GEOG0125
ADVANCED TOPICS IN SOCIAL AND GEOGRAPHIC DATA SCIENCE

INTRODUCTION TO BAYESIAN STATISTICS

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

About the course

Description of the GEOG0125 Course

- This module will introduce students to advanced topics and methods that are essential for social and geographic data science, i.e., advanced or specialised techniques e.g., applied **Bayesian Statistics and Machine Learning** applied to this domain.

- This module covers two areas:
  - How to use **Bayesian Statistics** for both evidence-based research in making spatial predictions as well as studying the associations between social-risk factors and outcomes
  - How to use **Machine Learning** for various image processing and network sciences for tackling future social geographical challenges

Meet the Team

Module tutors

Anwar Musah
Lecturer in Social & Geographic Data Science

Stephen Law
Lecturer in Social & Geographic Data Science

Seminar tutors

Thomas Keel
Post Graduate Teaching Assistant
PhD in UCL Geography

General format of the course

- **1-hour lecture**
  - Weekly sessions every Tuesday 10:00am – 11:00am
  - Delivered by the module tutors
- **2-hour computer practical seminar**
  - Followed immediately after lectures every Friday 11:00am – 01:00pm
  - These sessions are facilitated by both the module and seminar tutors

Location(s):

Lectures:	10:00am-11:00am North-West Wing Building Room G07
Computer practicals:	11:00am-01:00pm North-West Wing Building Room G07

Please bring your own laptops to the computer practicals as NWW G07 is not a PC Cluster room

NOTE: All lectures and tutorials sessions are compulsory & will be delivered in-person.

Module Outline & Assessment

Bayesian Statistics | Anwar [AM]
Machine Learning | Stephen [SL]

- **WK01**: Introduction to Bayesian Statistics [AM]
- **WK02**: Bayesian Generalised Linear Modelling (GLMs) [AM]
- **WK03**: Bayesian Generalised Additive Models (GAMs) [AM]
- **WK04**: Deep Learning [SL]
- **WK05**: Convolution Neural Networks (CNNs) [SL]
- **WK06**: GeoAI [SL]
- **WK07**: Bayesian Hierarchical Modelling [AM]
- **WK08**: Spatial Intrinsic Conditional Autoregressive Models [AM]
- **WK09**: Bayesian Updating & Spatiotemporal Areal Analysis [AM]
- **WK10**: Research Methodology, Study Design & Revision [AM]

All are lecture and teaching materials are posted on external websites to Moodle. Bayesian content is on a dedicated GitBook.

Machine Learning content is hosted on a Jupyter Notebook.

Details about assessment

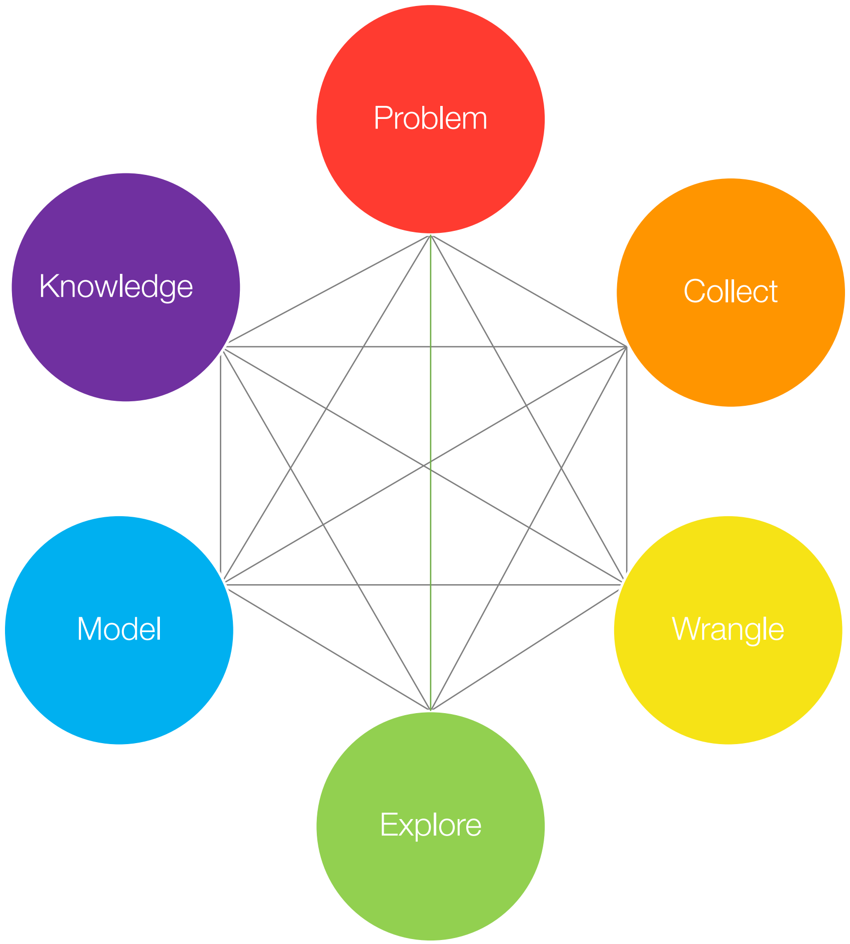
100% coursework with 2 x reports (each 1,500 words)

1 report focuses uses applied Bayesian methods for spatial analysis.

1 report focuses on Machine Learning for spatial analysis and image classification

Contents

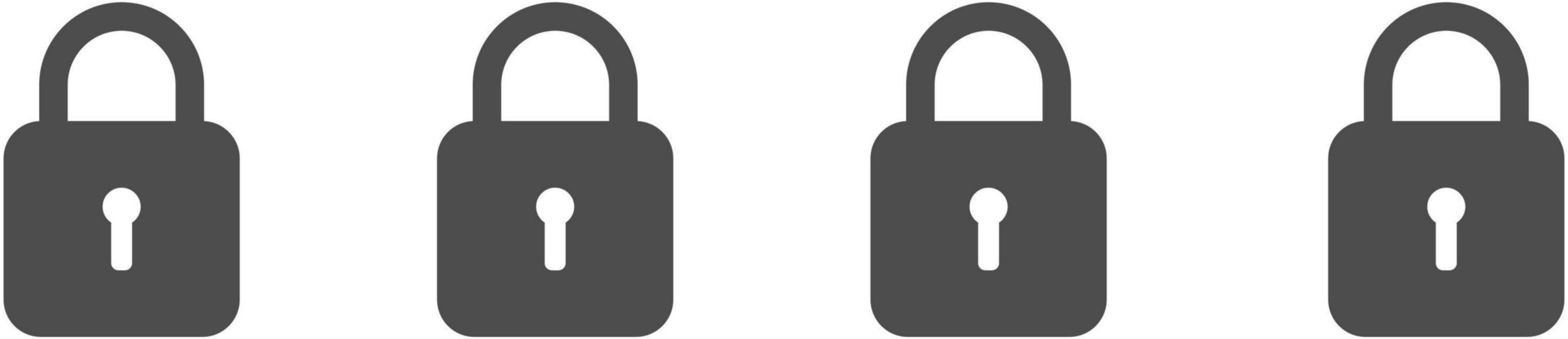
- What is Bayesian Statistics?
- What are the types of probabilities?
- Bayes' theorem
- What are the various Probability Distribution
- The basics of Bayesian Inference



What is a Bayesian Statistics?

Definition:

Bayesian statistics, is using **probabilities** to statistical problems (i.e., framing our research questions using probability statements).



Probabilities | Probability Distribution | Bayes’ Rule | Bayesian Inference

What is a Probability?

Definition:

Probability is the language of uncertainty. It quantifies the uncertainty for a defined event which takes any value between 0 and 1, whereby an event with probabilities near zero implies an event to be very unlikely, and probabilities near one implies an event to be very likely.

- The usual notation for probabilities are **Pr( )** or **P( )** or **Prob( )** (we will use the notation for **Pr( )** throughout this lecture)

- A probability is always a function of some event 'E' which has a set number of outcomes as well as the sample space (or population size) which is the total number of all possible outcomes that can occur for that event 'E'

- We can write this and say **Pr(E)**, which is read as "**the probability that event E will occur**"

Basic terminology: event, outcome & sample space

In probability, the following means:

- An event is referred to as the set of outcomes we can be observe in a study (or experiment).
- An outcome refers to one of the possible result that emerge from a study (or experiment).
- A sample space is the set of all possible outcomes observed in a study (or experiment)

- Example 1: We are conducting a vector-borne mosquito infestation survey in 7 households in an village.

