Statistics for Astronomers: Lecture 1, 2020.09.21

Prof. Sundar Srinivasan

IRyA/UNAM

About the course

- Lectures via Zoom; notes available 10 min before each lecture.
- Programming: Python (3+).
- In-class and homework assignments each week.
- One midterm and one final exam. Written exams starting 09:00 on Day 1, ending 17:00 on Day 2, with one Zoom session to address issues, and access to me via email for the entire duration.
- **Course webpage**
  Lectures notes, homework assignments, exams will be posted here.
  In addition, it will list a variety of resources - books, online courses, conferences/workshops, and papers.

Homework 1

Already on website, due next Monday (2020.09.28) at 17:00.

Overview

Probability Theory and Statistical Inference[1]

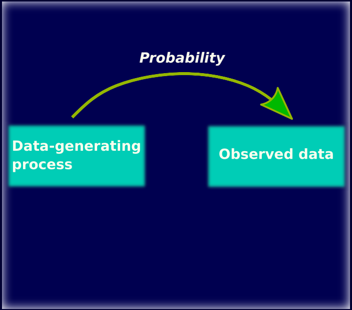
[1] Adapted from Wasserman, "All of Statistics"

Probability Theory and Statistical Inference[1]

- Given a data-generating process, what are the properties of the outcomes?

[1] Adapted from Wasserman, "All of Statistics"

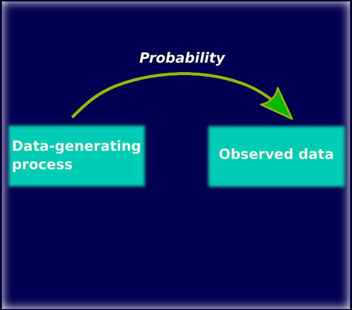
Probability Theory and Statistical Inference[1]



- Given a data-generating process, what are the properties of the outcomes? Probability theory.

[1] Adapted from Wasserman, "All of Statistics"

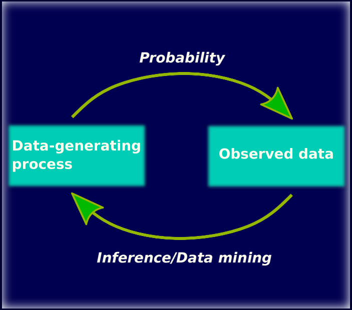
Probability Theory and Statistical Inference[1]



- Given a data-generating process, what are the properties of the outcomes? Probability theory.
- Given the outcome of an observation/experiment, what can be said about the process(es) that generated the data?

[1] Adapted from Wasserman, "All of Statistics"

Probability Theory and Statistical Inference[1]



- Given a data-generating process, what are the properties of the outcomes?
  Probability theory.
- Given the outcome of an observation/experiment, what can be said about the process(es) that generated the data?
  Statistical inference.

[1] Adapted from Wasserman, "All of Statistics"

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty.

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

[2] *e.g.*, Wall & Jenkins 2003, Feigelsen & Babu 2012.

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty.

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

- Astronomers have little to no control over their "test subjects".

[2] *e.g.*, Wall & Jenkins 2003, Feigelsen & Babu 2012.

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty.

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

- Astronomers have little to no control over their "test subjects".
- Experiments observe events with uncertain outcomes, and are affected by random error as well as uncertainty due to small samples.

[2] *e.g.*, Wall & Jenkins 2003, Feigelsen & Babu 2012.

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty.

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

- Astronomers have little to no control over their "test subjects".
- Experiments observe events with uncertain outcomes, and are affected by random error as well as uncertainty due to small samples.
- Typically either costly to repeat observations, or there are physical/instrumental constraints (e.g., resolution, weather).

[2] *e.g.*, Wall & Jenkins 2003, Feigelsen & Babu 2012.

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty.

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

- Astronomers have little to no control over their "test subjects".
- Experiments observe events with uncertain outcomes, and are affected by random error as well as uncertainty due to small samples.
- Typically either costly to repeat observations, or there are physical/instrumental constraints (e.g., resolution, weather).
- Can provide quantitative information about whether our measured interval captures the "true" value of a quantity. Can also quantitatively compare properties of two different samples.

