ECS7024 Statistics for Artificial Intelligence and Data Science

Topic 4: Probability

William Marsh

Outline

- Aim: Informal understanding of the rules of probability

- Boolean (binary) variables
- Probability
- Conditional probability
- Bayes theorem

Background

- We have already encountered probability concepts
- Probability underlies statistics, but connection often not obvious
- Probability can be hard – paradoxes and puzzles – not relevant for us

Basic Probability

A probability is a number between 0 and 1

An Event

- An event that does or does not occur
  - Does Arsenal score more goals than Man City?

- Can be represented as a Boolean (logical) variable
  - Arsenal = True (Arsenal has more)
  - Arsenal = False (Arsenal does NOT have more)

- Which is more probable? Will Arsenal win?

Probability Values

- We write P(A) to mean P(A is True)
- Probability P(A) or Pr(A)
- Value range of P(A)
  - 0 (or 0%) impossible
  - 1 (or 100%) certain

Statement (Belief)	Probability	
Arsenal cannot win	0	0 %
Arsenal is sure to win	1	100%
Arsenal is twice as likely to win	2/3	66%

Interpretation of Probability

- Objectivist (or frequentist)
  - Run repeated trials (e.g. flipping a coin)
  - Relative frequency of outcomes
  - Limitation: play the match multiple times?

- Subjectivist (or Bayesian)
  - Degree of belief (ensuring consistency)
  - Issue: different people will have different beliefs

Rules of Probability

Rules of Probability I

- We write
  - P(A) to mean p(A is True)
  - P(not A) to mean P(A is False)

1. 0 <= P(A) <= 1
2. P(A) + P(not A) = 1

- Rule 1: a probability is between ‘impossible’ and ‘certain’
- Rule 2: something must happen
  - Implies P(not A) = 1 – P(A)

Results and Events

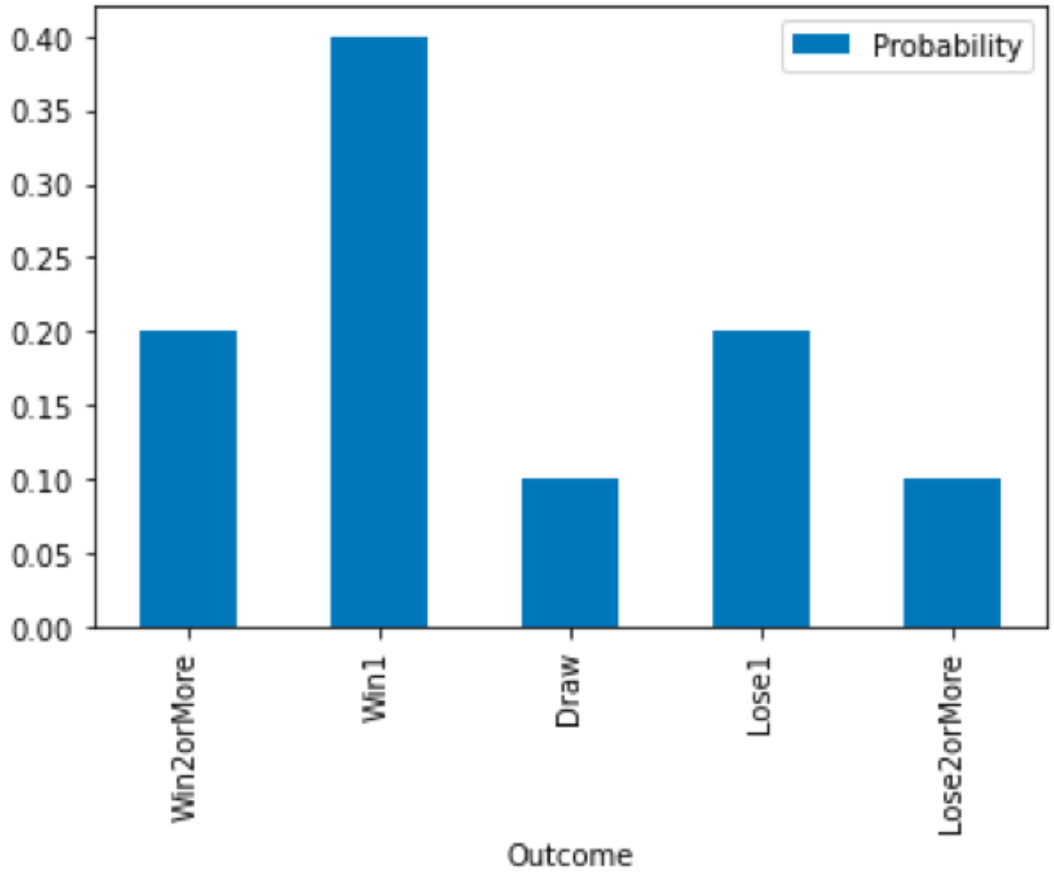
- What is the result of a football match (from Arsenal’s point of view)?
  - Distinguish 2 results: Win, Not win
  - Distinguish 3 results: : Win, Draw, Lose
  - Distinguish 5 results: Win by 2+, win by 1, draw, lose by 1, lose by 2+

