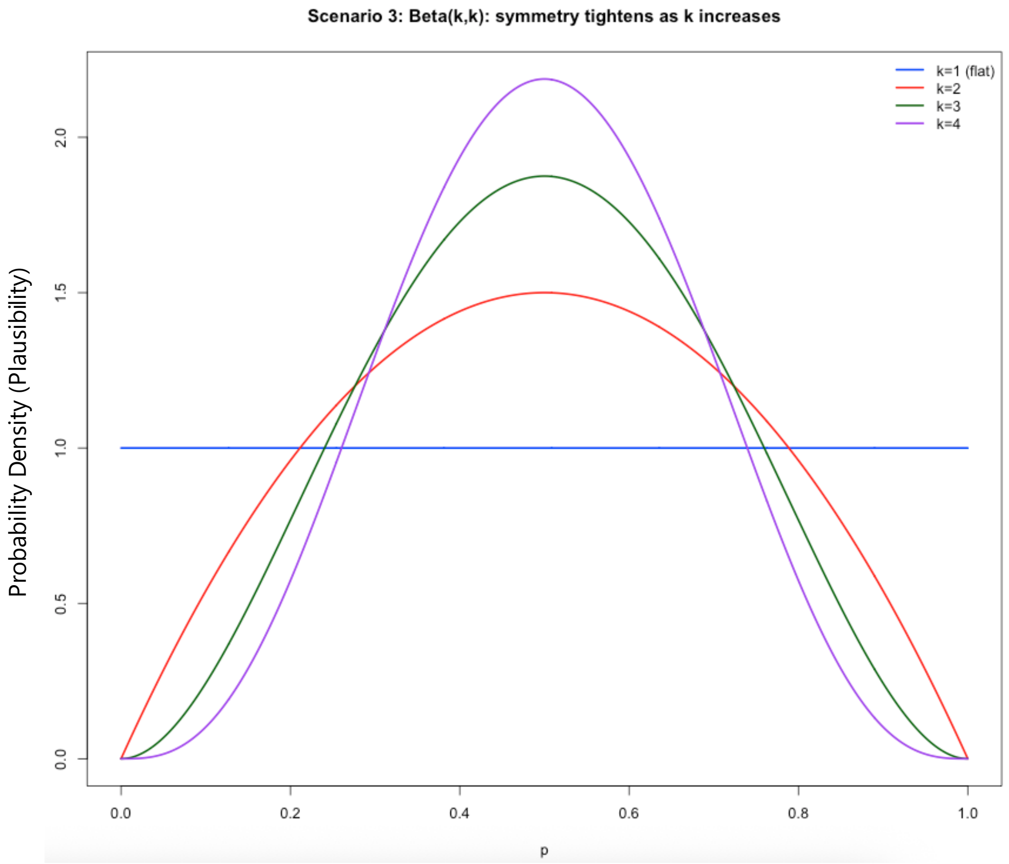
$$\alpha = \frac{\beta \times \mu}{(1-\mu)}$$

What is the formulation?

$$\Pr(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

Where:

- Shape is α
- Shape is β
- Denominator is a constant value – $\mathrm{B}(\alpha,\beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$
- p is a data input that is a **continuous measure that lies between 0 and 1, inclusively – it is a proportion (percentage).**
- Another way to write the above formula: $p \sim \mathrm{beta}(\alpha,\beta)$



How do probability distributions work?

Beta Distribution [4]

What is it?

It characterised by a smooth curve for **continuous data that are positive values between 0 and 1, inclusively.** It is a distribution best for describing **proportions with unknown numerator and/or denominator**. **You can control its shape and location of the Beta distribution**.

What do we need?

- **Scenario 1**: You need to specify both α and β. When α is greater than β, the resulting distribution is that of a negatively skewed distribution. If you assume higher proportion values closer to 1 (i.e., 100%) to be more plausible – let α be greater than β.

- **Scenario 2**: You need to specify both α and β. When β is greater than α, the resulting distribution is that of a positively skewed distribution. If you assume lower proportion values closer to 0 (i.e., 0%) to be more plausible – then let β be greater than α.

- **Scenario 3:** You need to specify both α and β. When α is equal to β, and increasing both at the same time, **the resulting distribution starts of as a uniform distribution and then converges to that which is akin to a normal distribution**. Uniform – every value for the proportion is equally plausible, while as both parameters increase in the same way, we assume that the proportion of 50% became more and more plausible.