[2] *e.g.*, Wall & Jenkins 2003, Feigelsen & Babu 2012.

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty.

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

- Astronomers have little to no control over their "test subjects".
- Experiments observe events with uncertain outcomes, and are affected by random error as well as uncertainty due to small samples.
- Typically either costly to repeat observations, or there are physical/instrumental constraints (e.g., resolution, weather).
- Can provide quantitative information about whether our measured interval captures the "true" value of a quantity. Can also quantitatively compare properties of two different samples.
- Uncertainty due to ignorance of the underlying physical processes could be reduced using past experience (e.g., "updating priors").

[2] *e.g.*, Wall & Jenkins 2003, Feigelsen & Babu 2012.

Why Must Astronomers Care About Probability?[2]

Probability quantifies uncertainty.

"Uncertain knowledge + Knowledge of the amount of uncertainty in it = Usable knowledge."
– C. R. Rao (1997)

- Astronomers have little to no control over their "test subjects".
- Experiments observe events with uncertain outcomes, and are affected by random error as well as uncertainty due to small samples.
- Typically either costly to repeat observations, or there are physical/instrumental constraints (e.g., resolution, weather).
- Can provide quantitative information about whether our measured interval captures the "true" value of a quantity. Can also quantitatively compare properties of two different samples.
- Uncertainty due to ignorance of the underlying physical processes could be reduced using past experience (e.g., "updating priors").
- Predicting exact behaviour of individual members of a class difficult, but easier to describe the class as a whole with some confidence.
  e.g.: stellar populations, globular clusters, kinetic theory of gases, radioactive decay, mortality, prevalence of a disease, the science of psychohistory in the **Foundation** series by Isaac Asimov.

[2] *e.g.*, Wall & Jenkins 2003, Feigelsen & Babu 2012.

Why Must Astronomers Care About Statistics?[3]

What is statistics?

"[T]he theory and methods of collecting, organizing, presenting, analyzing, and interpreting data sets so as to determine their essential characteristics."

– Panik (2005)

[3] see G. Djorgovsky's talk at the 2017 TIARA Summer School in Astrostatistics & Data Mining

Why Must Astronomers Care About Statistics?[3]

What is statistics?

"[T]he theory and methods of collecting, organizing, presenting, analyzing, and interpreting data sets so as to determine their essential characteristics."

– Panik (2005)

- Statistics: what to compute.
  Informatics: how to compute.

[3] see G. Djorgovsky's talk at the 2017 TIARA Summer School in Astrostatistics & Data Mining

Why Must Astronomers Care About Statistics?[3]

What is statistics?

"[T]he theory and methods of collecting, organizing, presenting, analyzing, and interpreting data sets so as to determine their essential characteristics."

– Panik (2005)

- Statistics: what to compute.
  Informatics: how to compute.
- Data-driven vs. hypothesis-driven science.

[3]see G. Djorgovsky's talk at the 2017 TIARA Summer School in Astrostatistics & Data Mining

Why Must Astronomers Care About Statistics?[3]

What is statistics?

"[T]he theory and methods of collecting, organizing, presenting, analyzing, and interpreting data sets so as to determine their essential characteristics."
– Panik (2005)

- Statistics: what to compute.
  Informatics: how to compute.
- Data-driven vs. hypothesis-driven science.
- Most data will never be seen by humans (already the case).
- Patterns in (multidimensional) data cannot be comprehended by humans directly.

[3] see G. Djorgovsky's talk at the 2017 TIARA Summer School in Astrostatistics & Data Mining

Why Must Astronomers Care About Statistics?[3]

What is statistics?

"[T]he theory and methods of collecting, organizing, presenting, analyzing, and interpreting data sets so as to determine their essential characteristics."
– Panik (2005)

- Statistics: what to compute.
  Informatics: how to compute.
- Data-driven vs. hypothesis-driven science.
- Most data will never be seen by humans (already the case).
- Patterns in (multidimensional) data cannot be comprehended by humans directly.
- **Why you should care:** Jobs! Jobs! Jobs! Data science! Big data! LSST! SKA! ngVLA! ...