- We can define a probably using any set of **mutually exclusive** results
  - Rule 1’: all probabilities are between 0 and 1
  - Rule 2’: total of all probabilities is 100%

Probability Distribution

- Assign a probability to each outcome
  - Outcomes are mutually exclusive
  - Outcomes are exhaustive – one must happen

Vertical axis with count or frequency can be converted to a probability



Rules: Multiple Independent Events

- Let A and B be independent (binary) events
  - E.g. A is coin shows heads
  - E.g. B is dice throw gets six
  - Result of A tells us nothing about B
- P(A and B) (often written P(A,B)) – probability of heads and a six
- P(A or B) – probability of heads or six

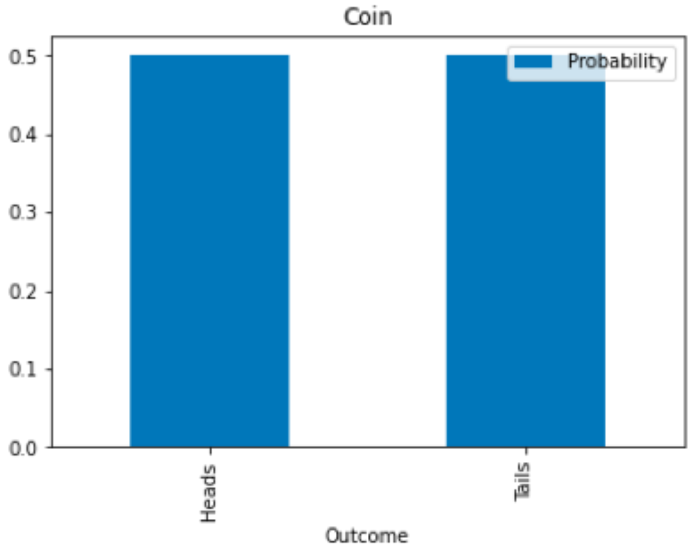
Rules: Multiple Independent Events

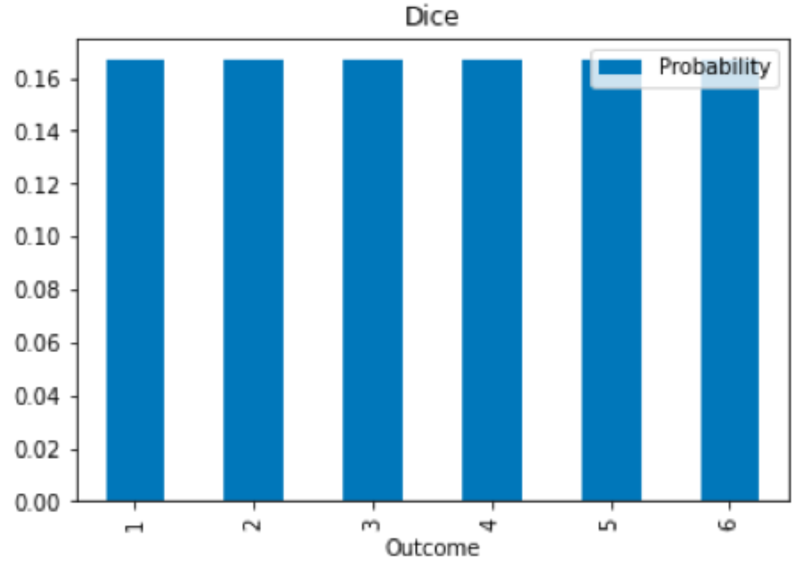
- Let A and B be independent (binary) events
  - E.g. A is coin shows heads
  - E.g. B is dice throw gets six
  - Result of A tells us nothing about B
- P(A and B) (often written P(A,B)) – probability of heads and a six
- P(A or B) – probability of heads or six

