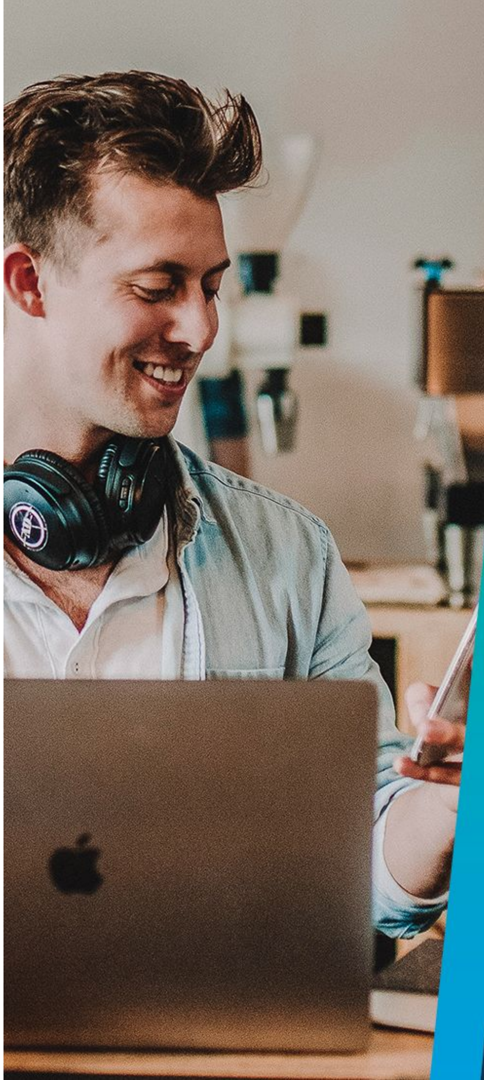


DS Workshop

Hyperiondev

Probability for Data Scientists

Welcome

Your Lecturer for this session

Sanana Mwanawina

Workshop – Housekeeping

- ❑ The use of disrespectful language is prohibited in the questions, this is a supportive, learning environment - please engage accordingly.
- ❑ No question is daft or silly - **ask them!**
- ❑ There are Q/A sessions midway and at the end of the session, should you wish to ask any follow-up questions.
- ❑ You can also submit questions here: www.hyperiondev.com/support
- ❑ For all non-academic questions, please submit a query: www.hyperiondev.com/support
- ❑ Report a safeguarding incident: hyperiondev.com/safeguardreporting
- ❑ We would love your feedback on lectures and workshops: https://hyperionde.wufoo.com/forms/zsqv4m40ui4i0q/

GitHub repo

Go to: github.com/HyperionDevBootcamps

Then click on the "**C4_DS_lecture_examples**" repository, do view or download the code.

Objectives

1. Learn the frequentist approach to probability, conditional probability, and the fundamentals of Bayes' Theorem

Probability Theory

Some terminology:

1. A random experiment is a procedure whose outcome in a particular trial cannot be predetermined
2. The set of all possible outcomes is the sample space
3. Each repetition of the steps for a random experiment is called a trial
4. An event is any subset of the sample space

Example: Tossing a fair coin.
Possible outcomes: $S = \{\text{heads, tails}\}$

Probability Theory

Definition of probability:

A measure or quantification of the likelihood of an event occurring.

$Pr(A) = f / N$

where $Pr(A)$ is the probability of event A occurring, f is the frequency (number of ways the event can occur) and N is the total number of possible outcomes.