[3] see G. Djorgovsky's talk at the 2017 TIARA Summer School in Astrostatistics & Data Mining

Probability

Estimating probability: an exercise

Example: In a two-coin toss, what is the probability of obtaining two tails?

Why?

Estimating probability: an exercise

Example: In a two-coin toss, what is the probability of obtaining two tails?

Why?

Two implicit assumptions:

1. All outcomes are equally likely, and
2. The sum of probabilities of all outcomes is 1.

Computing probability: three ways

- **Classical interpretation**: assume each outcome is equally likely.

  We just used this interpretation to compute the probability in the previous slide.

Computing probability: three ways

- **Classical interpretation**: assume each outcome is equally likely.

  We just used this interpretation to compute the probability in the previous slide.

- **Frequentist interpretation**: perform an infinite sequence of experiments, find relative frequency of favourable outcomes.

  Toss the two coins N times ($N \gg 1$) and count the number of times M that we get two tails. The required probability is then $\dfrac{M}{N}$.

Computing probability: three ways

- **Classical interpretation**: assume each outcome is equally likely.

  We just used this interpretation to compute the probability in the previous slide.

- **Frequentist interpretation**: perform an infinite sequence of experiments, find relative frequency of favourable outcomes.

  Toss the two coins N times ($N \gg 1$) and count the number of times M that we get two tails. The required probability is then $\dfrac{M}{N}$.

- **Bayesian interpretation**: use prior knowledge of the parameters of the problem, perform experiments, and update the priors to get posterior probabilities.

Computing probability: three ways

- **Classical interpretation**: assume each outcome is equally likely.

  We just used this interpretation to compute the probability in the previous slide.

- **Frequentist interpretation**: perform an infinite sequence of experiments, find relative frequency of favourable outcomes.

  Toss the two coins N times ($N \gg 1$) and count the number of times M that we get two tails. The required probability is then $\dfrac{M}{N}$.

- **Bayesian interpretation**: use prior knowledge of the parameters of the problem, perform experiments, and update the priors to get posterior probabilities.

  1. Select your priors (*e.g.*, are the coins known to be fair from experience? "Roberto performed the experiment 500 times yesterday and only got two tails 20 times!").

Computing probability: three ways

- **Classical interpretation**: assume each outcome is equally likely.

  We just used this interpretation to compute the probability in the previous slide.

- **Frequentist interpretation**: perform an infinite sequence of experiments, find relative frequency of favourable outcomes.

  Toss the two coins N times ($N \gg 1$) and count the number of times M that we get two tails. The required probability is then $\frac{M}{N}$.

- **Bayesian interpretation**: use prior knowledge of the parameters of the problem, perform experiments, and update the priors to get posterior probabilities.

  1. Select your priors (*e.g.*, are the coins known to be fair from experience? "Roberto performed the experiment 500 times yesterday and only got two tails 20 times!").
  2. Perform an experiment.

Computing probability: three ways

- **Classical interpretation**: assume each outcome is equally likely.

  We just used this interpretation to compute the probability in the previous slide.

- **Frequentist interpretation**: perform an infinite sequence of experiments, find relative frequency of favourable outcomes.

  Toss the two coins N times ($N \gg 1$) and count the number of times M that we get two tails. The required probability is then $\dfrac{M}{N}$.

- **Bayesian interpretation**: use prior knowledge of the parameters of the problem, perform experiments, and update the priors to get posterior probabilities.

  1. Select your priors (*e.g.*, are the coins known to be fair from experience? "Roberto performed the experiment 500 times yesterday and only got two tails 20 times!").
  2. Perform an experiment.
  3. Combine the resulting outcome with the prior and predict the probability of getting two tails on future trials. Update your prior belief about the coins.

Classical (naïve) interpretation of probability

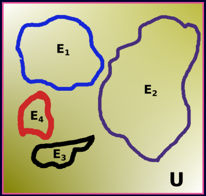
"The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible..."