- **Scenario 4:** When you have a specific value for the proportion which you want the Beta distribution to be centred on – you will have to solve for both α and β! Meaning that you will need to specify an assumed mean value (μ) for the proportion and a small value for the standard deviation (σ).

$$\beta = \mu - 1 + \left[\frac{\mu \times (1-\mu)^2}{\sigma^2}\right]$$

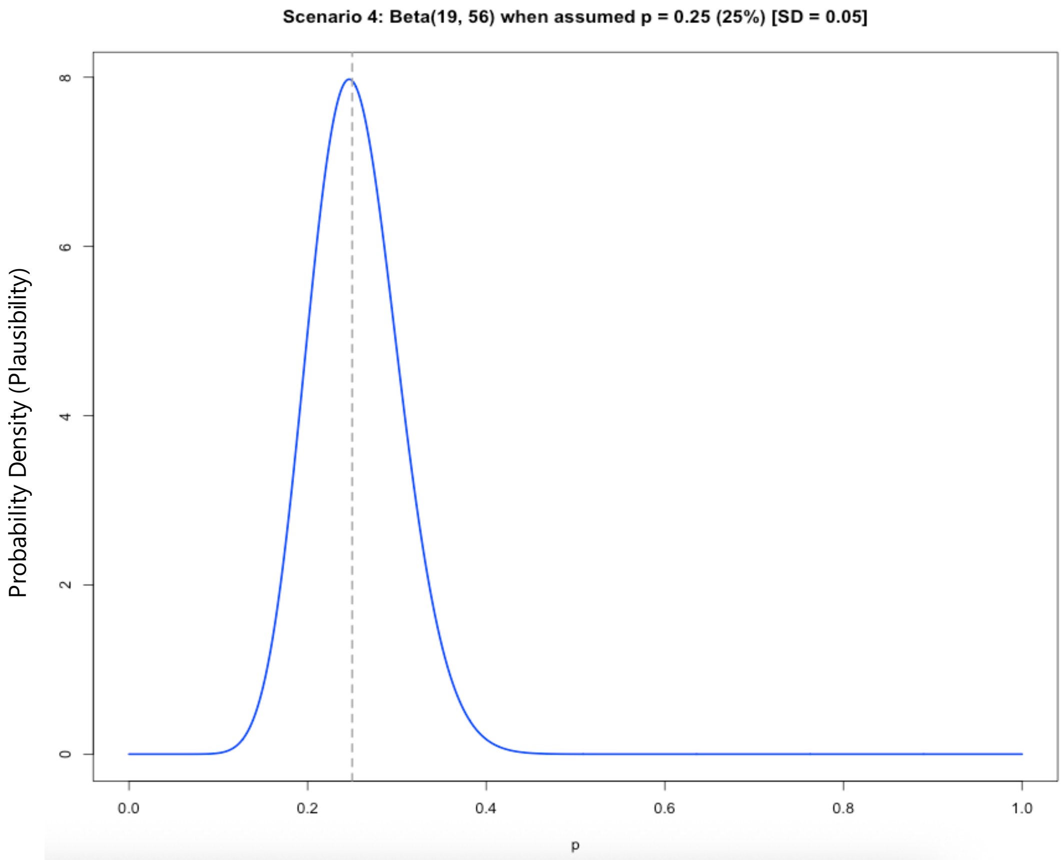
$$\alpha = \frac{\beta \times \mu}{(1-\mu)}$$

What is the formulation?

$$\Pr(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

Where:

- Shape is α
- Shape is β
- Denominator is a constant value – $\mathrm{B}(\alpha,\beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$
- p is a data input that is a **continuous measure that lies between 0 and 1, inclusively – it is a proportion (percentage).**
- Another way to write the above formula: $p \sim \mathrm{beta}(\alpha,\beta)$



How do probability distributions work?

Binomial Distribution [1]

What is it?

It characterised as a histogram or barplot for **discrete data that are positive values.** It is a distribution best for describing **proportions with known values for the numerator and denominator where the former represents successes, and the latter is a total.**

What do we need?