  Event: Infestation status of a household which takes discrete outcomes i.e., "Infested" or "Not Infested"
  Sample space: All observed 7 outcomes = {Infested, Not Infested, Not Infested, Not Infested, Infested, Not Infested, Infested}

- Example 2: We are performing a survey to know the burden of Obesity by measuring BMI status of 10 people.

  Event: Obesity status which takes continuous BMI measurements with any value in this range 15.0kg/m^2 to 66.0kg/m^2
  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}

The notation for writing the following:

- Probability that a house is infested with mosquitoes = Pr(Infested)
- Probability that a person from this cohort is obese = Pr(Obesity)

Example 1: We are conducting a vector-borne mosquito infestation survey in 7 households in an village.

Event: Infestation status of a household which takes discrete outcomes i.e., "Infested" or "Not 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.

Event: Obesity status which takes continuous BMI measurements with any value in this range 15.0kg/m^2 to 66.0kg/m^2
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≥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 (i.e., single) will occur without regards to external circumstances (i.e., past, present or future).

- Usually, the 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 results does not account for other conditions that may effect it.

- It is also called a **Marginal Probability**

- When there's data - unconditional (or marginal) probabilities are calculated accordingly as follows:

Probability that E will occur

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

Probability that E will NOT occur

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

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 deals with more than one event. Say, we have two events i.e., E_1 and E_2. It is the chance (or likelihood) that these two events will occur simultaneously at the same point in time.

- Simply putting it – its basically the joint probability E_1 and E_2 is the probability of event E_1 (i.e., Pr(E_1)) occurring at the same time as the event(s) for E_2 occurs.

- There are several notation for representing joint probabilities:

  $Pr(E_1 \cap E_2)$ [∩ intersection symbol from set theory] where: E_1, E_2 are two different events that intersect (i.e., E1 & E2)
  Pr(E_1 & E_2) or P(E_1E_2) which is joint probability of E_1& E_2

- When there's data – joint probabilities are calculated accordingly as follows:

Remember: Probability that E_1 OR E_2 will (i.e., single event)

$$Pr(E_1) = \frac{\text{Number of observed outcomes } (n_1)}{\text{Total sample space } (N_1)} = \frac{n_1}{N_1}$$

$$Pr(E_2) = \frac{\text{Number of observed outcomes } (n_2)}{\text{Total sample space } (N_2)} = \frac{n_2}{N_2}$$

The joint probabilities are:

$$Pr(E_1 \,\&\, E_2) = Pr(E_1)\ Pr(E_2) = \frac{n_1}{N_1} \times \frac{n_2}{N_2}$$

Note 1: You need to calculate the conditional probabilities and multiply them to get the joint.

Note 2: When you think of joint always think on multiplication

Note 3: Pr(AnB)/U

Conditional Probabilities

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

- Simply putting it – it's the likelihood of an event E_1 occurring, based on the occurrence of a previous event E_2.
- We say: "Probability of E_1 given E_2" – where the probability of event E_1 depends on that of E_2 has happened
- The notation for representing conditional probabilities:

  "|" to represent given
  $Pr(E_1 \mid E_2)$ which means the probability of E_1 given E_2
- Conditional probabilities are computed accordingly as follows:

Conditional probabilities:

$$\Pr(E_1 \mid E_2) = \frac{\Pr(E_1 \,\&\, E_2)}{\Pr(E_2)}$$

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

- Interested in the adult mosquito abundance from breeding sites within a urban/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 a urban setting)
Let the U' represent the event rural area (i.e., total number of breeding sites found in a rural setting)

UCL

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)			0.55 P(U)
Rural (U')			0.45 P(U')
Sum (Column)	0.14 P(B)	0.86 P(B')	Grand total: 1

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

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**

In the second table, we can convert these raw values to joint probabilities. The dark blue shaded cells are the joints probabilities by multiplying their corresponding probabilities.

Example: P(B & U) = P(U)*P(B) etc

Alternatively:

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 (/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)	0.08 Pr(B & U)	0.46 Pr(B' & U)	0.55 Pr(U)
Rural (U')	0.04 Pr(B & U')	0.39 Pr(B' & U')	0.45 Pr(U')
Sum (Column)	0.14 Pr(B)	0.86 Pr(B')	Grand total: 1

In the second table, we can convert these raw values to joint probabilities. The dark blue shaded cells are the joints probabilities by multiplying their corresponding probabilities.

Example: P(B & U) = P(U)*P(B) etc

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

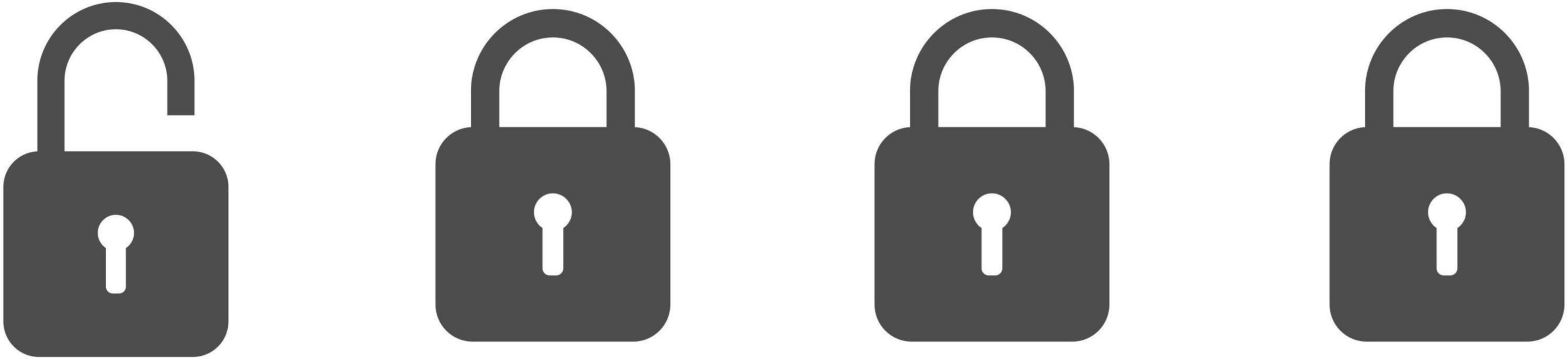
	Breeding sites: Aedes (B)	Breeding sites: No Aedes (B')	Sum (Row)
Urban (U)	0.08 Pr(B & U)	0.46 Pr(B' & U)	0.55 Pr(U)
Rural (U')	0.04 Pr(B & U')	0.39 Pr(B' & U')	0.45 Pr(U')
Sum (Column)	0.14 Pr(B)	0.86 Pr(B')	Grand total: 1

Suppose, we want to know what the probability breeding site have Aedes given the setting is urban i.e., Pr(B|U), a conditional probability.

Conditional probabilities:

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

$$\Pr(B|U) = \frac{\Pr(B \,\&\, U)}{\Pr(U)} = \frac{\left(\frac{B\cap U}{N}\right)}{\frac{U}{N}} = \frac{(B\cap U)}{U} = \frac{\left(\frac{182}{2183}\right)}{1190} = 0.1529 = 15.29\%$$



Probabilities | Probability Distribution | Bayes' Rule | Bayesian Inference

Probability Distributions

Definition:

Probability distribution is a mathematical function that churns out the probability of observing a particular value for a variable (or event). They allow us to also estimate the 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.

