# Probability

Michael Noonan

January 17, 2021

Biol 520C: Statistical modelling for biological data

# Table of contents

# Housekeeping

# Housekeeping



- Practical format. Do we want to move to something standardised?

# Review

# Clicker question: i

THE UNIVERSITY OF BRITISH COLUMBIA
Okanagan Campus

The following equation

$$y_i = \beta_0 + x_i\beta_1 + \varepsilon_i, \qquad \text{where} \qquad \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Describes:

**A** — A system where $y$ is proportional to $x$.

**B** — A model with a mean 0 Gaussian error

**C** — A model with both stochastic and deterministic components

**D** — All of the above.

# Clicker question: i



The following equation

$$y_i = \beta_0 + x_i\beta_1 + \varepsilon_i, \qquad \text{where} \qquad \varepsilon_i \sim \mathcal{N}(0,\, \sigma^2)$$

Describes:

**A** — A system where $y$ is proportional to $x$.

**B** — A model with a mean 0 Gaussian error

**C** — A model with both stochastic and deterministic components

**D** — All of the above.

# Simple linear regression



Last lecture we covered how simple linear regression models fit by least squares provides a formal description of the deterministic components of a system where $y$ is proportional to $x$:

$$y_i = \beta_0 + x_i\beta_1$$

# Simple linear regression

Last lecture we covered how simple linear regression models fit by least squares provides a formal description of the deterministic components of a system where $y$ is proportional to $x$:

$$y_i = \beta_0 + x_i\beta_1$$

Minimising the sum of the squares provided a path forward for the deterministic part, but provided no description of our model's stochastic component.

Last lecture we covered how simple linear regression models fit by least squares provides a formal description of the deterministic components of a system where $y$ is proportional to $x$:

$$y_i = \beta_0 + x_i\beta_1$$

Minimising the sum of the squares provided a path forward for the deterministic part, but provided no description of our model's stochastic component.

This led us to approaching the problem as probabilists.

Who even cares, how does that help?

# Probability Theory 101

# Preamble

# Preamble



If you were to look around UBC's libraries you'd probably find dozens if not hundreds of books on probability theory.

# Preamble



If you were to look around UBC's libraries you'd probably find dozens if not hundreds of books on probability theory. There's no way I could teach you all of this material in a single lecture.

# Preamble



If you were to look around UBC's libraries you'd probably find dozens if not hundreds of books on probability theory. There's no way I could teach you all of this material in a single lecture.

I'm not trying to teach you 'probability theory', but I am trying to give you those pieces you'll need to understand the concepts relevant to this course.

# Preamble



If you were to look around UBC's libraries you'd probably find dozens if not hundreds of books on probability theory. There's no way I could teach you all of this material in a single lecture.

I'm not trying to teach you 'probability theory', but I am trying to give you those pieces you'll need to understand the concepts relevant to this course.

Hopefully this will also motivate you to take a deeper dive into probability theory and probability distributions outside of this course.

# Why probability theory?

# Why probability theory?



Biological data are very noisy, and full of ‘randomness’.

# Why probability theory?



Biological data are very noisy, and full of ‘randomness’.

Being a good modeller means being able to understand not only the deterministic part of a model, but also the stochastic part.

# Why probability theory?



Biological data are very noisy, and full of 'randomness'.

Being a good modeller means being able to understand not only the deterministic part of a model, but also the stochastic part.

Most introductory statistics courses teach you methods that assume the randomness of a process is normally distributed

# Why probability theory?

Biological data are very noisy, and full of ‘randomness’.

Being a good modeller means being able to understand not only the deterministic part of a model, but also the stochastic part.

Most introductory statistics courses teach you methods that assume the randomness of a process is normally distributed, in many cases this is totally fine

# Why probability theory?

Biological data are very noisy, and full of 'randomness'.

Being a good modeller means being able to understand not only the deterministic part of a model, but also the stochastic part.

Most introductory statistics courses teach you methods that assume the randomness of a process is normally distributed, in many cases this is totally fine, in many others though this is not an appropriate assumption.

# Why probability theory?



Biological data are very noisy, and full of 'randomness'.

Being a good modeller means being able to understand not only the deterministic part of a model, but also the stochastic part.

Most introductory statistics courses teach you methods that assume the randomness of a process is normally distributed, in many cases this is totally fine, in many others though this is not an appropriate assumption.

In order to make sense of a system's stochasticity, we need to rely on probability distributions.

# Why probability theory?



Biological data are very noisy, and full of 'randomness'.

Being a good modeller means being able to understand not only the deterministic part of a model, but also the stochastic part.

Most introductory statistics courses teach you methods that assume the randomness of a process is normally distributed, in many cases this is totally fine, in many others though this is not an appropriate assumption.

In order to make sense of a system's stochasticity, we need to rely on probability distributions. In order to work with probability distributions, we need to understand some probability theory.

# History of probability theory

# History of probability theory



In probability theory we are concerned with the occurrence of random events.

# History of probability theory



In probability theory we are concerned with the occurrence of random events.

More specifically, we're interested not in single outcomes, but average outcomes over a large number of replicates (flipping coins, rolling die, picking names from a hat, etc...)

# History of probability theory



In probability theory we are concerned with the occurrence of random events.

More specifically, we're interested not in single outcomes, but average outcomes over a large number of replicates (flipping coins, rolling die, picking names from a hat, etc...)

What group of people have a lot of experience with the outcomes of random chance events?

In probability theory we are concerned with the occurrence of random events.

More specifically, we're interested not in single outcomes, but average outcomes over a large number of replicates (flipping coins, rolling die, picking names from a hat, etc...)