Rule 3$_{\text{(Independent)}}$: P(A and B) = P(A).P(B)

Rule 4: P(A or B) = P(A) + P(B) – P(A and B)

Example





- Flip a coin and roll a dice
  - P(Heads and ‘1’) = 1/2 x 1/6 = 1/12
  - P(Heads or ‘3’) = 1/2 + 1/6 – 1/12 = 7/12
  - P(Heads or ‘>3’) = 1/2 + 1/2 – 1/4 = 3/4

Quiz 1

Expected Value

We previously learnt about averages for continuous distribution. What about categorical distributions?

Summarising Categorical Distribution

- Categorical distribution
  - Has a mode
  - No median: categories not ordered
  - No mean: category values cannot be added

- Two exceptions
  1. If the category is associated with a value
     - Expected value
  2. Ordinal categories
     - Median

Expected Value: Example

- Manager of a fleet of cars wants to estimate the annual cost of repairs
  - All cars have an annual service (fixed cost)
  - Some cars require a further repair

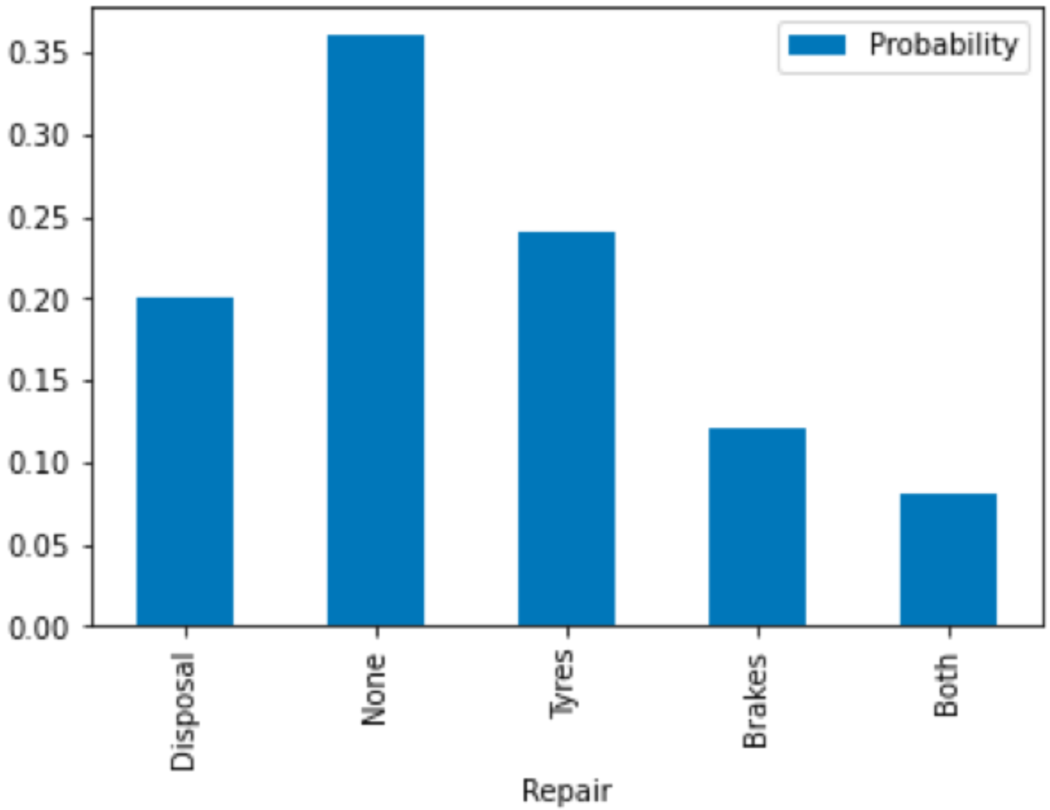
- Annual estimates
  - 20% cars disposed of
  - Other 80%
    - P(T) = 40%, P(B) = 25%
    - Repairs can be combined – independent