Probability Density Function (PDFs) (Continuous)

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

Note: These distributions can handle variables that are continuous in nature

Important note:

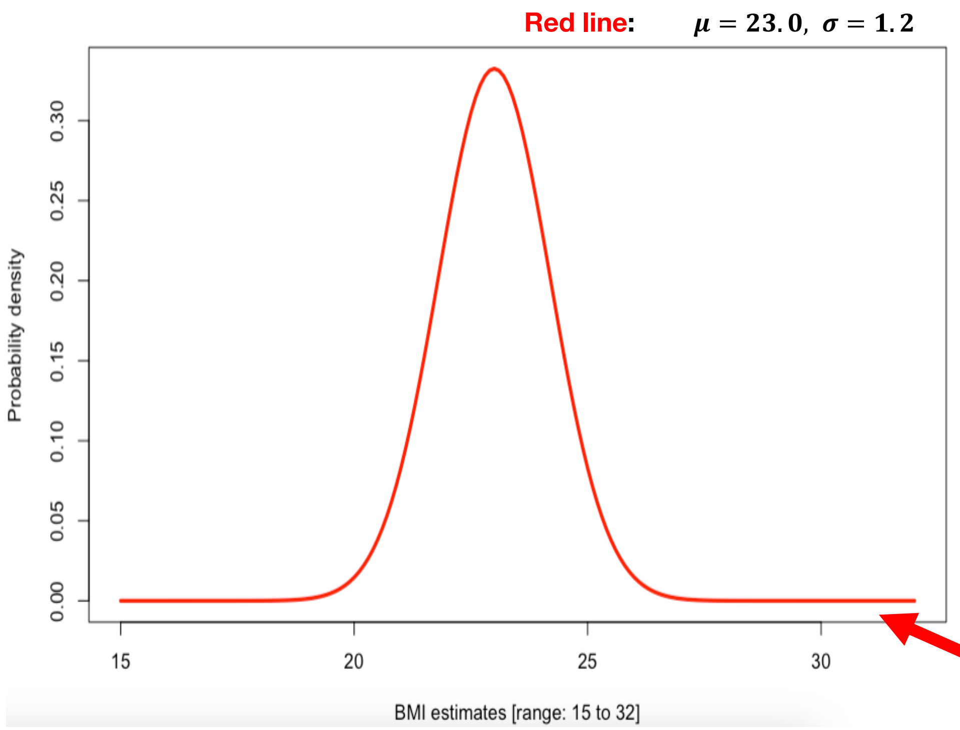
In Bayesian inference, we will have apply probability distributions to the various components of the Bayes' rule formula for the likelihood function and prior probability to derive the posterior probability.

We impose a probability distribution to model an outcome (likelihood), while we subjectively choose another probability distribution (according to our believes) that best describes the shape of the parameters we want to estimate from our outcome

Understanding how to generate a probability distribution [Part 1]:

The key information is knowing the <u>characteristic</u> of the variable, as well as the <u>parameters</u> to be specified or estimated. This will dictate how you will select the appropriate **probability distribution.**

For instance, BMI in children (an <u>interval measure</u>) ranging from 15.0 to 32.0 from school reception, and for the parameters, we have knowledge about them (i.e., say mean and SD)



When determining the probability distribution of such variable. Consider the following:

- **Characteristic:** Here, the BMI measure is an interval and hence continuous; **we are using a PDF (probability density function)**
- **Shape:** Plot of the BMI data strictly resembles a **bell curve** (which is either centred or shifted) – indicates normality
- **Parameters:** The parameter's you have knowledge of a **mean** and **SD**.
- **Probability Density Function**: **<u>Gaussian (or Normal) Distribution</u>**

How was this PDF created?

UCL

Gaussian (or Normal) Distribution [1]

What is this used for?

1. To get a point probability estimate for a measured value (i.e. a probability density)
2. To get a cumulative probability.

$$P(y, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} \longleftrightarrow \mathrm{norm}(\mu, \sigma)$$

- Estimate point probability estimate for some measure of BMI value

FIXED INPUTS

- $\mu = 23.0$ (i.e., assumed mean for BMI)
- $\sigma = 1.2$ (i.e., assumed standard deviation for BMI)
- $\pi = 3.141593$

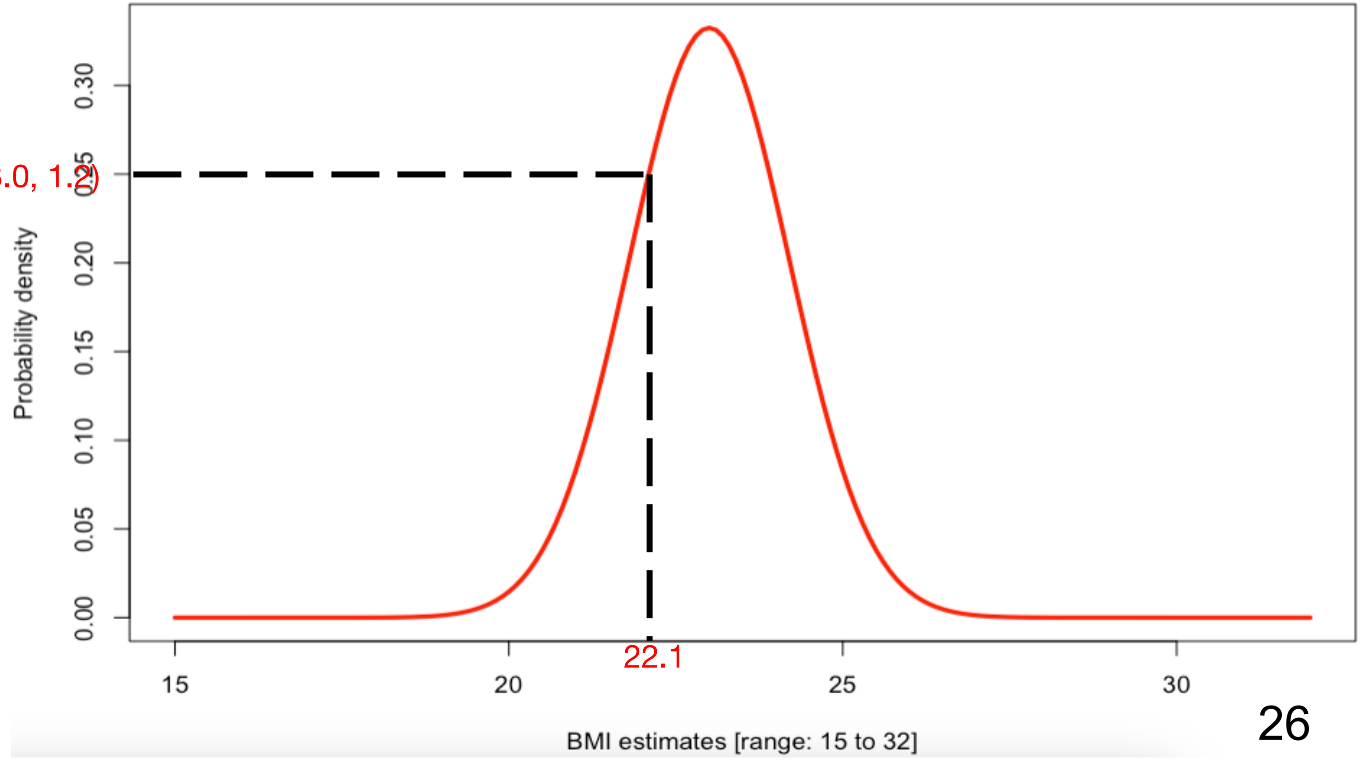
INPUTS

$y =$ BMI data ranging from 15.0 to 32.0 (at increments of 0.1, so 15.0, 15.1, 15.2… 31.8, 31.9 and 32.0)

$$P(y, 23.0, 1.2) = \frac{1}{1.2\sqrt{2(3.1415193)}} e^{-\frac{1}{2}\left(\frac{y-23.0}{1.2}\right)^2}$$

Finding the probability that someone's BMI is 22.1

`dnorm(22.1, mean = 23, sd=1.2)`



For instance, we estimate the probability of someone's BMI being exactly 22.1 which is 0.2509 ≈ 25.1%

Gaussian (or Normal) Distribution [2]

What is this used for?

1. To get a point probability estimate for a measured value (i.e. a probability density)
2. To get a cumulative probability.