What group of people have a lot of experience with the outcomes of random chance events? Gamblers.

# History of probability theory

# History of probability theory



Pascal Source: Wikipedia

de Fermat Source: Wikipedia

In the mid 1600s when a professional gambler asked French mathematician Pierre de Fermat why if he bet on rolling at least one six in four throws of a die he won in the long term, whereas betting on throwing at least one double-six in 24 throws of two dice resulted in his losing on average.

# History of probability theory



Pascal Source: Wikipedia

de Fermat Source: Wikipedia

In the mid 1600s when a professional gambler asked French mathematician Pierre de Fermat why if he bet on rolling at least one six in four throws of a die he won in the long term, whereas betting on throwing at least one double-six in 24 throws of two dice resulted in his losing on average.

de Fermat worked with Blaise Pascal to show mathematically why this was the case...

# History of probability theory cont.

# History of probability theory cont.



de Fermat worked out that

Prob. of one 6 in 4 throws = 1- Prob. of no 6 in 4 throws
$= 1 - (5/6)^4$
$= 0.518$ (i.e., winning on average)

Whereas

Prob. of 6-6 in 24 = 1- Prob. of no 6-6 in 24
$= 1 - (35/36)^{24}$
$= 0.491$ (i.e., losing on average)

de Fermat worked out that

Prob. of one 6 in 4 throws = 1- Prob. of no 6 in 4 throws
$= 1 - (5/6)^4$
$= 0.518$ (i.e., winning on average)

Whereas

Prob. of 6-6 in 24 = 1- Prob. of no 6-6 in 24
$= 1 - (35/36)^{24}$
$= 0.491$ (i.e., losing on average)

...and this work became the foundation of modern probability theory.

# Probability theory 101



In probability theory we are concerned with the occurrence of random events.

# Probability theory 101



In probability theory we are concerned with the occurrence of random events.

(Think of an event as the outcome of an experiment)

# Probability theory 101



In probability theory we are concerned with the occurrence of random events.

(Think of an event as the outcome of an experiment)

We write this:

$$\Pr\{A\} = \text{Probability that event A occurs,}$$

# Probability theory 101



In probability theory we are concerned with the occurrence of random events.

(Think of an event as the outcome of an experiment)

We write this:

$$\Pr\{A\} = \text{Probability that event A occurs},$$

$$\Pr\{B\} = \text{Probability that event B occurs},$$

etc...

# Probability of an event

# Probability of an event



Let's say S is the collection of all possible outcomes of our 'experiment' (sides on a coin, numbers on a die, possible ages, whatever)

# Probability of an event



Let's say S is the collection of all possible outcomes of our 'experiment' (sides on a coin, numbers on a die, possible ages, whatever)

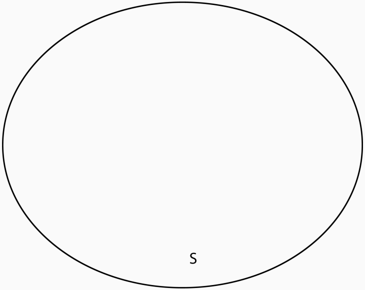
This collection of outcomes is termed the 'sample space'

# Probability of an event



Let's say S is the collection of all possible outcomes of our 'experiment' (sides on a coin, numbers on a die, possible ages, whatever)

This collection of outcomes is termed the 'sample space', the sum of all the probabilities in the sample space is 1 ($\Pr\{S\} = 1$)

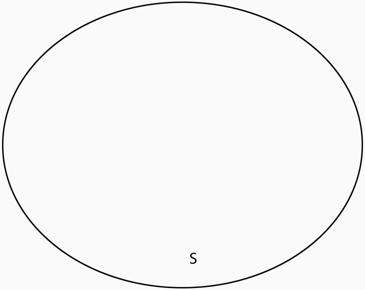


# Probability of an event



Let's say S is the collection of all possible outcomes of our 'experiment' (sides on a coin, numbers on a die, possible ages, whatever)

This collection of outcomes is termed the 'sample space', the sum of all the probabilities in the sample space is 1 ($\Pr\{S\} = 1$)




For tossing a single six-sided die, the sample space is $\{1, 2, 3, 4, 5, 6\}$.

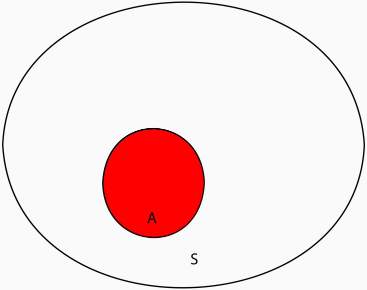
# Probability of an event, cont.



We carry out our experiment and we observe event ‘A’

# Probability of an event, cont.

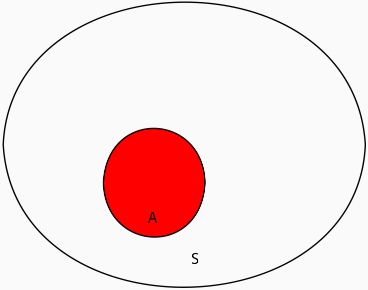
We carry out our experiment and we observe event 'A'



# Probability of an event, cont.



We carry out our experiment and we observe event 'A'

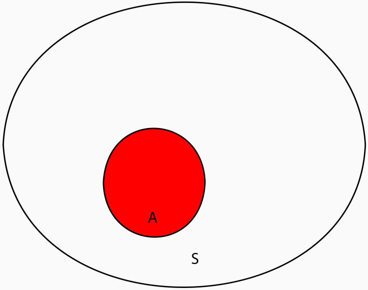



$\Pr\{A\} = \text{Probability of event A}$

# Probability of an event, cont.



We carry out our experiment and we observe event 'A'