Example: In the coin tossing experiment

$$Pr(\text{Heads}) = \tfrac{1}{2} = 0.5$$

Probability Theory

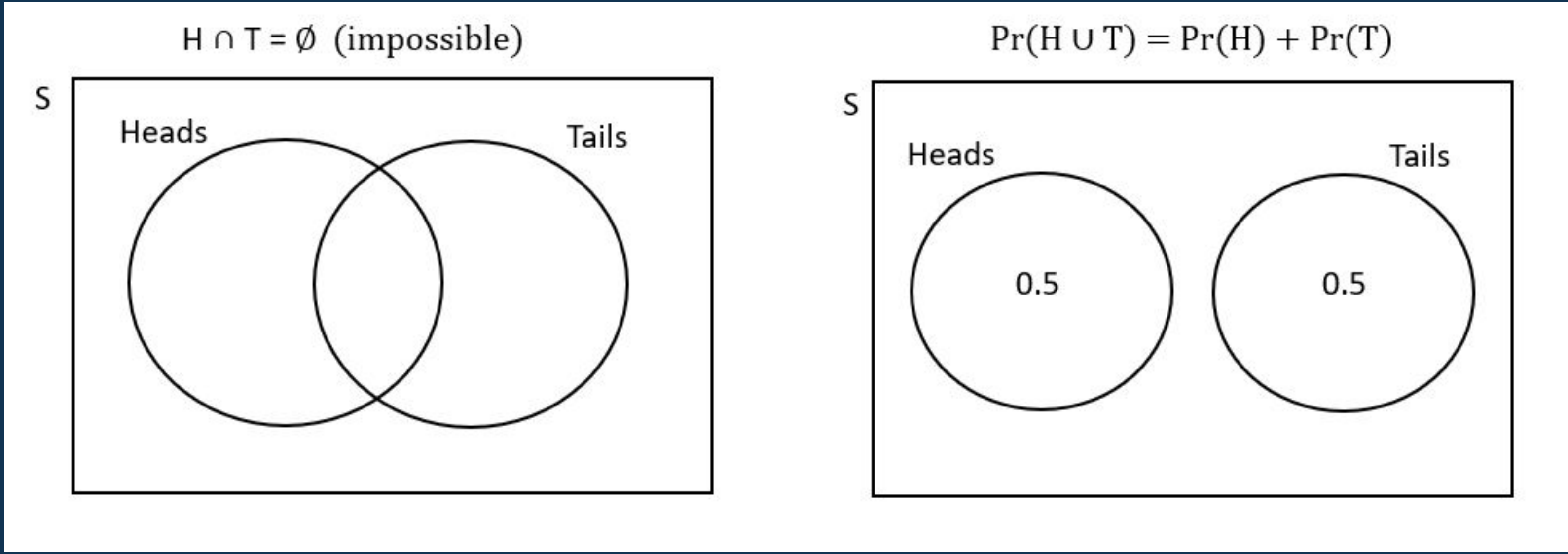
"Andrey Nikolaevich Kolmogorov was a Russian mathematician who, in 1933, published the axioms of probability, and established the theoretical foundation for the rigorous mathematical study of probability theory."

Kolmogorov's axioms of probability:

1. $0 \leq Pr(A) \leq 1$ for all $A \subset S$
2. $Pr(S) = 1$
3. If $A \cap B = \emptyset$ (i.e. if A and B are mutually exclusive events) then $Pr(A \cup B) = Pr(A) + Pr(B)$

Probability Theory

Third probability axiom using venn diagrams:



Conditional Probability

Conditional probability is a measure of the probability of an event occurring given that another event has already occurred.

Let A and B be two events in a sample space S. Then the conditional probability of the event B given that the event A has occurred, denoted by $Pr(B \mid A)$, is:

$$Pr(B \mid A) = Pr(A \cap B)/Pr(A)$$

provided that $Pr(A)$ 6= 0. $Pr(B \mid A)$ is read "the probability of B given A".

Conditional Probability

Example: You, an economist, are one of the 198 applicants applying for an M.B.A programme of whom 81 will be accepted. You hear along the grapevine that there were 70 economist applicants, of whom 38 were accepted. Assess your probabilities of being accepted before and after you receive the information.

Events of interest:

Pr(being an economist) = Pr(E)
Pr(being accepted) = Pr(A)

Conditional Probability

Before you got the inside scoop:

$Pr(A) = n(A) / n(\text{applicants}) = 81 / 198 = 0.409$

After you received the grapevine information:

$Pr(A \mid E) = Pr(A \cap E) / Pr(E) = (38 / 198) / (70 / 198) = 0.1919 / 0.3535$

$Pr(A \mid E) = 0.543$

Bayesian statistics

For any two events A and B, there are two conditional probabilities that can be considered:

$$\Pr(B \mid A) = \Pr(A \cap B)/\Pr(A)$$
$$\Pr(A \mid B) = \Pr(A \cap B)/\Pr(B)$$

A useful tool that connects the two probabilities is Bayes' theorem:

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

Bayesian statistics

Example: You wake up feeling like you are coming down with a flu. We know that 6.5% of people in your city have the flu. So, the probability that you are developing a flu, with the information we have so far, is 0.065

The symptoms get worse. So you decide to take an at-home test. This test is not perfect, it correctly identifies positive tests 96.2% of the time. But it is the best you can do for now.