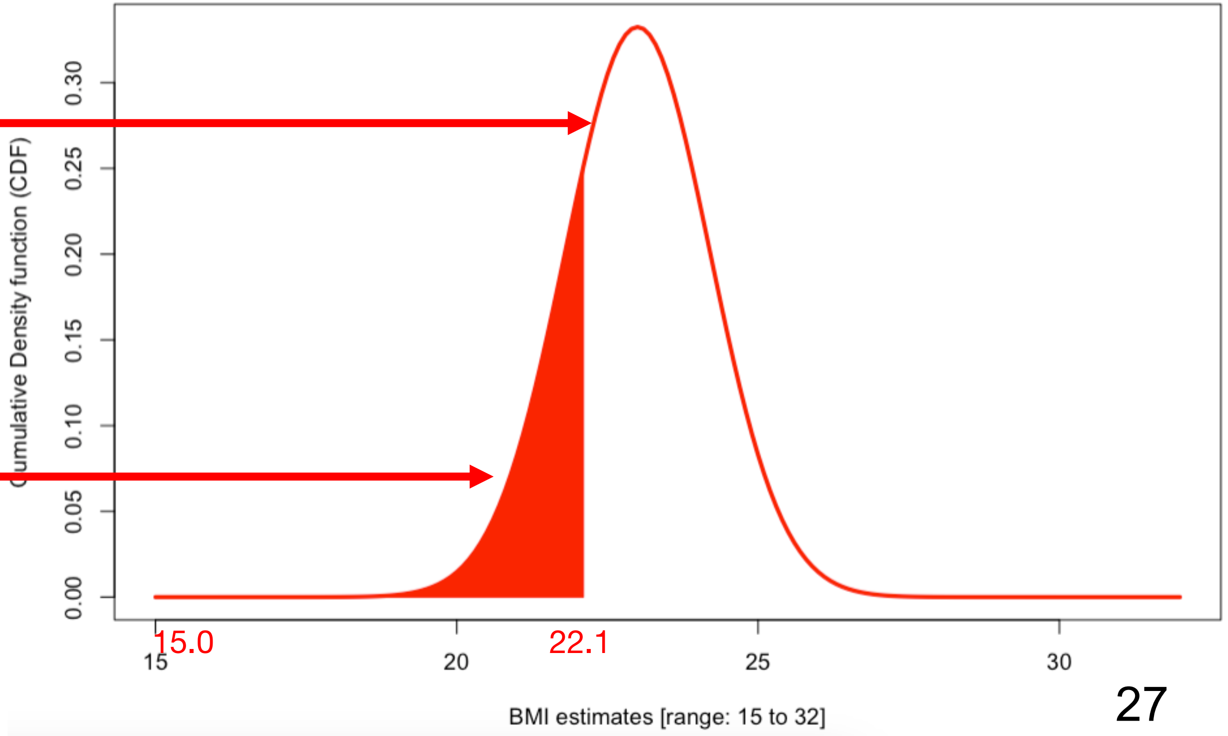
$$P(y,\mu,\sigma) = \int \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy \longleftrightarrow \mathrm{norm}(\mu,\sigma)$$

- Estimate the cumulative probability (**cumulative density function [CDF]**) for some range of BMI values

These INPUTS were used to generate this curve

- $\mu = 23.0$ (i.e., assumed mean for BMI)
- $\sigma = 1.2$ (i.e., assumed standard deviation for BMI)
- $\pi = 3.141593$
- BMI data ranging from 15.0 to 32.0 (at increments of 0.1, so 15.0, 15.1, 15.2… 31.8, 31.9 and 32.0)

```
pnorm(22.1, mean = 23, sd=1.2, lower.tail = TRUE)
```



FINDING TO AREA UNDER THE CURVE FOR BMI being 15.0 to 22.1

$$P(15.0 \leq y \leq 22.1, 23.0, 1.2) = \int \frac{1}{1.2\sqrt{2(3.1415193)}} e^{-\frac{1}{2}\left(\frac{y-23.0}{1.2}\right)^2}$$

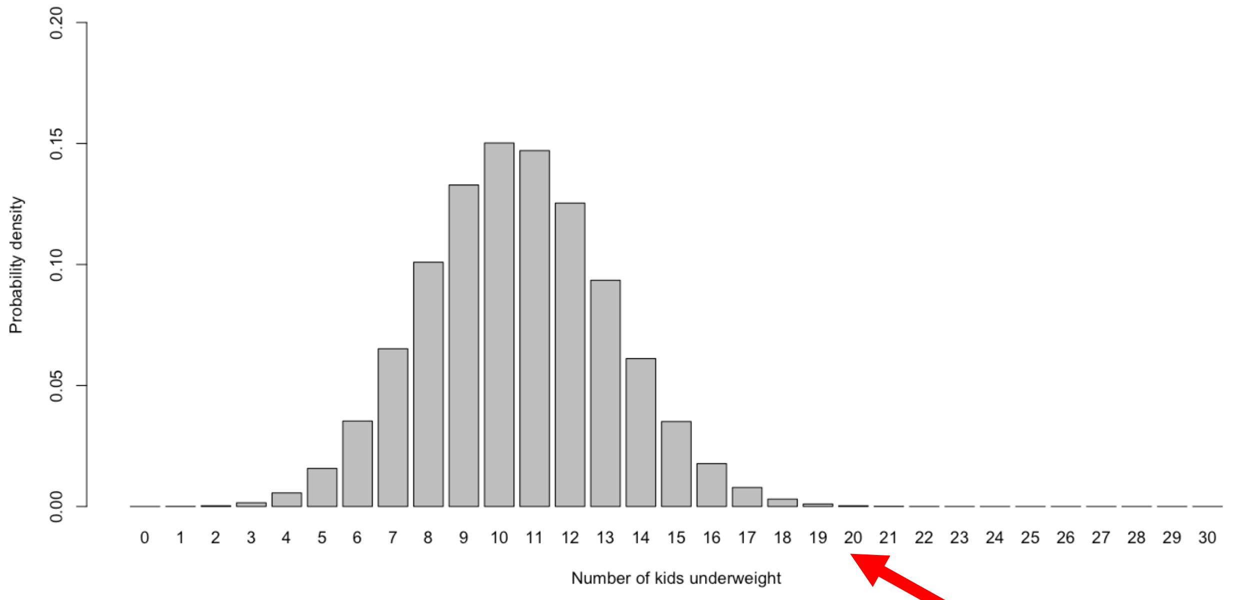
Finding the probability that someone's BMI is at most 22.1

For instance, we estimated the cumulative probability (CDF) of someone's BMI being between 15.0 and 22.1 is 0.2266 ≈ 22.7%. The area of the curve is the estimated CDF.

Understanding how to generate a probability distribution [Part 2]:

The key information is knowing the <u>characteristic</u> of the variable, as well as the <u>parameters</u> to be specified or estimated. This will dictate how you will select the appropriate **probability distribution.**

For instance, we want find out the prevalence of children in reception who are underweight (BMI < 17.5). Without data, we have prior knowledge that 35% of kids are underweight. We understand what the probability will be in observing 7 kids (out of 30 in classroom) being underweight.



When determining the probability distribution of such variable. Consider the following:

- **Characteristic:** Here, the outcome measure is an proportion (i.e., 7 out of 30) with details on numerator and denominator; and hence it discrete; it's aggregated binary measure and so **we are using a PMF (probability mass function)**

- **Parameters:** The parameter's you have knowledge of is a proportion.

- **Probability Mass Function**: **<u>Binomial Distribution</u>**

How was this PMF created?

Binomial Distribution [1]

What is this used for?

1. To get a point probability estimate for a measured value (i.e. a probability density)
2. To get a cumulative probability.

$$P(y, n, \theta) = \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \longleftrightarrow \mathrm{bin}(n, \theta)$$

- Estimate point probability estimate for some number of children who are underweight

FIXED INPUTS

- $n = 30$ (i.e., total number of kids in a class)
- $\sigma = 0.35$ (i.e., prior knowledge of prevalence of children underweight in reception)