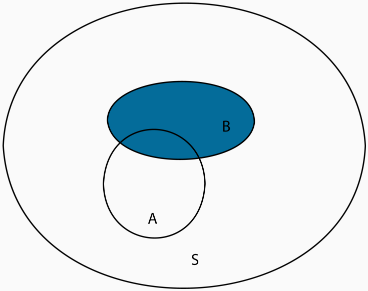



$$\begin{aligned}\Pr\{\text{A}\} &= \text{Probability of event A}\\ &= (\text{area of A}) \;/\; (\text{area of S})\end{aligned}$$

# Probability of an event, cont.



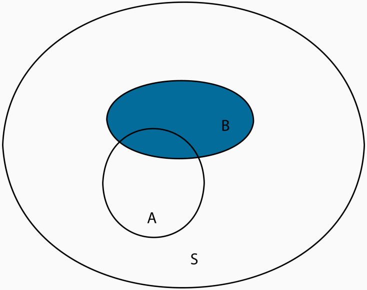
We carry out our experiment again and we observe event 'B'



# Probability of an event, cont.



We carry out our experiment again and we observe event 'B'




$$\Pr\{B\} = (\text{area of B}) \; / \; (\text{area of S})$$

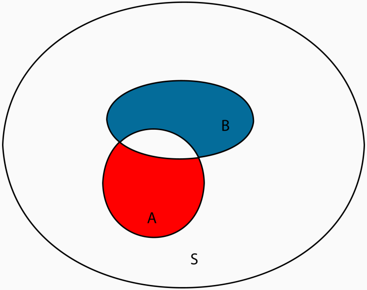
# Probability of an event, cont.



What about the probability of either ‘A’ **or** ‘B’ ?

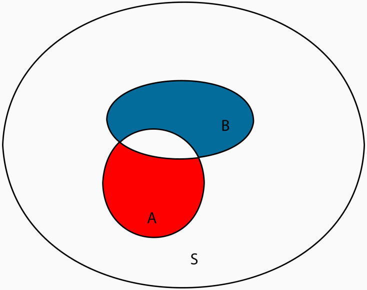
# Probability of either ‘A’ or ‘B’





# Probability of either 'A' or 'B'

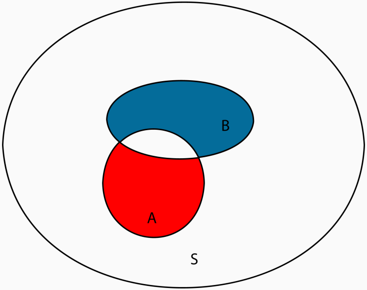





$\Pr\{\text{A or B}\} =$

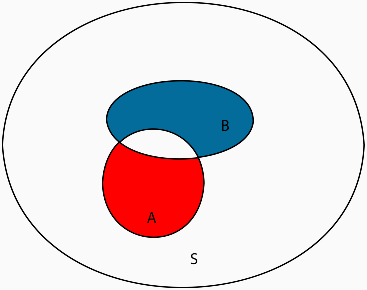
# Probability of either 'A' or 'B'






$$\Pr\{A \text{ or } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \text{ and } B\}$$

# Probability of either 'A' or 'B'




$$\Pr\{A \text{ or } B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \text{ and } B\}$$

Note: more formally the Pr{A and B} is denoted as Pr{A,B}

What about the probability of ‘B’ given ‘A’ occurred?

# Conditional probability



What about the probability of 'B' given 'A' occurred?

This is termed conditional probability (i.e., the probability of an event under the condition that another event occurred)

# Conditional probability



What about the probability of ‘B’ given ‘A’ occurred?

This is termed conditional probability (i.e., the probability of an event under the condition that another event occurred)

Events follow each other all the time in reality.

# Conditional probability cont.



The probability of event B = Pr{B} = (area of B) / (area of S)

# Conditional probability cont.

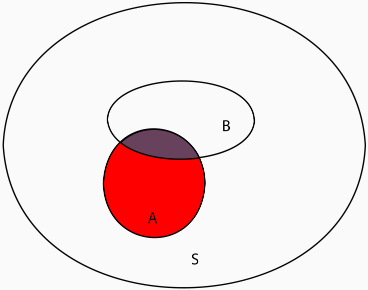


The probability of event B = $\Pr\{B\}$ = (area of B) / (area of S), but if we know that 'A' happened...

# Conditional probability cont.

The probability of event B = Pr{B} = (area of B) / (area of S), but if we know that ‘A’ happened...

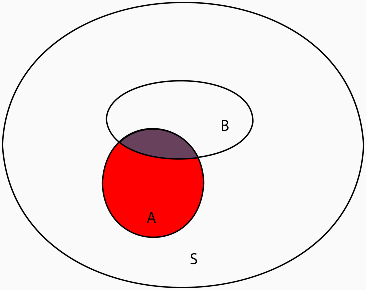
Pr{B given that A occurred} = (area common to A and B) / (area of A)



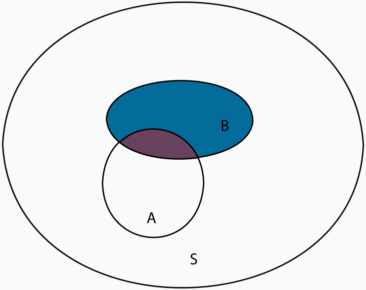

or more formally, Pr{B|A} = Pr{A,B}/Pr{A}

# Conditional probability cont.

$\Pr\{B|A\} = \Pr\{A,B\}/\Pr\{A\}$

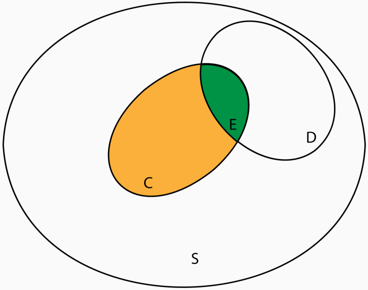



$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\}$



# Clicker question: ii






How would you write down the probability of the green event occurring, given the yellow event was observed?