- Similar to a mean
  - Also ‘weighted average’

Expected Value: Example

- Manager of a fleet of cars wants to estimate the annual cost of repairs
  - All cars have an annual service (fixed cost)
  - Some cars require a further repair

- Annual estimates
  - 20% cars disposed of
  - Other 80%
    - P(T) = 40%, P(B) = 25%
    - Repairs can be combined – independent



- Similar to a mean
  - Also ‘weighted average’

Expected Value: Example II

- Expected value is:

sum (probability x cost)

Repair Type	Cost (incl labour)
Tyres	£400
Brakes	£200
Disposal	£7,000

Expected Value: Example II

- Expected value is:

$\boxed{\text{sum (probability x cost)}}$

Repair Type	Cost (incl labour)
Tyres	£400
Brakes	£200
Disposal	£7,000

Repair	Probability	Total Cost	Expected Cost
Disposal	0.2	7000	1400
None	0.36	0	0
Tyres only	0.24	400	96
Brakes only	0.12	200	24
Both	0.08	600	48
		Total	1568

Expected Value: Notation

- If X is a numeric variable
  - Values are x_1, x_2 and x_3
  - Probabilities $p(x_1)$, $p(x_2)$, and $p(x_3)$
- Expected value E[X]
  - $E[X] = x_1.p(x_1) + x_2.p(x_2) + x_3.p(x_3)$

Expected Value: Notation

- If X is a numeric variable
  - Values are x_1, x_2 and x_3
  - Probabilities $p(x_1)$, $p(x_2)$, and $p(x_3)$
- Expected value E[X]
  - $E[X] = x_1.p(x_1) + x_2.p(x_2) + x_3.p(x_3)$
- Dice example:
  - Values {1, 2, 3, 4, 5, 6}

$$\mathrm{E}[\mathrm{dice}] = \left(\frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6}\right) = 3.5$$

Expected Value: Notation

- If X is a numeric variable
  - Values are x_1, x_2 and x_3
  - Probabilities $p(x_1)$, $p(x_2)$, and $p(x_3)$
- Expected value E[X]
  - $E[X] = x_1.p(x_1) + x_2.p(x_2) + x_3.p(x_3)$

Sum up

$$E[X] = \sum_{i=0}^{N} x_i . p(x_i)$$

- Dice example:
  - Values {1, 2, 3, 4, 5, 6}

$$\mathrm{E[dice]} = \left(\frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6}\right) = 3.5$$