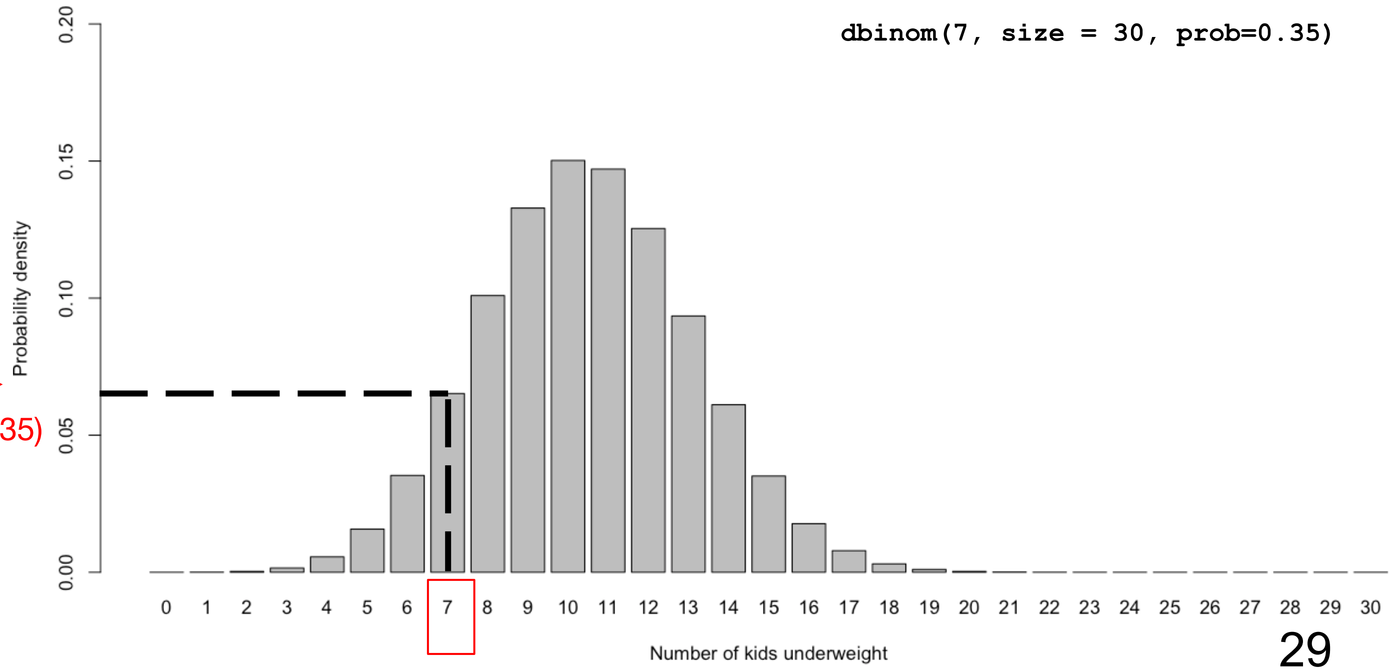
INPUTS

$y =$ number of kids in a class who might be underweight. 0 kids to up to 30 maximum (at increments of 1, so 0, 1, 2, 3, ..., 27, 28, 29 and 30)

$$P(y, 30, 0.35) = \binom{30}{y} 0.35^{y} (1-0.35)^{30-y}$$

P(7, 30, 0.35)

Finding the probability 7 children being underweight (i.e., y = 7)



For instance, we estimate the probability of 7 kids (out of 30 in classroom) being underweight is 0.06519 ≈ 6.52%

Binomial Distribution [2]

What is this used for?

$$P(y,n,\theta) = \sum \binom{n}{y} \theta^{y} (1-\theta)^{n-y} \longleftrightarrow \mathrm{bin}(n,\theta)$$

1. To get a point probability estimate for a measured value (i.e. a probability density)
2. To get a cumulative probability.

- Estimate the cumulative probability (**cumulative mass function [CMF]**) for some range of values

These INPUTS were used to generate this curve

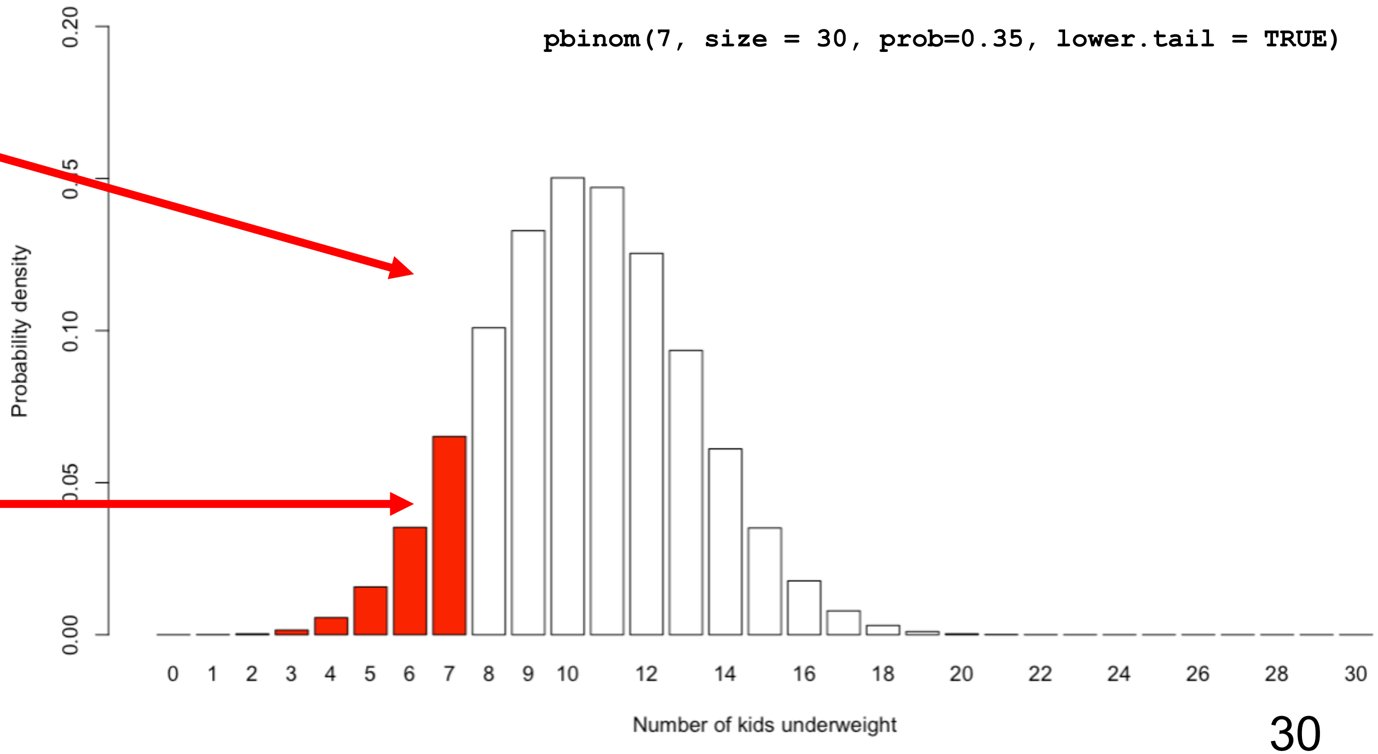
- n = 30 (i.e., total number of kids in a class)
- σ = 0.35 (i.e., prior knowledge of prevalence of children underweight in reception)
- y = number of kids in a class who might be underweight. 0 kids to up to 30 maximum (at increments of 1, so 0, 1, 2, 3, …, 27, 28, 29 and 30)

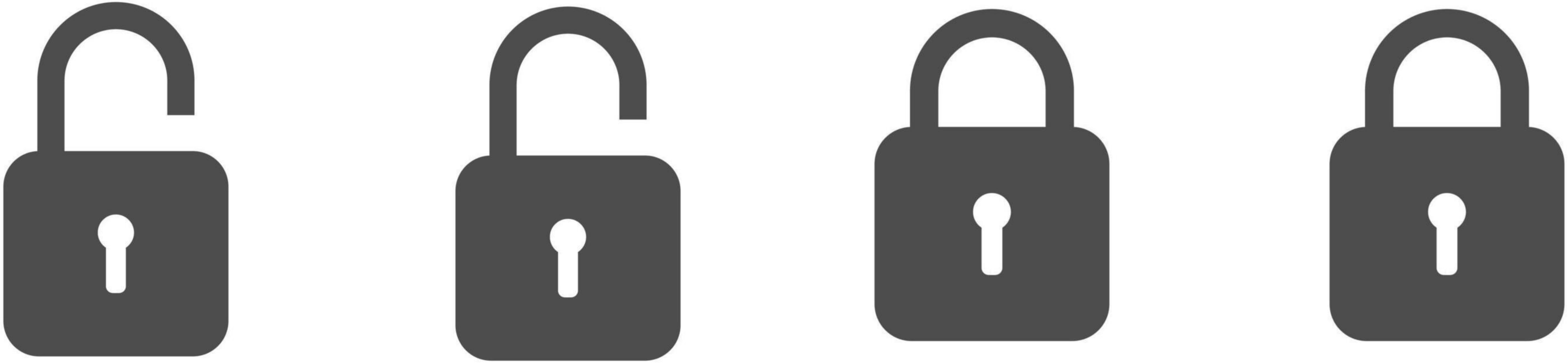
SUM EACH DISCRETE BAR from 0 to 7

$$P(0 \le y \le 7, 30, 0.35) = \sum \binom{30}{y} 0.35^{y} (1-0.35)^{30-y}$$

Finding the probability with y being at most up to 7

For instance, we estimated the cumulative mass probability (CMF) of someone being underweight when y ranges from 0 to 7 is 0.1237 $\approx$ 12.4%. The shaded portion is the estimated CMF.





Probabilities

Probability Distribution

Bayes’ Rule

Bayesian Inference

Bayes' Theorem

Definition:

Bayes’ theorem (or law) determines the conditional probability of a hypothesis (i.e., event) being true given the evidence (i.e., data).