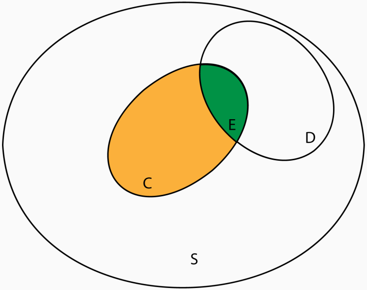
**A** — $\Pr\{C|D\} = \Pr\{E\}/\Pr\{S\}$

**B** — $\Pr\{D|C\} = \Pr\{C,D\}/\Pr\{D\}$

**C** — $\Pr\{D|C\} = \Pr\{C,D\}/\Pr\{C\}$

**D** — $\Pr\{C|D\} = \Pr\{E\}/\Pr\{S\}$

# Clicker question: ii




How would you write down the probability of the green event occurring, given the yellow event was observed?

**A** — $\Pr\{C|D\} = \Pr\{E\}/\Pr\{S\}$

**B** — $\Pr\{D|C\} = \Pr\{C,D\}/\Pr\{D\}$

**C** — $\Pr\{D|C\} = \Pr\{C,D\}/\Pr\{C\}$

**D** — $\Pr\{C|D\} = \Pr\{E\}/\Pr\{S\}$

# Clicker question: iii



Source: Wikipedia

In populations of spirit bears (*Ursus americanus kermodei*), the percentage of animals that are both female and white is $\sim 7.5\%$. If I see a female spirit bear, what is the chance that it is white?

**A** — 7.5%

**B** — 50%

**C** — 15%

**D** — 3.75%

# Clicker question: iii



Source: Wikipedia

In populations of spirit bears (*Ursus americanus kermodei*), the percentage of animals that are both female and white is $\sim 7.5\%$. If I see a female spirit bear, what is the chance that it is white?

**A** — 7.5%

**B** — 50%

**C** — 15% $\frac{0.075}{0.5} \times 100$

**D** — 3.75%

# Independent events



Events are independent when knowing that one occurs does nothing to change our idea about the probability of another event occurring.

# Independent events



Events are independent when knowing that one occurs does nothing to change our idea about the probability of another event occurring.

If two events are independent of one another, then

$$\Pr\{A|B\} = \Pr\{A\} \qquad \text{and} \qquad \Pr\{B|A\} = \Pr\{B\}$$

# Independent events



Events are independent when knowing that one occurs does nothing to change our idea about the probability of another event occurring.

If two events are independent of one another, then

$$\Pr\{A|B\} = \Pr\{A\} \qquad \text{and} \qquad \Pr\{B|A\} = \Pr\{B\}$$

And if we remember that

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\} \qquad \text{and} \qquad \Pr\{B|A\} = \Pr\{A,B\}/\Pr\{A\}\}$$

# Independent events



Events are independent when knowing that one occurs does nothing to change our idea about the probability of another event occurring.

If two events are independent of one another, then

$$\Pr\{A|B\} = \Pr\{A\} \qquad \text{and} \qquad \Pr\{B|A\} = \Pr\{B\}$$

And if we remember that

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\} \qquad \text{and} \qquad \Pr\{B|A\} = \Pr\{A,B\}/\Pr\{A\}\}$$

then

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\}$$
$$\Pr\{A\} = \Pr\{A,B\}/\Pr\{B\}$$
$$\Pr\{A,B\} = \Pr\{A\}\Pr\{B\}$$

# Joint Probability

# Joint Probability



Pr{A,B} is referred to as the joint probability of A and B.

# Joint Probability



$\Pr\{A,B\}$ is referred to as the joint probability of A and B.

You'll often see this written as $\Pr\{A \cap B\}$

# Joint Probability



$\Pr\{A,B\}$ is referred to as the joint probability of A and B.

You'll often see this written as $\Pr\{A \cap B\}$

The $\cap$ symbol refers to the 'intersection' of A and B.

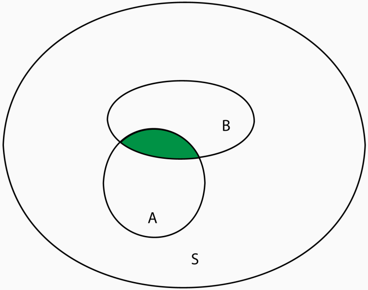
# Joint Probability



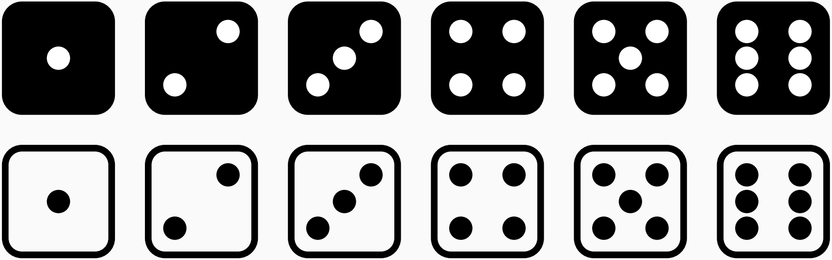
Pr{A,B} is referred to as the joint probability of A and B.