- We need to specify the assumed proportion (or percentage) about the outcome, represented as p
- We need to know the observed total – this is drawn from the dataset once the outcome has been observed. This is represented as n
- We need to know the observed number of successes y (e.g., people diseased, damage items from a cargo etc.,) from the total n

What is the formulation?

$$\Pr(y) = \binom{n}{y} p^{y}(1-p)^{n-y}$$

Where:

- The assumed proportion p
- The total is represented as n
- The number of successes y (from the total of n)
- Combinatorial notation which computes the total number of different ways (combinations) of observing the number of successes y (from the total of n): $\binom{n}{y} = \frac{n!}{y!(n-y)!}$
- Another way to write the above formula: $y \sim \mathrm{binomial}(n, p)$

What is it used for?

- To directly get a probability estimate for specific value of the discrete outcome. No plausibility here!
- To get a cumulative mass probability.
- Generating random data points that represents the number of successes from a binomial distribution

FIXED INPUTS

Before survey

- $p = 0.35$ (i.e., prior knowledge of prevalence household infestation in a particular neighbourhood)

After survey

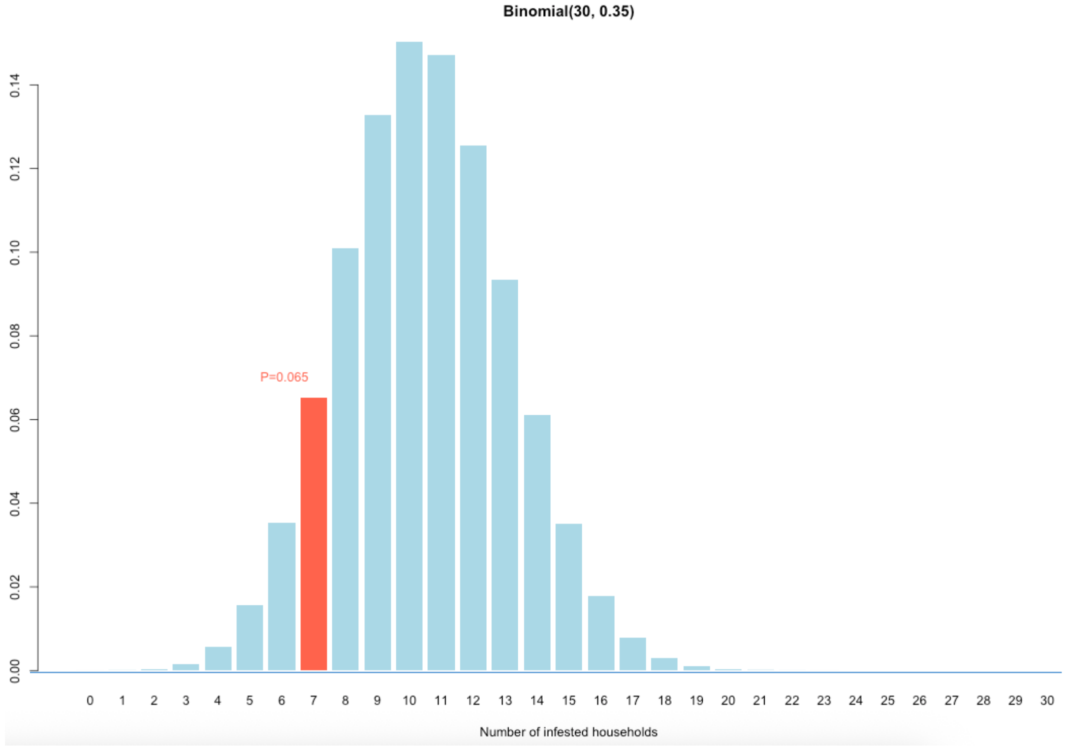
- $\mathrm{infested}\ (y) =$ There are 7 households infested
- $\mathrm{total}(n) =$ There are 30 households in total surveyed

INPUTS

$\mathrm{infested}\ (y) =$ 0 to 30 (maximum) (i.e., 0, 1, 2, 3, …, 27, 28, 29 and 30)

Each value in this range is inserted into this formula to get a probability estimate.

$$\Pr(\mathrm{infested}) = \binom{30}{\mathrm{infested}} 0.35^{\mathrm{infested}}(1-0.35)^{30-\mathrm{infested}}$$



The probability of 7 households being infested with mosquitoes is 0.0652 (6.52%).

How do probability distributions work?

Binomial Distribution [2]

What is it?

It characterised as a histogram or barplot for **discrete data that are positive values.** It is a distribution best for describing **proportions with known values for the numerator and denominator where the former represents successes, and the latter is a total.**

What do we need?