– P.-S. Laplace (1812)

Principle of Indifference

If N events are mutually exclusive and collectively exhaustive,

(1) they are equally likely and (2) the probability of any one occurring is $\frac{1}{N}$.



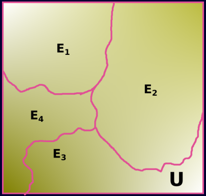


Figure: Case (1): E1, E2, E3, and E4 are mutually exclusive but not collectively exhaustive. Case (2): the events are now also collectively exhaustive.

Classical (naïve) interpretation of probability

"The probability of an event is the ratio of the number of cases favorable to it, to the number of all cases possible..."

– P.-S. Laplace (1812)

Principle of Indifference

If N events are mutually exclusive and collectively exhaustive,

(1) they are equally likely and (2) the probability of any one occurring is $\frac{1}{N}$.

Two-coin toss example

Four possible outcomes: (TT), (TH), (HT), and (HH).

The outcomes are mutually exclusive and collectively exhaustive
$\Longrightarrow$ probability of any one outcome $= 1/4$.

Classical (naïve) interpretation of probability (contd.)

Criticism of the Principle of Indifference: not all sample spaces consist of equally likely events.

Excerpt from a Daily Show segment in which John Oliver interviews a male high-school mathematics teacher about the LHC:

John: What is the probability that the Large Hadron Collider destroys the Universe?
Teacher: 50%
John: How do you figure?
Teacher: It will either destroy the Universe, or it won't.
John: Do you know how probability works?!

Axioms of Probability (Kolmogorov, 1933)

For a given problem, let Ω be the set of all possible events (the sample space).

1. The probability that an event has occurred is always a non-negative real number.

Axioms of Probability (Kolmogorov, 1933)

For a given problem, let Ω be the set of all possible events (the sample space).

1. The probability that an event has occurred is always a non-negative real number.

   in particular, $P(\varnothing) = 0$. (At least one event in Ω occurs.)

Axioms of Probability (Kolmogorov, 1933)

For a given problem, let Ω be the set of all possible events (the sample space).

1. The probability that an event has occurred is always a non-negative real number.

   in particular, $P(\varnothing) = 0$. (At least one event in Ω occurs.)

2. **Unitarity:** The probability that at least one event in the sample space will occur is unity.
   $P(\Omega) = 1$

Axioms of Probability (Kolmogorov, 1933)

For a given problem, let Ω be the set of all possible events (the sample space).

1. The probability that an event has occurred is always a non-negative real number.

   in particular, $P(\varnothing) = 0$. (At least one event in Ω occurs.)

2. **Unitarity:** The probability that at least one event in the sample space will occur is unity.
   $P(\Omega) = 1$

3. **Countable additivity:** The probability that at least one event among a set of (pairwise) disjoint events occurs is the sum of the probabilities of each of those events occurring.

   Given A_j $(j = 1, \ldots)$ such that $A_j \cap A_k = \varnothing$ for $j \neq k$,

   $$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j), \text{ if } A_j \cap A_k = \varnothing \text{ for } j \neq k$$

Some properties derived from the axioms

For two events $A, B \in \Omega$,

- $P(A) + P(A^c) = 1$ (either A occurs, or doesn't).
- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $A \subseteq B \Rightarrow P(B) = P(B \cap A) + P(B \cap A^c)$
- $A \cup B = A + B - A \cap B$, which can be used to derive the "Inclusion-Exclusion Principle": $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Inclusion-Exclusion Principle (generalised):

$$P\left(\bigcup_{j=0}^{n} A_j\right) = \sum_{j=0}^{n} P(A_j) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots$$
$$+(-1)^{n+1} P\left(\bigcap_{j=1}^{n} A_j\right)$$

Conditional probabilities and Bayes' Theorem

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.

Examples for a two-coin flip:

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.

Examples for a two-coin flip:

1. $A =$ "Second coin lands heads.", $B =$ "First coin lands heads."
   $\implies A \cap B =$ "Get two heads."

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.

Examples for a two-coin flip:

1. $A =$ "Second coin lands heads.", $B =$ "First coin lands heads."
   $\implies A \cap B =$ "Get two heads."
2. $A =$ "Get two heads.", $B =$ "Get two tails."
   $\implies A \cap B = \varnothing$.