You'll often see this written as Pr{A $\cap$ B}

The $\cap$ symbol refers to the 'intersection' of A and B.



# Clicker question: iv



If I roll a pair of six sided die, what's $\Pr\{6 \cap 6\}$?

**A** — $\Pr\{6 \cap 6\} = 2/6$

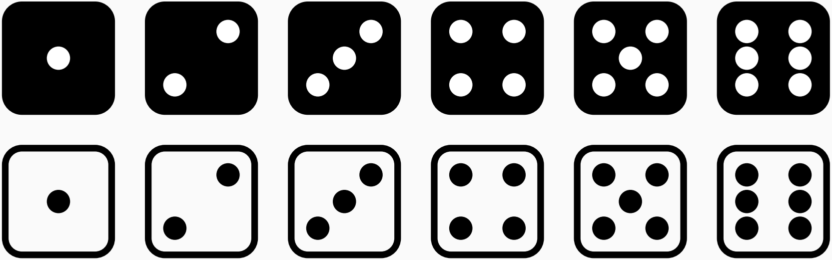
**B** — $\Pr\{6 \cap 6\} = 1/12$

**C** — $\Pr\{6 \cap 6\} = 1/36$

**D** — $\Pr\{6 \cap 6\} = 1/6$

# Clicker question: iv





If I roll a pair of six sided die, what's $\Pr\{6 \cap 6\}$?

**A** — $\Pr\{6 \cap 6\} = 2/6$

**B** — $\Pr\{6 \cap 6\} = 1/12$

**C** — $\Pr\{6 \cap 6\} = 1/36 = 1/6 \times 1/6$

**D** — $\Pr\{6 \cap 6\} = 1/6$

# Bayes' Theorem

# Bayes' Theorem

# Bayes' Theorem

Wikipedia: In probability theory and statistics, Bayes' theorem (alternatively Bayes's theorem, Bayes's law or Bayes's rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

# Bayes' Theorem



Wikipedia: In probability theory and statistics, Bayes' theorem (alternatively Bayes's theorem, Bayes's law or Bayes's rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

But what does that actually mean?

# Bayes' Theorem Cont.



From earlier, we had:

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\}$$

# Bayes' Theorem Cont.



From earlier, we had:

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\}$$

Rearranging this, we get:

$$\Pr\{A,B\} = \Pr\{A|B\}\ \Pr\{B\}$$

# Bayes' Theorem Cont.



From earlier, we had:

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\}$$

Rearranging this, we get:

$$\Pr\{A,B\} = \Pr\{A|B\}\ \Pr\{B\}$$

Remembering that the probability of B given A is given by:

$$\Pr\{B|A\} = \Pr\{A,B\}/\Pr\{A\}$$

From earlier, we had:

$$\Pr\{A|B\} = \Pr\{A,B\}/\Pr\{B\}$$

Rearranging this, we get:

$$\Pr\{A,B\} = \Pr\{A|B\} \Pr\{B\}$$

Remembering that the probability of B given A is given by:

$$\Pr\{B|A\} = \Pr\{A,B\}/\Pr\{A\}$$

We can get to:

$$\Pr\{B|A\} = \Pr\{A|B\} \Pr\{B\}/\Pr\{A\}$$

# Bayes' Theorem Cont. Cont.



The mathematical description of Bayes' Theorem is given as:

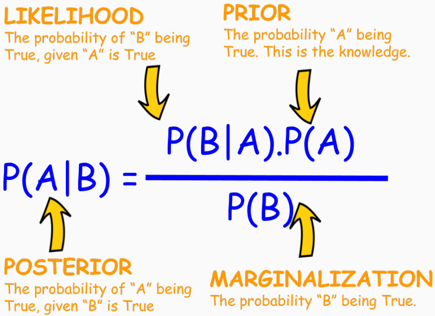
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

# Bayes' Theorem Cont. Cont.



The mathematical description of Bayes' Theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



# Bayes' Theorem Multiple Outcomes

# Bayes' Theorem Multiple Outcomes

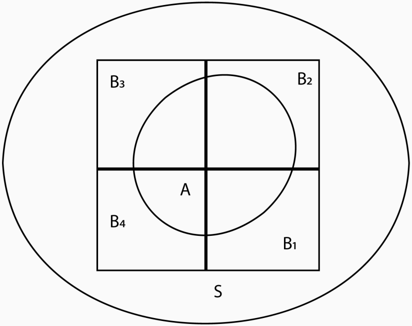


Bayes' theorem is most useful when there are multiple, exclusive possible outcomes, $B_1$,$B_2$... $B_N$

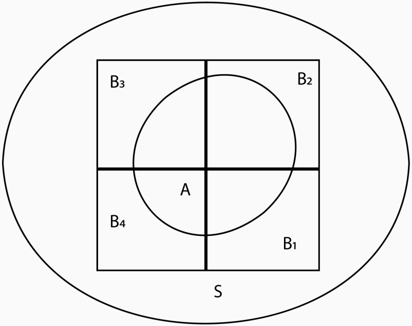
# Bayes' Theorem Multiple Outcomes



Bayes' theorem is most useful when there are multiple, exclusive possible outcomes, $B_1$,$B_2$... $B_N$, and one must occur when A occurs.