Bayesian statistics

The test result comes back positive. So, what is the probability of you having the flu?

Possibilities:

1. Pr(flu) = 1? The test is not perfect, so we cannot be certain.
2. Pr(flu) > 0.065? Yes, but greater by how much? Bayes Theorem can help.

Bayesian statistics

The test result comes back positive. So, what is the probability of you having the flu?

Possibilities:

1. Pr(flu) = 1? The test is not perfect, so we cannot be certain.
2. Pr(flu) > 0.065? Yes, but greater by how much? Bayes Theorem can help.

Events:

A is the event that you genuinely are positive for the flu

B is the event that you get a positive result from the test

Bayesian statistics

Looking at Bayes Theorem:

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

We are interested in Pr(A | B), the probability that you are genuinely positive given that the test result is positive

We need to figure out what Pr(A), Pr(B) and Pr(B | A) are.

Bayesian statistics

$Pr(A)$ is the probability of being positive for the flu. In other words, the probability that you actually have the flu. We know that 6.5% of people in your city have the flu. So this is easy.

$$Pr(A) = 0.065$$

Events:
A is the event that you genuinely are positive for the flu
B is the event that you get a positive result from the test

Bayesian statistics

Pr(B | A) is the probability that we test positive given that we have the flu. We know that our at-home test kit correctly identifies positive tests 96.2% of the time. Therefore:

$$Pr(B \mid A) = 0.967$$

Events:
A is the event that you genuinely are positive for the flu
B is the event that you get a positive result from the test

Bayesian statistics

Finally Pr(B) is the overall probability that you test positive. Upon further research, we find out that 7% of all tests given to participants were positive, regardless of whether they were genuinely infected or not. So,

$$Pr(B) = 0.07$$

Events:
A is the event that you genuinely are positive for the flu
B is the event that you get a positive result from the test

Bayesian statistics

We have all we need!

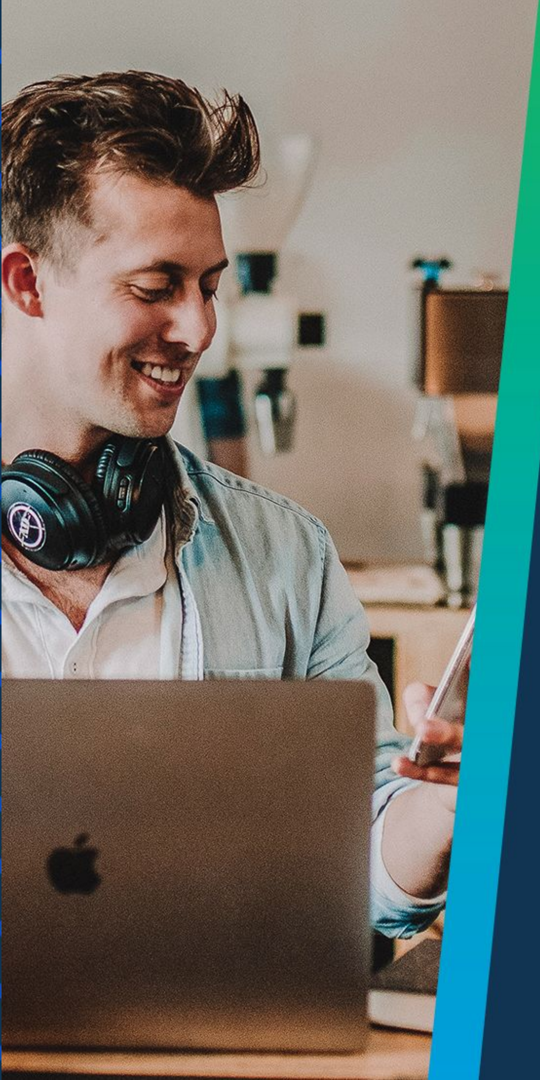
$$Pr(A \mid B) = (Pr(B|A) \; . \; Pr(A)) \; / \; Pr(B)$$
$$Pr(A \mid B) = ((0.967)(0.065)) \; / \; (0.07) = 0.897$$

So the probability of being genuinely infected given that the test result came back positive is 0.897

Hyperiondev

Q & A Section

Please use this time to ask any questions relating to the topic explained, should you have any



Hyperiondev

Thank you for joining us