- 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 evidence

The basic formulation for Bayes’ Theorem (or Law):

Probability the hypothesis is true (before any evidence is present)

Probability that a hypothesis is true given the evidence

$$\boxed{\Pr(\mathrm{H} \mid \mathrm{E})} = \frac{\boxed{\Pr(\mathrm{H})}\,\boxed{\Pr(\mathrm{E} \mid H)}}{\boxed{\Pr(\mathrm{E})}}$$

Probability of seeing the evidence **IF** the hypothesis is true

Probability of observing hypothesis in light of the evidence

Thomas Bayes (1701 – 1761)

Prior probability (our belief, we usually don’t know)

Probability the hypothesis is true (before any evidence is present)

Likelihood probability (derived from data)

Probability of seeing the evidence **IF** the hypothesis is true

Probability that a hypothesis is true given the evidence

$$\boxed{\Pr(\mathrm{H} \mid \mathrm{E})} = \frac{\boxed{\Pr(\mathrm{H})}\,\boxed{\Pr(\mathrm{E} \mid H)}}{\boxed{\Pr(\mathrm{E})}}$$

Posterior probability

Probability of observing the evidence

Model prior probability (our belief, we usually don’t know)

- H is the hypothesis whose probability may be affected by data (i.e., evidence)
- Pr(H), the unconditional probability pertained to hypothesis is our prior probability
- Pr(E|H) is the conditional probability of observing the data (i.e., evidence) given the hypothesis. Here, it indicates the compatibility between the data and hypothesis. This is our likelihood.
- Pr(H|E), is conditional probability i.e., what we want to know – the probability of a hypothesis given after we observed the data. This is the posterior probability.

The basic formulation for Bayes' Theorem (or Law):

Prior probability (our belief, we usually don't know)

Probability the hypothesis is true (before any evidence is present)

Likelihood probability (derived from data)

Probability that a hypothesis is true given the evidence

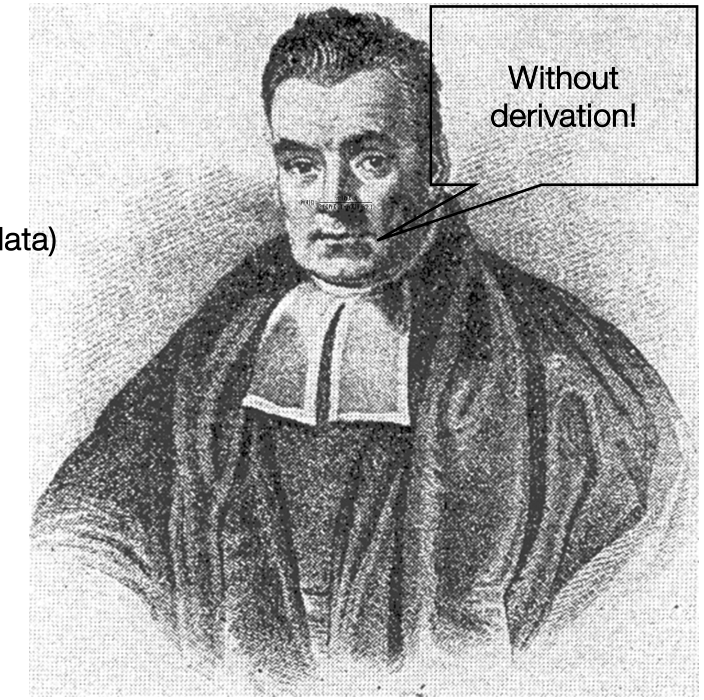
Posterior probability

$$\Pr(\mathrm{H} \mid \mathrm{E}) = \frac{\Pr(\mathrm{H})\,\Pr(\mathrm{E} \mid H)}{\Pr(\mathrm{E})}$$

Probability of seeing the evidence **IF** the hypothesis is true

Probability of observing the evidence

Model prior probability (our belief, we usually don't know)



The expanded version of the Bayes' Theorem (or Law):

$$\Pr(\mathrm{H} \mid \mathrm{E}) = \frac{\Pr(\mathrm{H})\,\Pr(\mathrm{E} \mid \mathrm{H})}{\Pr(\mathrm{H})\,\Pr(\mathrm{E} \mid \mathrm{H}) \;+ \Pr(\mathrm{E} \mid \mathrm{H}')\,\Pr(\mathrm{H}')}$$

This is still Pr(E) i.e., the probability of observing the evidence

The great thing about this equation is the fact that it allows us to posit specific hypotheses for some phenomena and in turn express our current belief that each hypothesis is true. Then, as new information (i.e., data) becomes available, we update our belief in the hypothesis.

Let's see the Bayes' theorem in action!

Example with Bayes' Rule: Study on measuring abundance of Adult mosquitoes in Location A within urban setting

	Breeding sites: Aedes	Breeding sites: No Aedes	Sum (Row)
Urban	182 (/2183) = 0.08	1008	1190
Rural	132	861	993
Sum (Column)	314	1869	Grand total: 2183

$$\Pr(H \mid E) = \frac{\Pr(H)\Pr(E \mid H)}{\Pr(E)}$$

$$\Pr(H \mid E) = \frac{\Pr(H)\Pr(E \mid H)}{\Pr(H)\Pr(E \mid H) + \Pr(E \mid H')\Pr(H')}$$

After engaging with community workers about mosquito infestation in Location A. We suspect that households in Location A have a 1 in 5 risk are infested with the Aedes mosquitoes. From a previous survey, the prevalence of infestation based on an urban setting is 8.0%. How probable is it that households within an urban slum setting are infested with Aedes mosquito?

- Prior: We hypotheses that 1 in 5 households are infested: $\Pr(H) = 1/5 = 0.2$
- Likelihood: Within the urban setting, the prevalence of infestation is 8.0%: $\Pr(E \mid H) = 0.08$
- Model Priors: $\Pr(H') = 1 - 0.2 = 0.8$; $\Pr(E \mid H') = 1 - 0.08 = 0.92$
- Posterior:

$$\Pr(H \mid E) = \frac{\Pr(H)\Pr(E \mid H)}{\Pr(H)\Pr(E \mid H) + \Pr(E \mid H')\Pr(H')} = \frac{0.2(0.08)}{0.2(0.08) + 0.8(0.92)} = 0.0212 = 2.12\%$$

Back to the Bayes' Theorem and re-writing the formula

Prior probability (our belief, we usually don't know)

Probability the hypothesis is true (before any evidence is present)

Likelihood probability (derived from data)

Probability that a hypothesis is true given the evidence

$$\boxed{\Pr(\mathrm{H} \mid \mathrm{E})} = \frac{\boxed{\Pr(\mathrm{H})}\,\boxed{\Pr(\mathrm{E} \mid H)}}{\boxed{\Pr(\mathrm{E})}}$$

Probability of seeing the evidence **IF** the hypothesis is true

Posterior probability

Probability of observing the evidence

Model prior probability (our belief, we usually don't know)

- Re-write the formula and expressing it in terms of y (data or evidence) and θ (hypothesis or parameter) to estimate

$$\Pr(\mathrm{H} \mid \mathrm{E}) = \frac{\Pr(\mathrm{H})\,\Pr(\mathrm{E} \mid H)}{\Pr(\mathrm{E})} \equiv \Pr(\theta \mid \mathrm{y}) = \frac{\Pr(\theta)\,\Pr(\mathrm{y} \mid \theta)}{\boxed{\Pr(\mathrm{y})}} \quad >> \quad \Pr(\theta \mid \mathrm{y}) \propto \Pr(\theta)\,\Pr(\mathrm{y} \mid \theta)$$

This can removed since we only care about the ones with θ

Posterior ∝ Prior × Likelihood

This forms the framework for Bayesian Statistics

Probabilities

Probability Distribution

Bayes’ Rule

Bayesian Inference

Bayesian Inference

Performing Bayesian Inference

- For point or interval estimation of a parameter θ, the inference is based off:

$$\Pr(\theta \mid y) = \frac{\Pr(\theta)\Pr(y \mid \theta)}{\Pr(y)} \propto \Pr(\theta)\Pr(y \mid \theta)$$

Where

- $\Pr(\theta)$ is the prior density for the parameter θ, which must be given a prior probability function
- $\Pr(y \mid \theta)$ is the statistical model (or likelihood), which must be given a probability function depending y
- $\Pr(\theta \mid y)$ is the posterior density for the parameter θ

- Specifying the likelihood's probability function is very easy. For the usual you only need to use normal, Poisson or binomial distribution because we often deal with these types of outcomes.

- We take the easier way out when specifying the probability function for the priors – here, we want to choose a function which, in turn, will lead to the posterior distribution PF being the same as the prior. This is what we call a Conjugate.