- ❖ We need to specify the assumed proportion (or percentage) about the outcome, represented as p
- ❖ We need to know the observed total – this is drawn from the dataset once the outcome has been observed. This is represented as n
- ❖ We need to know the observed number of successes y (e.g., people diseased, damage items from a cargo etc.,) from the total n

What is the formulation?

$$\Pr(y) = \binom{n}{y} p^{y}(1-p)^{n-y}$$

Where:

- ❖ The assumed proportion p
- ❖ The total is represented as n
- ❖ The number of successes y (from the total of n)
- ❖ Combinatorial notation which computes the total number of different ways (combinations) of observing the number of successes y (from the total of n): $\binom{n}{y} = \frac{n!}{y!(n-y)!}$
- ❖ Another way to write the above formula: $y \sim \mathrm{binomial}(n, p)$

What is it used for?

- ❖ To directly get a probability estimate for specific value of the discrete outcome. No plausibility here!
- ❖ To get a cumulative mass probability.
- ❖ Generating random data points that represents the number of successes from a binomial distribution

FIXED INPUTS

Before survey

- $p = 0.35$ (i.e., prior knowledge of prevalence household infestation in a particular neighbourhood)

After survey

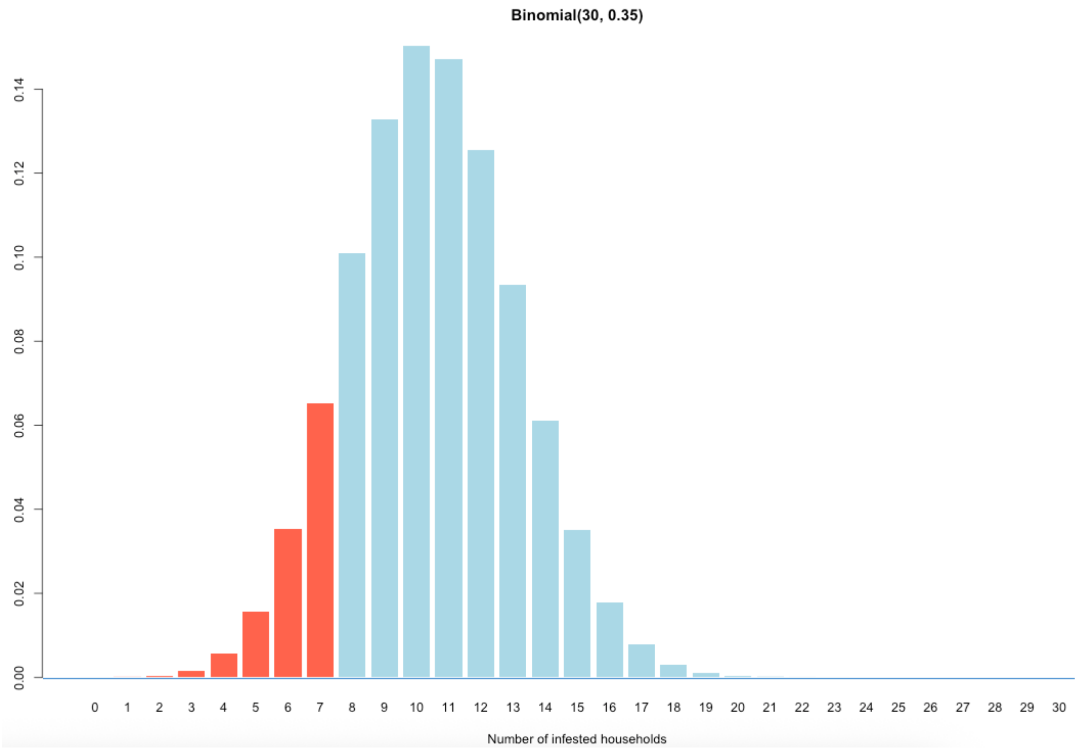
- $\mathrm{infested}\ (y) =$ There are 7 households infested
- $\mathrm{total}(n) =$ There are 30 households in total surveyed

INPUTS

$\mathrm{infested}\ (y) =$ 0 to 30 (maximum) (i.e., 0, 1, 2, 3, ..., 27, 28, 29 and 30)

Each value in this range is inserted into this formula to get a probability estimate.

$$\Pr(\mathrm{infested}) = \binom{30}{\mathrm{infested}} 0.35^{\mathrm{infested}}(1-0.35)^{30-\mathrm{infested}}$$



The cumulative mass probability of (at most) 7 households being infested with mosquitoes is 0.1237 (12.37%).