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.

Examples for a two-coin flip:

1. $A =$ "Second coin lands heads.", $B =$ "First coin lands heads."
   $\implies A \cap B =$ "Get two heads."
2. $A =$ "Get two heads.", $B =$ "Get two tails."
   $\implies A \cap B = \varnothing$.

Definition (Conditional probability)

The probability that an event A occurs, given that another event B has already occurred.
Representation: $P(A|B)$ ("probability of A given B").

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.

Examples for a two-coin flip:

1. $A =$ "Second coin lands heads.", $B =$ "First coin lands heads."
   $\implies A \cap B =$ "Get two heads."
2. $A =$ "Get two heads.", $B =$ "Get two tails."
   $\implies A \cap B = \varnothing$.

Definition (Conditional probability)

The probability that an event A occurs, given that another event B has already occurred.
Representation: $P(A|B)$ ("probability of A given B").

Example 1: $A|B =$ "Second coin lands heads, given that first coin landed heads."
Example 2: $A|B =$ "Get two heads, given that we get two tails." (contradiction).

Conditional probability

If A and B are two events, $A \cap B$ represents the event that both A and B occur.

Examples for a two-coin flip:

1. $A =$ "Second coin lands heads.", $B =$ "First coin lands heads."
   $\Longrightarrow A \cap B =$ "Get two heads."
2. $A =$ "Get two heads.", $B =$ "Get two tails."
   $\Longrightarrow A \cap B = \varnothing$.

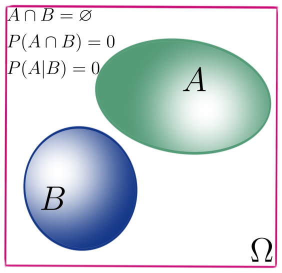
Definition (Conditional probability)

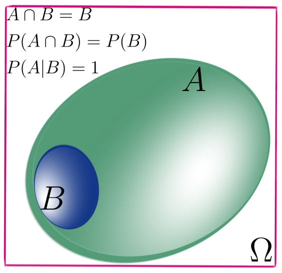
The probability that an event A occurs, given that another event B has already occurred.
Representation: $P(A|B)$ ("probability of A given B").

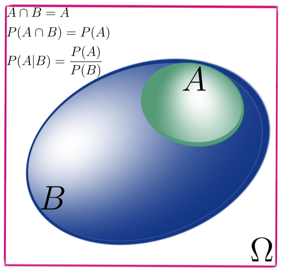
Example 1: $A|B =$ "Second coin lands heads, given that first coin landed heads."
Example 2: $A|B =$ "Get two heads, given that we get two tails." (contradiction).

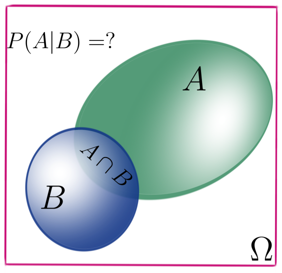
In general, the events $A \cap B$ and $A|B$ are related.
(given that B has occurred, A can only occur if $A \cap B \neq \varnothing$).

Conditional probability - 2









Conditional probability - 3

Let $P(A \cap B) = P(A|B) \times \lambda$, for some constant λ.

Conditional probability - 3

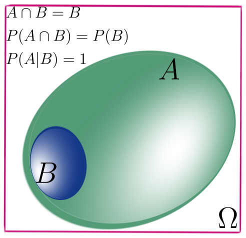
Let $P(A \cap B) = P(A|B) \times \lambda$, for some constant λ.

To solve for λ, consider the special case $B \subseteq A$. Then,

(1) A always occurs if B occurs $\Rightarrow P(A|B) = 1$ in this case.

(2) $A \cap B = B \implies P(A \cap B) = P(B)$ in this case.

From (1) and (2), $\lambda = P(B)$.