Back to the Bayes' Theorem and re-writing the formula [2]

- For point or interval estimation of a parameter θ, the inference is based off:

$$\Pr(\theta \mid y) = \frac{\Pr(\theta)\,\Pr(y \mid \theta)}{\Pr(y)} \propto \Pr(\theta)\,\Pr(y \mid \theta)$$

Where
- $\Pr(\theta)$ is the prior density for the parameter θ, which must be given a prior probability function
- $\Pr(y \mid \theta)$ is the statistical model (or likelihood), which must be given a probability function depending y
- $\Pr(\theta \mid y)$ is the posterior density for the parameter θ

- Specifying the likelihood's probability function is very easy. Often, you will be using normal, Poisson or binomial distribution to deal continuous, counts, or binary outcomes respectively.

- The difficulty lies in the specification of the prior probability function. Here its entirely subjective.

Motivating example [1]:

$$\Pr(\theta \mid y) = \frac{\Pr(\theta)\,\Pr(y \mid \theta)}{\Pr(y)} \propto \Pr(\theta)\,\Pr(y \mid \theta)$$

- What is the probability (or prevalence) of infestation in Recife this year?

Survey year: 05/2023	Number properties detected with Aedes	Overall number of properties surveyed
* Most recent data collection effort.	428	976

Important Information:
- y represent the number of infested properties (428)
- n represent the overall number of properties surveyed (976)
- θ represent the unknown probability (or prevalence) of infestation
- Prior information for θ (i.e., our knowledge or belief) is assumed 0.20 (in most cases, the prevalence from past research is often this value of 20-25%.

Likelihood function: $P(y \mid \theta) = \binom{n}{y}\,\theta^{y}(1-\theta)^{n-y}$

Alternatively, instead type this mathematical notation, you use this statistical notation: $y \sim \mathrm{Bin}(n, \theta)$

- We have specified the probability function for the likelihood, what about the function for the prior i.e., $P(\theta)$?

```
dbeta(prop, α − 1, β − 1)
pbeta(prop, α − 1, β − 1)
```

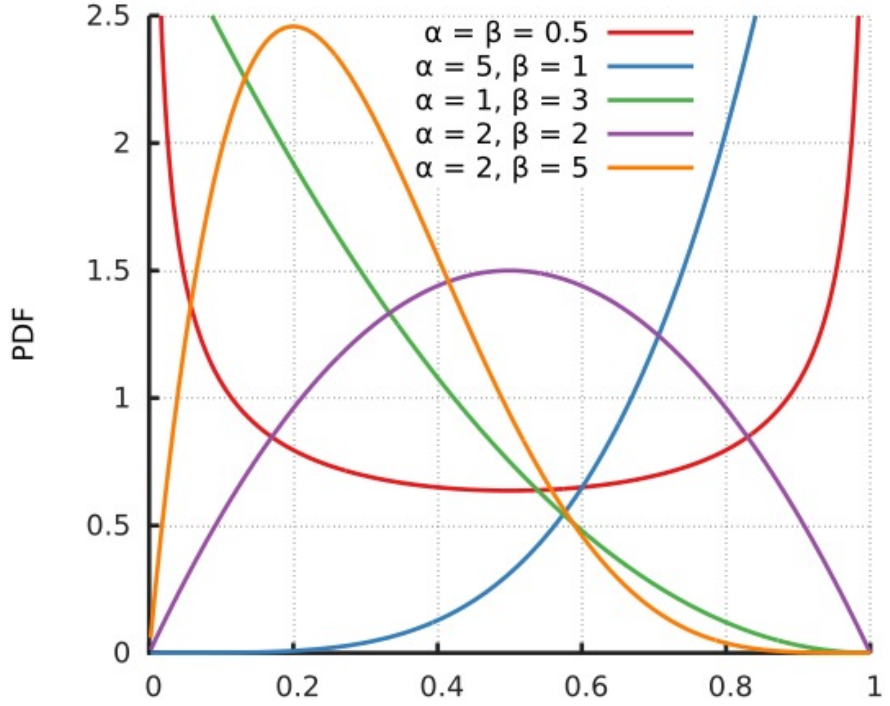
Motivating example [2]:

- The probability distribution needed in this situation is a Beta distribution. It is the best probability function to use a prior distribution for unknown parameter that's a proportion.

$$P(\theta) = \frac{1}{B(\alpha, \beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, (0 < \theta < 1)$$

$$P(\theta) = \text{beta}(\theta|\alpha - 1, \beta - 1)$$

```
Posterior mean:           qbeta(0.5,   α − 1,  β − 1)
Posterior lower limit:    qbeta(0.025, α − 1,  β − 1)
Posterior upper limit:    qbeta(0.975, α − 1,  β − 1)
```



- Example of a flexible PDF function as we can bend it to accordingly be setting values to α and β
- Here, we need to use values for α and β which gives us a distribution with a shape that's concentrated on 20-25%.
- This type of prior is an **informative prior**, because we've assigned a distribution with information that's specific.

Solutions [1]:

- Likelihood function: $P(y \mid \theta) = \binom{n}{y}\theta^{y}(1-\theta)^{n-y}$
- Prior: $P(\theta) = \frac{1}{B(\alpha,\beta)}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$

- For a binomial model specification, with Beta prior, the posterior is:

$$
\begin{aligned}
P(\theta \mid y) &\propto P(\theta)\,P(y \mid \theta) \\
&\propto [\theta^{\alpha-1}(1-\theta)^{\beta-1}][\theta^{y}(1-\theta)^{n-y}] \\
&\Rightarrow \theta^{\alpha-1}\theta^{y}(1-\theta)^{\beta-1}(1-\theta)^{n-y} \\
&\Rightarrow \theta^{(\alpha+y)-1}\,(1-\theta)^{(\beta+n-y)-1} \\
&= \mathrm{Beta}(\theta \mid (\alpha+y)-1,\ (\beta+n-y)-1)
\end{aligned}
$$

All constants with the parameter of interest can be removed i.e., $\binom{n}{y}$ and $\frac{1}{B(\alpha,\beta)}$

Arranging like terms so θ's and $(1-\theta)$'s together

Using indices $a^{m}a^{n} = a^{m+n}$

- As you can see the posterior distribution is a Beta distribution, but the parameters for it have been updated.
- When the posterior is in the same family as the prior, we say the prior is a conjugate for the model.
- Here, the Beta prior is a conjugate for the binomial model.

Solutions [2]:

Updated Bayesian model: $P(\theta|Y) \propto \theta^{(\alpha+y)-1}(1-\theta)^{(\beta+n-y)-1}$

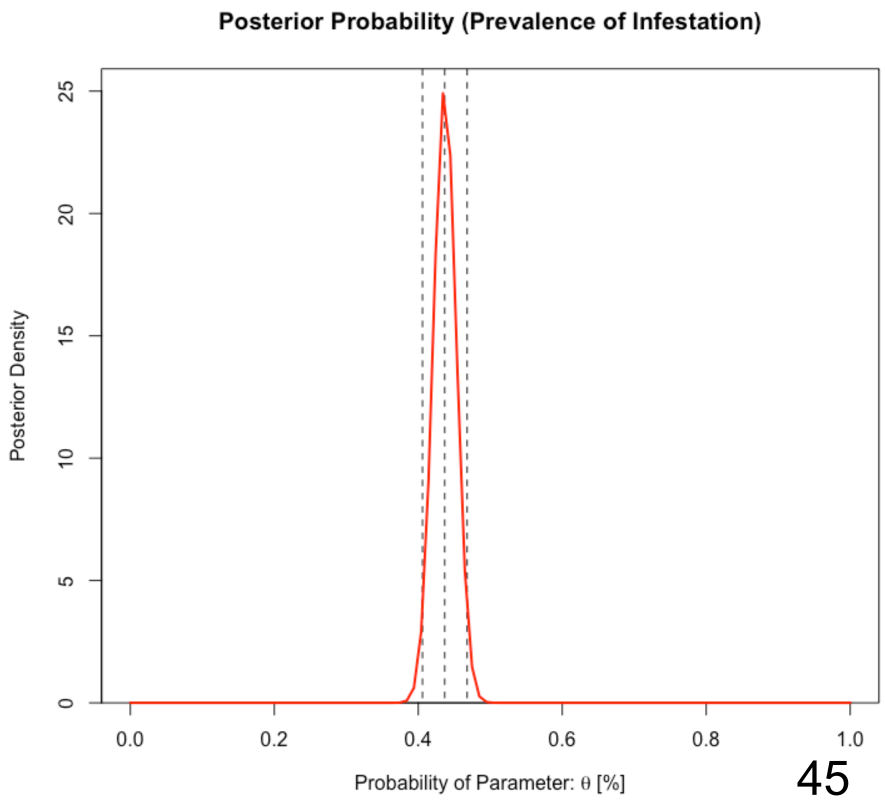
- Combining this with infestation data, the posterior is $\text{Beta}(\theta|\ (\alpha+y)-1, (\beta+n-y)-1)$
- Using a informative prior in this situation, where α = 1 and β = 5, we get the following

$$\text{Beta}(\theta|\ (1+428)-1, (5+976-428)-1)$$

- Above model will generate a posterior distribution where θ is from 0 to 1.