Bayes' theorem is most useful when there are multiple, exclusive possible outcomes, $B_1, B_2 \ldots B_N$, and one must occur when A occurs.

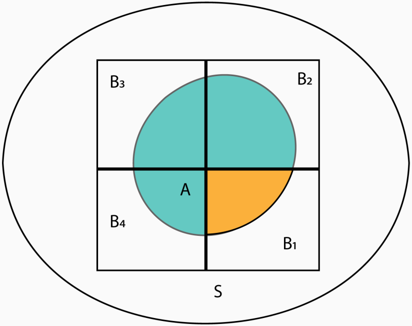



$$Pr\{B_i|A\} = \frac{Pr\{A|B_i\}Pr\{B_i\}}{\sum_{j=1}^{N} Pr\{A|B_j\}Pr\{B_j\}}$$

# Bayes' Theorem Multiple Outcomes



$$Pr\{B_1|A\} = \frac{Pr\{A|B_1\}Pr\{B_1\}}{\sum_{j=1}^{4} Pr\{A|B_j\}Pr\{B_j\}}$$



# Bayes' Theorem in Action

# Bayes' Theorem in Action

Question: I've flipped two coins, and I tell you 1 came up heads. What's the probability the other flip was heads?

# Bayes' Theorem in Action



Question: I've flipped two coins, and I tell you 1 came up heads. What's the probability the other flip was heads?

Approach #1: If each coin flip is independent, and heads/tails are equally probable, then:

$$\mathrm{Pr}\{\mathrm{Heads}_2 \mid \mathrm{Heads}_1\} = \mathrm{Pr}\{\mathrm{Heads}_2\} = 1/2$$

# Bayes' Theorem in Action



Question: I've flipped two coins, and I tell you 1 came up heads. What's the probability the other flip was heads?

Approach #1: If each coin flip is independent, and heads/tails are equally probable, then:

$$\Pr\{\text{Heads}_2 \mid \text{Heads}_1\} = \Pr\{\text{Heads}_2\} = 1/2$$

Approach #2: There are 4 possible outcomes: {HH, HT, TH, TT}. If 1 flip is heads, TT is impossible. If each combination is equally likely, then:

$$\Pr\{\text{HH}\} = 1/3$$

Question: I've flipped two coins, and I tell you 1 came up heads. What's the probability the other flip was heads?

Approach #1: If each coin flip is independent, and heads/tails are equally probable, then:

$$\Pr\{\mathrm{Heads}_2 \mid \mathrm{Heads}_1\} = \Pr\{\mathrm{Heads}_2\} = 1/2$$

Approach #2: There are 4 possible outcomes: {HH, HT, TH, TT}. If 1 flip is heads, TT is impossible. If each combination is equally likely, then:

$$\Pr\{\mathrm{HH}\} = 1/3$$

Both approaches make intuitive sense

# Bayes' Theorem in Action

Question: I've flipped two coins, and I tell you 1 came up heads. What's the probability the other flip was heads?

Approach #1: If each coin flip is independent, and heads/tails are equally probable, then:

$$\Pr\{\text{Heads}_2 \mid \text{Heads}_1\} = \Pr\{\text{Heads}_2\} = 1/2$$

Approach #2: There are 4 possible outcomes: {HH, HT, TH, TT}. If 1 flip is heads, TT is impossible. If each combination is equally likely, then:

$$\Pr\{\text{HH}\} = 1/3$$

Both approaches make intuitive sense, but both can't be right.

# Bayes' Theorem in Action

The challenge is how to use the information I've given to you that 1 flip is heads.

The challenge is how to use the information I've given to you that 1 flip is heads.

What we want to know is:

$$\Pr\{\text{HH} \mid \text{knowing one flip is H}\}$$

# Bayes' Theorem in Action



The challenge is how to use the information I've given to you that 1 flip is heads.

What we want to know is:

$$\text{Pr}\{\text{HH} \mid \text{knowing one flip is H}\} = \frac{PR\{\text{HH, knowing one flip is H}\}}{PR\{\text{knowing one flip is H}\}}$$

The challenge is how to use the information I've given to you that 1 flip is heads.

What we want to know is:

$$\text{Pr}\{\text{HH} \mid \text{knowing one flip is H}\} = \frac{PR\{\text{HH, knowing one flip is H}\}}{PR\{\text{knowing one flip is H}\}}$$

Allowing all 4 sets of possible outcomes, we have:

| **Flip Results** | **Prior probability** | **Pr{H given flip results}** |
|---|---|---|
| HH | 1/4 | 1 |
| HT | 1/4 | 1/2 |
| TH | 1/4 | 1/2 |
| TT | 1/4 | 0 |

# Bayes' Theorem in Action Cont.

| Flip Results | Prior probability | Pr{H given flip results} |
|---|---|---|
| HH | 1/4 | 1 |
| HT | 1/4 | 1/2 |
| TH | 1/4 | 1/2 |
| TT | 1/4 | 0 |

# Bayes' Theorem in Action Cont.



| Flip Results | Prior probability | Pr{H given flip results} |
|---|---|---|
| HH | 1/4 | 1 |
| HT | 1/4 | 1/2 |
| TH | 1/4 | 1/2 |
| TT | 1/4 | 0 |

Next we need to calculate the joint probability of each outcome and you knowing I flipped 1 heads:

| Flip Results | Prior probability | Pr{H given flip results} |
|---|---|---|
| HH | 1/4 | 1 |
| HT | 1/4 | 1/2 |
| TH | 1/4 | 1/2 |
| TT | 1/4 | 0 |