Conditional probability - 3

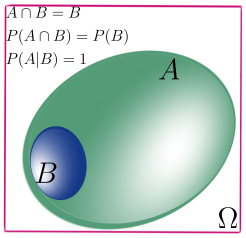
Let $P(A \cap B) = P(A|B) \times \lambda$, for some constant λ.

To solve for λ, consider the special case $B \subseteq A$. Then,

(1) A always occurs if B occurs $\Rightarrow P(A|B) = 1$ in this case.

(2) $A \cap B = B \Longrightarrow P(A \cap B) = P(B)$ in this case.

From (1) and (2), $\lambda = P(B)$.



Since λ is a constant, this is valid for all cases

$\Longrightarrow P(A \cap B) = P(A|B) \times P(B)$.

Conditional probability - 3

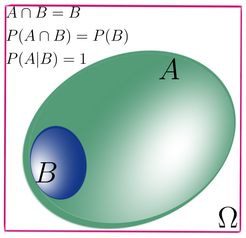
Let $P(A \cap B) = P(A|B) \times \lambda$, for some constant λ.

To solve for λ, consider the special case $B \subseteq A$. Then,

(1) A always occurs if B occurs $\Rightarrow P(A|B) = 1$ in this case.

(2) $A \cap B = B \implies P(A \cap B) = P(B)$ in this case.

From (1) and (2), $\lambda = P(B)$.



Since λ is a constant, this is valid for all cases

$\implies P(A \cap B) = P(A|B) \times P(B)$.

Conditional probability

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

$$P(A \text{ occurs, given that } B \text{ already occurred}) = \frac{P(A \text{ and } B \text{ both occur})}{P(B \text{ occurs})}$$

Note: From the axioms of probability (unitarity), $P(A|B) + P(A^c|B) = 1$

$\implies P(B \cap A) = P(B|A) \times P(A)$.

Conditional probability - 3

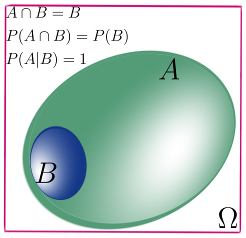
Let $P(A \cap B) = P(A|B) \times \lambda$, for some constant λ.

To solve for λ, consider the special case $B \subseteq A$. Then,

(1) A always occurs if B occurs $\Rightarrow P(A|B) = 1$ in this case.

(2) $A \cap B = B \Longrightarrow P(A \cap B) = P(B)$ in this case.

From (1) and (2), $\lambda = P(B)$.



Since λ is a constant, this is valid for all cases

$\Longrightarrow P(A \cap B) = P(A|B) \times P(B)$.

Conditional probability

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

$$P(A \text{ occurs, given that } B \text{ already occurred}) = \frac{P(A \text{ and } B \text{ both occur})}{P(B \text{ occurs})}$$

Note: From the axioms of probability (unitarity), $P(A|B) + P(A^c|B) = 1$

$\Longrightarrow P(B \cap A) = P(B|A) \times P(A)$.

Since $P(B \cap A) = P(A \cap B)$, this means $P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A)$.

Bayes' Theorem

Recall: $P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A)$

Bayes' Theorem

Recall: $P(A \cap B) = P(A|B) \times P(B) = P(B|A) \times P(A)$

Definition (Bayes' Theorem)

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

Under the Bayesian Interpretation of probability, this is read as

$$\text{Updated deg. of belief in } A = \text{Support for } A \text{ from evidence } B \times \text{Original deg. of belief in } A.$$

or

$$\text{Posterior prob. of } A \text{ given evidence } B = \frac{\text{Cond. prob. of } B \text{ given } A}{\text{Marginal prob. of } B} \times \text{Prior prob. of} A.$$

or

Posterior prob. of A given evidence B $= \dfrac{\text{Likelihood of } A \text{ given } B}{\text{Evidence } B} \times \text{Prior prob. of } A.$

Independence

Two events A and B are said to be independent ("$A \perp B$") if the occurrence of one does not affect the probability of occurrence of the other.

Independence

Two events A and B are said to be independent ("$A \perp B$") if the occurrence of one does not affect the probability of occurrence of the other.

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Independence

Two events A and B are said to be independent ("$A \perp B$") if the occurrence of one does not affect the probability of occurrence of the other.