Summary of the common distributions

Characteristic	Function	Distribution	Statistical Notation	Usage***
• Continuous measure (scale, interval)	PDF	Gaussian	normal(mu, sigma)	• Set it as a likelihood for statistical model • Can be used as a prior
• Proportion (with information on the numerator and denominator)	PMF	Binomial	binomial(n, p)	• Set it as a likelihood for statistical model (only)
• Counts	PMF	Poisson	poisson(rate)	• Set it as a likelihood for statistical model (only)
• Binary measure (Yes, No)	PMF	Bernoulli	bernoulli(p)	• Set it as a likelihood for statistical model (only)
• Continuous measure (scale, interval) • Proportion (with information on the numerator and denominator) • Counts	PDF/PMF	Uniform	uniform(a, b)	• Use ONLY as a prior • Use it if you don’t have any prior information • This is an uninformative prior
• Continuous measure (scale, interval) • Counts	PDF/PMF	Gamma	gamma(alpha, beta)	• Use ONLY as a prior • It’s flexible
• Proportion (with information on the numerator and denominator)	PDF	Beta	beta(alpha, beta)	• Use ONLY as a prior • It’s flexible

Basic Probabilities | Probability Distribution | Bayes' Rule | Bayesian Inference

A Brief on Bayes' Theorem

Definition:

Bayes’ theorem (or law) determines the posterior probability of the parameters we estimate given the observed dataset.

- It’s a mathematical formula comprising of conditional and marginal probabilities
- It expresses how a subjective degree of belief should rationally change to account for new evidence.
- The basic formulation for Bayes’ Theorem (or Law):

Plausibility of the observed data we have, given the parameter has a specific value(s)

Our belief about what the parameter is, before observing the data

Probability of our new parameter being true after we’ve observed dataset

$$\boxed{\Pr(\text{Parameter} \mid \text{Data})} = \frac{\boxed{\Pr(\text{Data} \mid \text{Parameter})}\,\boxed{\Pr(\text{Parameter})}}{\boxed{\Pr(\text{Data})}}$$

This corresponds to something out the data that’s been observed

Thomas Bayes (1701 – 1761)

Posterior probability

- This is the result we get

Probability of our new parameter being true after we've observed dataset

Likelihood probability

- Define our statistical model on

Plausibility of the observed data we have, given the parameter has a specific value(s)

Prior probability

- Uninformative
- Weakly informative
- Informative

Our belief about what the parameter is, before observing the data

$$\boxed{\Pr(\text{Parameter} \mid \text{Data})} = \frac{\boxed{\Pr(\text{Data} \mid \text{Parameter})}\,\boxed{\Pr(\text{Parameter})}}{\boxed{\Pr(\text{Data})}}$$

This corresponds to something out the data that's been observed

Model prior probability linked with the observed dataset

- We don't really care about the part of the formula. Since it is not a function of the parameters we intend to predict.
- As far as I am concerned – it serves as constant value. Therefore, it can be removed from this equation.
- Once it is removed, the left-hand side of the equation is no longer "equal to" to the right-hand side of this equation. Its now proportional to each other.
- Also, admittedly – it makes the math of multiplying just the prior and likelihood a lot easier!

This is what is derived:

Posterior probability ∝ Likelihood x Prior Probability

This equation is the backbone of Bayesian Statistics

This equation is the backbone of Bayesian Statistics

❖ User friendly formulation:

Posterior probability ∝ Likelihood Function x Prior Probability

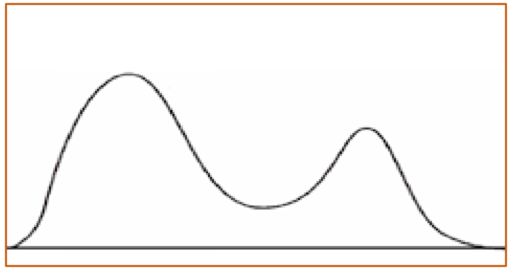
❖ Mathematical formulation:

Posterior: Probability of our new parameter being true **after** observing the data.