Next we need to calculate the joint probability of each outcome and you knowing I flipped 1 heads:

$$\text{Pr}\{\text{HH, H}\} = \text{Pr}\{\text{HH}\}\ \text{Pr}\{\text{H given flip result}\} = 1/4 \times 1 = 1/4$$

# Bayes' Theorem in Action Cont.



| Flip Results | Prior probability | Pr{H given flip results} |
|---|---|---|
| HH | 1/4 | 1 |
| HT | 1/4 | 1/2 |
| TH | 1/4 | 1/2 |
| TT | 1/4 | 0 |

Next we need to calculate the joint probability of each outcome and you knowing I flipped 1 heads:

$$\Pr\{\text{HH, H}\} = \Pr\{\text{HH}\}\Pr\{\text{H given flip result}\} = 1/4 \times 1 = 1/4$$

$$\Pr\{\text{HT, H}\} = 1/4 \times 1/2 = 1/8$$

$$\Pr\{\text{TH, H}\} = 1/4 \times 1/2 = 1/8$$

$$\Pr\{\text{TT, H}\} = 1/4 \times 0 = 0$$

# Bayes' Theorem in Action Cont.



| Flip Results | Prior probability | Pr{H given flip results} |
|---|---|---|
| HH | 1/4 | 1 |
| HT | 1/4 | 1/2 |
| TH | 1/4 | 1/2 |
| TT | 1/4 | 0 |

Next we need to calculate the joint probability of each outcome and you knowing I flipped 1 heads:

$$\Pr\{\text{HH, H}\} = \Pr\{\text{HH}\}\,\Pr\{\text{H given flip result}\} = 1/4 \times 1 = 1/4$$

$$\Pr\{\text{HT, H}\} = 1/4 \times 1/2 = 1/8$$

$$\Pr\{\text{TH, H}\} = 1/4 \times 1/2 = 1/8$$

$$\Pr\{\text{TT, H}\} = 1/4 \times 0 = 0$$

So $\Pr\{\text{of knowing 1 flip is heads}\} = 1/4 + 1/8 + 1/8 = 1/2$

# Bayes' Theorem in Action Cont. Cont.



$$\Pr\{\text{HH} \mid \text{knowing one flip is H}\} = \frac{\Pr\{\text{HH, knowing one flip is H}\}}{\Pr\{\text{knowing one flip is H}\}}$$

# Bayes' Theorem in Action Cont. Cont.



$$\text{Pr}\{\text{HH} \mid \text{knowing one flip is H}\} = \frac{\text{Pr}\{\text{HH, knowing one flip is H}\}}{\text{Pr}\{\text{knowing one flip is H}\}}$$

$$\text{Pr}\{\text{HH, knowing one flip is H}\} = 1/4$$

$$\text{Pr}\{\text{of knowing 1 flip is H}\} = 1/2$$

# Bayes' Theorem in Action Cont. Cont.



$$\Pr\{\text{HH} \mid \text{knowing one flip is H}\} = \frac{\Pr\{\text{HH, knowing one flip is H}\}}{\Pr\{\text{knowing one flip is H}\}}$$

$$\Pr\{\text{HH, knowing one flip is H}\} = 1/4$$

$$\Pr\{\text{of knowing 1 flip is H}\} = 1/2$$

$$\Pr\{\text{HH} \mid \text{knowing one flip is H}\} = \frac{1/4}{1/2} = 1/2$$

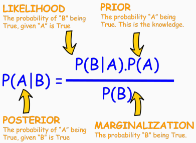
So, our first approach from earlier was correct.

# A Note on Bayes' Theorem

# A Note on Bayes' Theorem



You'll often see people argue that the strength of Bayesian methods is the ability to make use of 'prior' information (e.g., previously collected data).

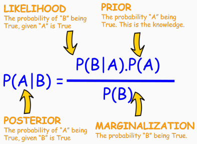



$$Pr\{B_i|A\} = \frac{Pr\{A|B_i\}Pr\{B_i\}}{\sum_{j=1}^{N} Pr\{A|B_j\}Pr\{B_j\}}$$

# A Note on Bayes' Theorem



You'll often see people argue that the strength of Bayesian methods is the ability to make use of 'prior' information (e.g., previously collected data).



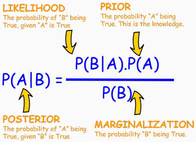

$$Pr\{B_i|A\} = \frac{Pr\{A|B_i\}Pr\{B_i\}}{\sum_{j=1}^{N} Pr\{A|B_j\}Pr\{B_j\}}$$

But in practice, people usually calculate the prior based on some existing dataset, so you could skip this if you just used the original data.

# A Note on Bayes' Theorem



You'll often see people argue that the strength of Bayesian methods is the ability to make use of 'prior' information (e.g., previously collected data).




$$Pr\{B_i|A\} = \frac{Pr\{A|B_i\}Pr\{B_i\}}{\sum_{j=1}^{N} Pr\{A|B_j\}Pr\{B_j\}}$$

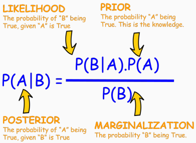
But in practice, people usually calculate the prior based on some existing dataset, so you could skip this if you just used the original data.

A lot of the time you will also see people using 'flat uninformative prior', which means the prior isn't really doing anything meaningful.

# A Note on Bayes' Theorem



You'll often see people argue that the strength of Bayesian methods is the ability to make use of 'prior' information (e.g., previously collected data).




$$Pr\{B_i|A\} = \frac{Pr\{A|B_i\}Pr\{B_i\}}{\sum_{j=1}^{N} Pr\{A|B_j\}Pr\{B_j\}}$$

But in practice, people usually calculate the prior based on some existing dataset, so you could skip this if you just used the original data.