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.
Therefore, from the definition of conditional probability,

Definition (Independence)

$A \perp B \Rightarrow P(A \cap B) = P(A) \times P(B)$.

Independence

Two events A and B are said to be independent ("$A \perp B$") if the occurrence of one does not affect the probability of occurrence of the other.

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.
Therefore, from the definition of conditional probability,

Definition (Independence)

$A \perp B \Rightarrow P(A \cap B) = P(A) \times P(B)$.

Example (Two-coin toss)

Let A = "Second coin turns up heads" and B = "First coin turns up heads".

If both coins are fair, then the outcome of flipping the second coin should not depend on the outcome of flipping the first one.

$\Rightarrow P(A|B) = P(A) = 1/2$.
$\Rightarrow P(\text{two heads}) = P(A|B) \times P(B) = P(A) \times P(B) = 1/4$.

Independence

Two events A and B are said to be independent ("$A \perp B$") if the occurrence of one does not affect the probability of occurrence of the other.

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.
Therefore, from the definition of conditional probability,

Definition (Independence)

$A \perp B \Rightarrow P(A \cap B) = P(A) \times P(B)$.

Example (Two-coin toss)

Let $A =$ "Second coin turns up heads" and $B =$ "First coin turns up heads".

If both coins are fair, then the outcome of flipping the second coin should not depend on the outcome of flipping the first one.

$\Rightarrow P(A|B) = P(A) = 1/2$.

$\Rightarrow P(\text{two heads}) = P(A|B) \times P(B) = P(A) \times P(B) = 1/4$.

Are two mutually exclusive events mutually independent?

Independence

Two events A and B are said to be independent ("$A \perp B$") if the occurrence of one does not affect the probability of occurrence of the other.

If A and B are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.
Therefore, from the definition of conditional probability,

Definition (Independence)

$A \perp B \Rightarrow P(A \cap B) = P(A) \times P(B)$.

Example (Two-coin toss)

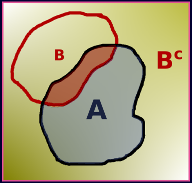
Let $A =$ "Second coin turns up heads" and $B =$ "First coin turns up heads".

If both coins are fair, then the outcome of flipping the second coin should not depend on the outcome of flipping the first one.

$\Rightarrow P(A|B) = P(A) = 1/2$.
$\Rightarrow P(\text{two heads}) = P(A|B) \times P(B) = P(A) \times P(B) = 1/4$.

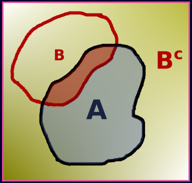
Are two mutually exclusive events mutually independent?
Mutual exclusivity $\neq$ mutual independence!!

Conditionality and Marginalisation



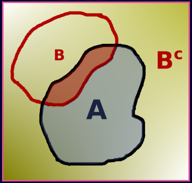
In general, $A = (A \cap B) \cup (A \cap B^c)$.

Conditionality and Marginalisation



In general, $A = (A \cap B) \cup (A \cap B^c)$. Using the Inclusion-Exclusion Principle, $P(A) = P(A \cap B) + P(A \cap B^c) - P((A \cap B) \cap (A \cap B^c))$.

Conditionality and Marginalisation

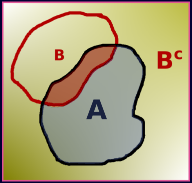


In general, $A = (A \cap B) \cup (A \cap B^c)$. Using the Inclusion-Exclusion Principle, $P(A) = P(A \cap B) + P(A \cap B^c) - P((A \cap B) \cap (A \cap B^c))$.

Using conditional probabilities, $P(A) = P(A|B) \times P(B) + P(A|B^c) \times P(B^c)$. $P(A)$ is obtained by "marginalising over B".

Do not confuse with the result from unitarity: $P(A|B) + P(A^c|B) = 1$.

Conditionality and Marginalisation

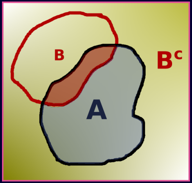


In general, $A = (A \cap B) \cup (A \cap B^c)$. Using the Inclusion-Exclusion Principle, $P(A) = P(A \cap B) + P(A \cap B^c) - P((A \cap B) \cap (A \cap B^c))$.