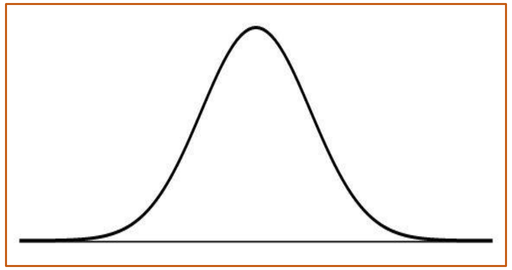
Likelihood: Plausibility of the observed data we have, given the parameter has a specific value(s)

Prior: Our belief about what the parameter is, <u>before</u> observing the data

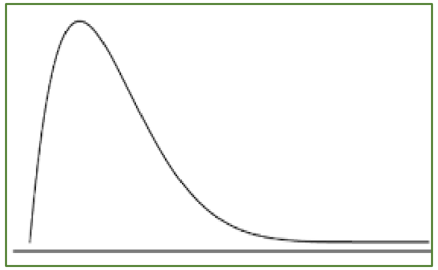
$$\Pr(\text{Parameter} \mid \text{Data}) \propto \Pr(\text{Data} \mid \text{Parameter}) \times \Pr(\text{Parameter})$$



$\propto$



$\times$



Notes:

- This equation comprise of conditionals, and marginal probability notation. So, an understanding of what **basic probabilities** are helps a lot.
- These marginal or conditional probabilities are expressed as **distributions**. In essence, you are defining, transforming and manipulating **probability distributions** to express and update uncertainty.
- This formulation is derived from the **Bayes' Rule**

❖ The usual mathematical notation for this is: $P(\theta|Y) \propto P(Y|\theta) \times P(\theta)$

Example

Goal: We need to estimate the coefficients and standard deviation for the regression model using a Bayesian framework.

Stan code

```
data {
  int<lower=0> N;         // sample size N
  int<lower=0> k;         // number of variables 3
  matrix[N, k] X;         // matrix: independent variables
  vector[N] y;            // vector/array for outcome
}

parameters {
  real beta0;              // Intercept
  vector[k] beta;          // beta coefficients
  real<lower=0> sigma;     // standard deviation
}

transformed parameters {
   vector[N] mu;
   mu = beta0 + X*beta;
}

model {
  beta0 ~ normal(0, 20);    // Prior for beta0
  beta ~ normal(0, 5);      // Prior for beta1 & beta2
  sigma ~ cauchy(0, 2.5);   // Prior for sigma
  y ~ normal(mu, sigma);    // Likelihood function
}
```

Model formulation

- Simple GLM (Linear case)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

- Specify likelihood function. The outcome is continuous – thus it normal distribution.

$$y \sim \mathbf{norm}(\boldsymbol{\mu}, \boldsymbol{\sigma})$$
$$\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

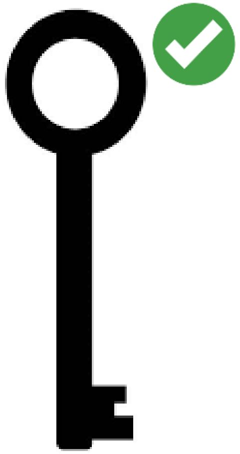
- Define the priors for the intercept, coefficients and other parameters, e.g., standard deviation

$$\beta_0 \sim \mathrm{Norm}(0, 20)$$
$$\beta_1 \sim \mathrm{Norm}(0, 5)$$
$$\beta_2 \sim \mathrm{Norm}(0, 5)$$
$$\sigma \sim \mathrm{cauchy}(0, 2.5)$$

- Build Bayesian model

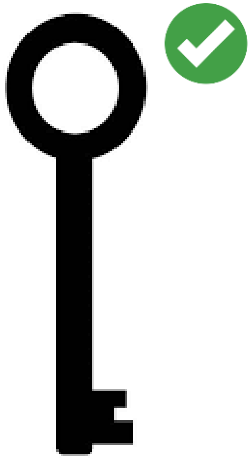
Recall the Bayes' Rule: $P(\theta|Y) \propto P(Y|\theta)P(\theta)$

$$\underbrace{\Pr(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma|\mu)}_{\text{Posterior}} \propto \underbrace{\Pr(\mu|\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \sigma)}_{\text{Likelihood}} \underbrace{\Pr(\boldsymbol{\beta}_0) Pr(\boldsymbol{\beta}_1) Pr(\boldsymbol{\beta}_2) Pr(\sigma)}_{\text{Priors}}$$



Basic Probabilities

Probability Distribution



Bayes’ Rule

Bayesian Inference

Any questions?