A lot of the time you will also see people using 'flat uninformative prior', which means the prior isn't really doing anything meaningful.

The biggest benefit (in my opinion) comes from being able to use computer algorithms to calculate the denominator (marginal).

# Probability Distributions

# Probability Distributions

We're going to finish by briefly reviewing a number of commonly used probability distributions.

We're going to finish by briefly reviewing a number of commonly used probability distributions.

This list is not exhaustive, but it should be sufficient for allowing you to calculate $\Pr\{\text{model}|\text{data}\}$ and $\Pr\{\text{data}|\text{model}\}$ for many ecological scenarios.

We're going to finish by briefly reviewing a number of commonly used probability distributions.

This list is not exhaustive, but it should be sufficient for allowing you to calculate $\Pr\{\text{model}|\text{data}\}$ and $\Pr\{\text{data}|\text{model}\}$ for many ecological scenarios.

You **do not** need to memorise the formulae, but you should be able to recognise them, and understand their basic properties and use cases.

# Binomial distribution



The binomial distribution describes the probability of obtaining $k$ yes/no successes in a sample of size $n$, or in other words, the distribution of the number of successful trials among a defined number of trials.

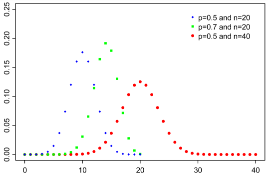
**Parameters**: $n$ and $p$

**Type**: Discrete

**Biological scenarios**: Mark recapture data, live vs dead survival data, killed by a predator or not, yes/no behavioural outcomes, anything with a discrete yes/no outcome.

**PMF**: $\binom{n}{k} p^k (1-p)^{n-k}$
where
$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

**Range**: discrete $(0 \leq x \leq n)$
**Mean**: $np$
**Variance**: $np(1-p)$




Source: Wikipedia

The Poisson distribution describes the probability of a given number of events occurring in a fixed interval of time or space.

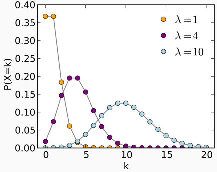
**Parameters**: $\lambda$

**Type**: Discrete

**Biological scenarios**: Counts of a species per unit time, the number of mutations on a strand of DNA per unit length, number of births/deaths per year in a given age group, prey caught per unit time.

**PMF**: $\Pr(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

**Range**: discrete $(0, \infty)$
**Mean**: $\lambda$
**Variance**: $\lambda$




Source: Wikipedia

# Negative binomial distribution

The negative binomial distribution describes the number of *failures* in a sequence of independent and identically distributed trials.

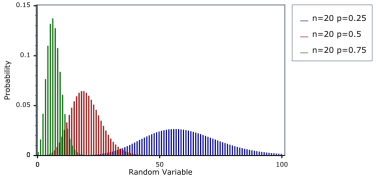
**Parameters**: $p$ Probability per trial, $k$ Overdispersion parameter

**Type**: Discrete

**Biological scenarios**: Same as the Poisson distribution, but allowing for more heterogeneity because variance $\neq$ mean.

**PMF**: $\frac{\Gamma(k+r)}{k!\cdot\Gamma(r)} p^k (1-p)^r$

**Range**: discrete $(x \geq 0)$
**Mean**: $\frac{pr}{1-p}$
**Variance**: $\frac{pr}{(1-p)^2}$



# Gaussian distribution

The Gaussian (or normal) distribution is a continuous, symmetrical distribution that applies frequently in practice.

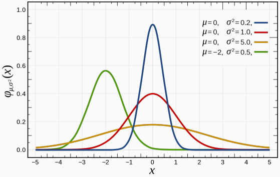
**Parameters**: $\mu$ and $\sigma$

**PDF**: $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

**Type**: Continuous

**Range**: $(-\infty, \infty)$
**Mean**: $\mu$
**Variance**: $\sigma^2$

**Biological scenarios**: Many. Almost any measurement that is continuous and symmetrically distributed.



The log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed.

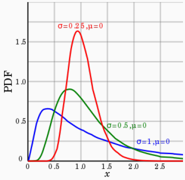
**Parameters**: $\mu$ and $\sigma$

**Type**: Continuous

**Biological scenarios**: Many continuous variables that can not take negative values (e.g., weight, height).

**PDF**: $\frac{1}{x\sigma\sqrt{2\pi}}\ \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right)$

**Range**: $(0, \infty)$
**Mean**: $\exp\left(\mu + \frac{\sigma^2}{2}\right)$
**Var**: $[\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$



The gamma distribution is a continuous probability distribution that describes waiting times until a certain number of events take place. For example a gamma distribution with shape = 3 and scale = 2 is the distribution of the length of time (in years) you'd have to wait for 3 deaths to occur in a population with an average survival time of 2 years.

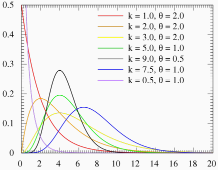
**Parameters**: shape = $k$ and scale = $\theta$ (both >0)

**Type**: Continuous

**Biological scenarios**: Survival time, the age distribution of cancer incidence, highly variable data where negative numbers don't make sense.

**PDF**: $\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$

**Range**: $(0, \infty)$
**Mean**: $k\theta$
**Var**: $k\theta^2$



What does all this have to do with fitting a straight line to some data you ask?

What does all this have to do with fitting a straight line to some data you ask?

We'll get to that next lecture...