Using conditional probabilities, $P(A) = P(A|B) \times P(B) + P(A|B^c) \times P(B^c)$. $P(A)$ is obtained by "marginalising over B".

Do not confuse with the result from unitarity: $P(A|B) + P(A^c|B) = 1$.

Conditionality and Marginalisation



In general, $A = (A \cap B) \cup (A \cap B^c)$. Using the Inclusion-Exclusion Principle, $P(A) = P(A \cap B) + P(A \cap B^c) - P((A \cap B) \cap (A \cap B^c))$.

Using conditional probabilities, $P(A) = P(A|B) \times P(B) + P(A|B^c) \times P(B^c)$. $P(A)$ is obtained by "marginalising over B".

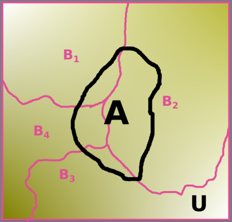
Do not confuse with the result from unitarity: $P(A|B) + P(A^c|B) = 1$.

Generalisation: Law of Total Probability

(Connects conditional probabilities to marginal probability)

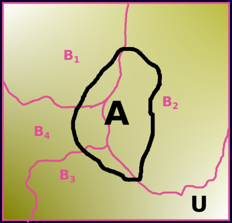
Given N pairwise disjoint and collectively exhaustive events B_i $(i = 1, 2, ..., N)$, the probability of occurrence of an event A is given as the weighted average of the conditional probabilities $P(A|B_i)$, with weights $P(B_i)$:

$$P(A) = \sum_{i=1}^{N} P(A \cap B_i) = \sum_{i=1}^{N} P(A|B_i) \times P(B_i)$$



$P(A)$ is then the probability of A marginalised over the events B_i.

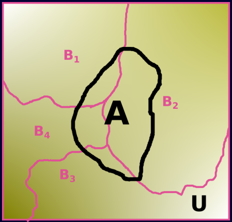
Bayes' Theorem with Marginalisation



We can use the Law of Total Probability to convert the marginal probability into conditional probabilities:

$$P(A) = \sum_{i=1}^{N} P(A|B_i) \times P(B_i) = P(A|B_j) \times P(B_j) + P(A|B_j^c) \times P(B_j^c)$$

Bayes' Theorem with Marginalisation

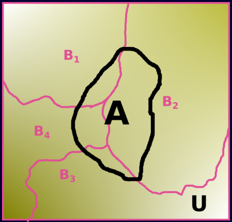


We can use the Law of Total Probability to convert the marginal probability into conditional probabilities:

$$P(A) = \sum_{i=1}^{N} P(A|B_i) \times P(B_i) = P(A|B_j) \times P(B_j) + P(A|B_j^c) \times P(B_j^c)$$

$$\implies P(B_j|A) = \frac{P(A|B_j) \times P(B_j)}{\sum_{i=1}^{N} P(A|B_i) \times P(B_i)}$$

Bayes' Theorem with Marginalisation



We can use the Law of Total Probability to convert the marginal probability into conditional probabilities:

$$P(A) = \sum_{i=1}^{N} P(A|B_i) \times P(B_i) = P(A|B_j) \times P(B_j) + P(A|B_j^c) \times P(B_j^c)$$

$$\implies P(B_j|A) = \frac{P(A|B_j) \times P(B_j)}{\sum_{i=1}^{N} P(A|B_i) \times P(B_i)}$$

$$= \frac{P(A|B_j) \times P(B_j)}{P(A|B_j) \times P(B_j) + P(A|B_j^c) \times P(B_j^c)}$$

Bayesian inference versus frequentist inference

Bayesian inference relies on assuming prior knowledge of the hypotheses/parameters of interest.

In the Bayesian interpretation, probability is a *degree of belief*.
In this sense, Bayesian probabilities are subjective.

Frequentist inference is, by contrast, considered objective as it does not incorporate/assume priors for the parameters that are the subject of inference.