Posterior mean: qbeta(0.5, $(1+428-1,\ (5+976-428)\ -1)$
Posterior lower limit: qbeta(0.025, $(1+428-1,\ (5+976-428)\ -1)$
Posterior upper limit: qbeta(0.975, $(1+428-1,\ (5+976-428)\ -1)$

- The mean prevalence is approximately 44% with 95% credible intervals (40% to 47%)



Selection of priors for Likelihood $\mathbf{P}(\boldsymbol{Y}|\boldsymbol{\theta})$ or Priors $\mathbf{P}(\boldsymbol{\theta})$

Characteristic	Check shape	Function	Distribution	Statistical Notation	Usage
• Continuous measure (scale, interval)	Bell shape (centre, or shifted)	PDF	Gaussian	norm(mu, sigma)	• Set it as a likelihood • Use as prior
• Proportion (with information on the numerator and denominator)	Does not matter	PMF	Binomial	bin(n, p)	• Set it as a likelihood (only)
• Counts	Does not matter	PMF	Poisson	poisson(rate)	• Set it as a likelihood (only)
• Binary measure (Yes, No)	Does not matter	PMF	Bernoulli	bern(p)	• Set it as a likelihood (only)
• Continuous measure (scale, interval) • Proportion (with information on the numerator and denominator) • Counts	Flat shape	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	Skewed shape	PDF/PMF	Gamma	Gamma(rate, shape)	• Use ONLY as a prior • It's flexible
• Proportion (with information on the numerator and denominator)	Does not matter	PDF	Beta	Beta(theta, shape1, shape2)	• Use ONLY as a prior • It's flexible

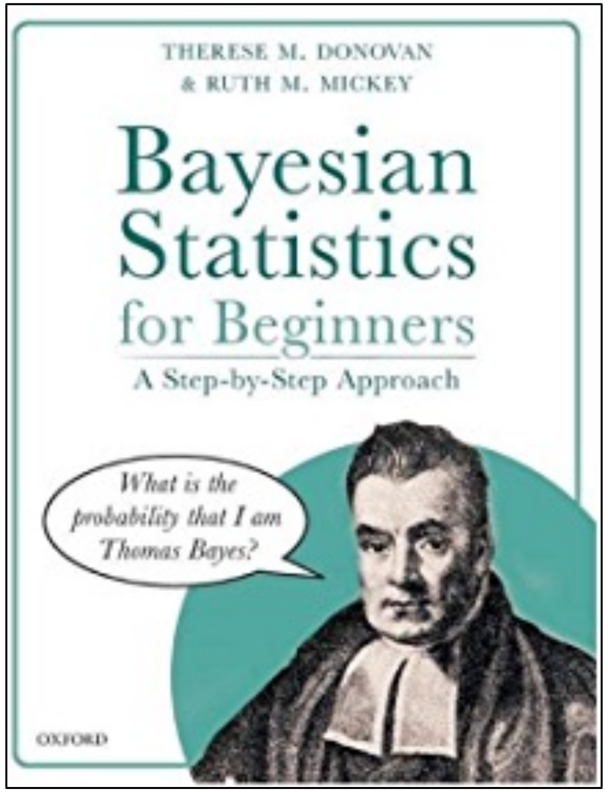
Probabilities

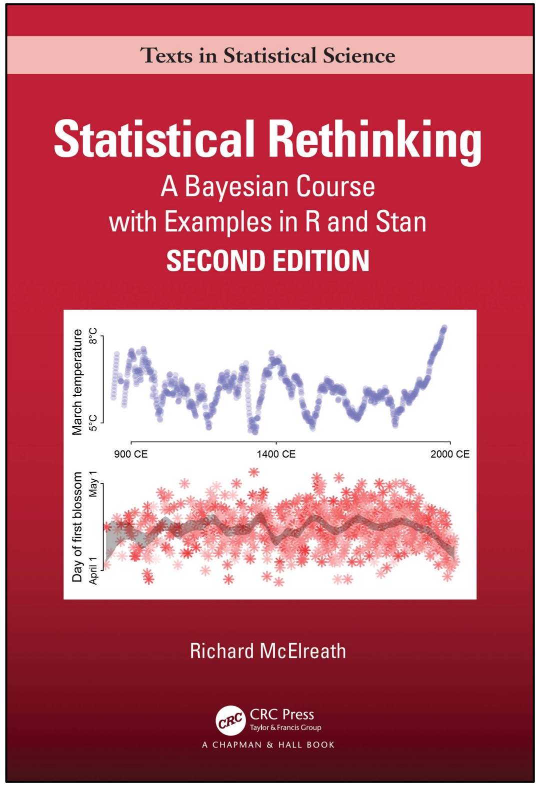
Bayes’ Rule

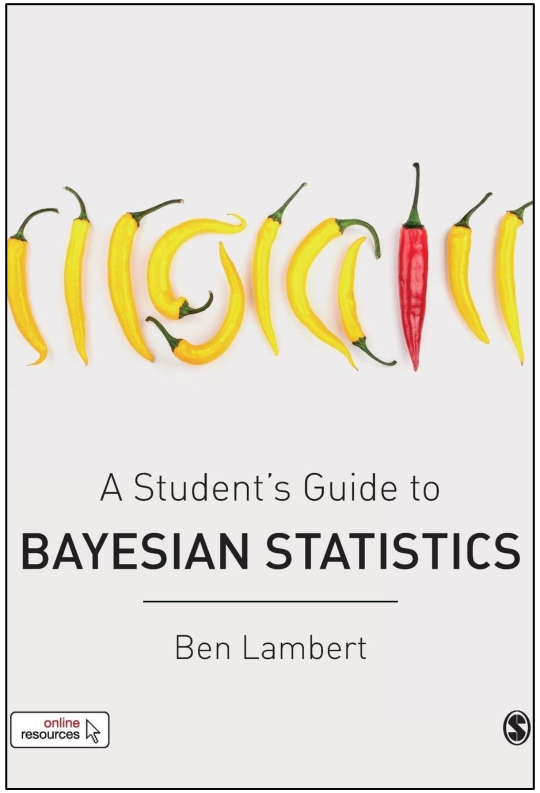
Bayesian Inference

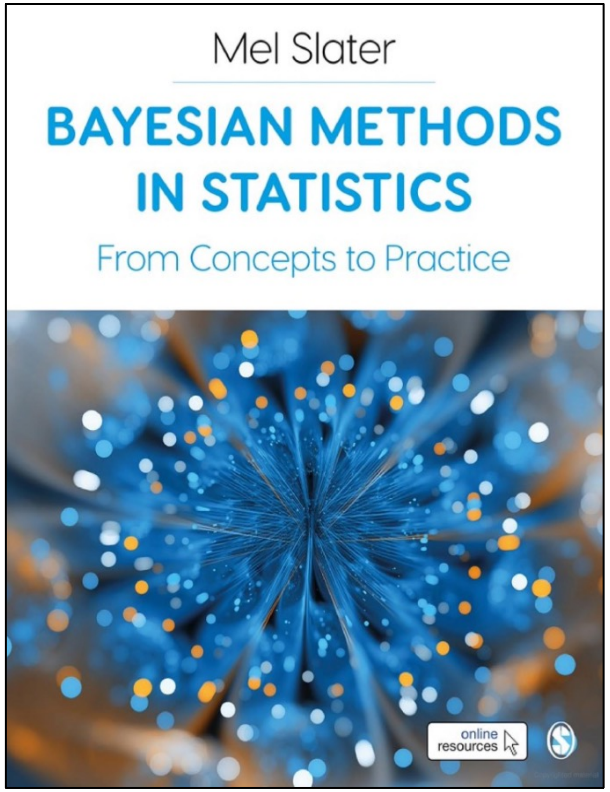
Probability Functions

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

Any questions?