Expected Value: Notation

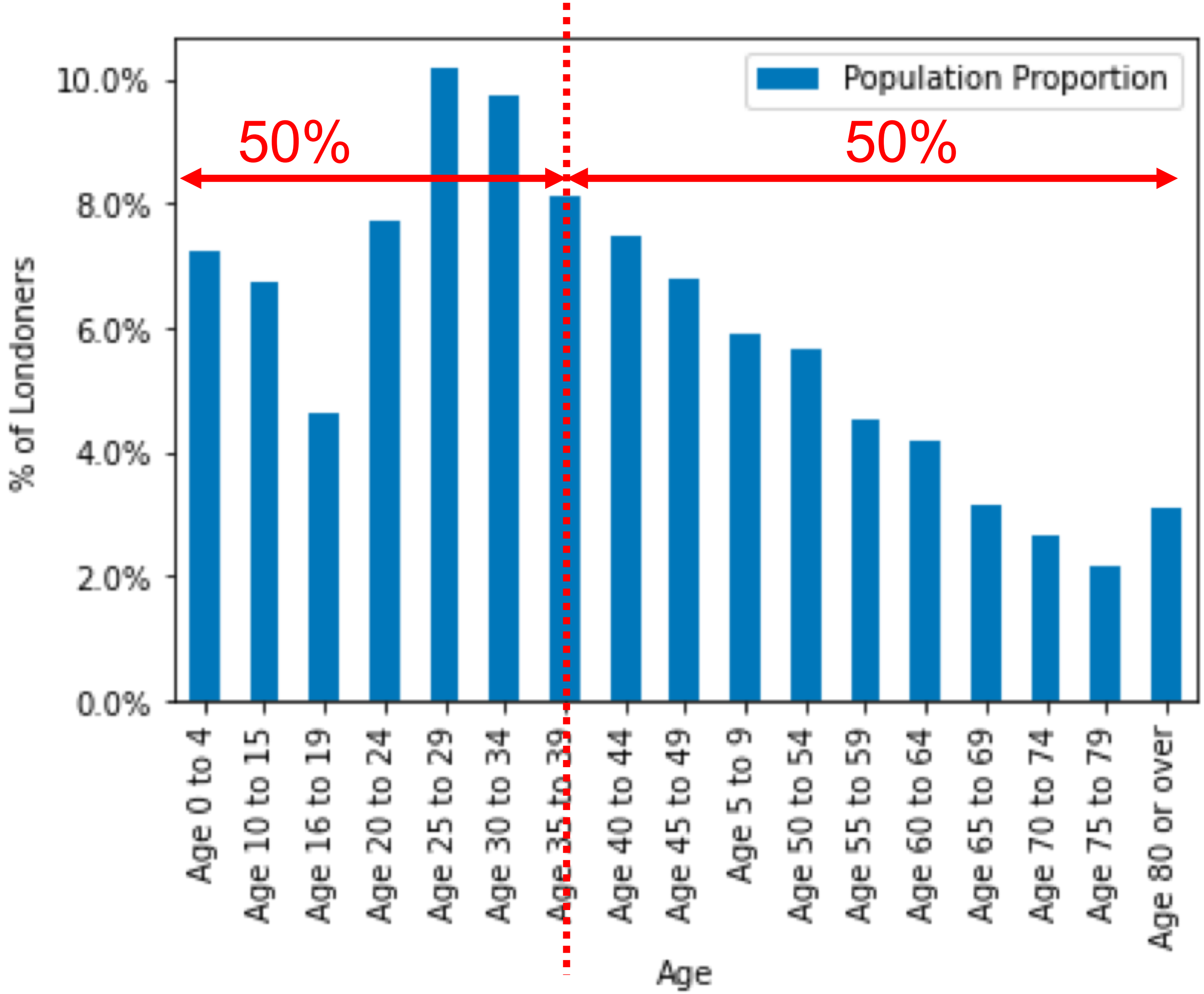
- If C is a **categorical** variable
  - Values are c_1, c_2 and c_3
- Cannot add value of C
  - Function f(C) give a (numeric) value to each category
  - 'Repair cost' example
- Expected value

$$E[f(C)] = \sum_{i=0}^{N} f(c_i).p(c_i)$$

sum (cost x probability)

Ordinal Example: Median

- London ages again



Quiz 2, 3

Every lecture will have a ‘learning reflection’ slide

Metacognition

Thinking about your thinking.

Remember: you are expert thinkers!

Metacognition

Metacognition is thinking about one's thinking. More precisely, it refers to the processes used to plan, monitor, and assess one's understanding and performance. Metacognition includes a critical awareness of a) one's thinking and learning and b) oneself as a thinker and learner.

https://cft.vanderbilt.edu/guides-sub-pages/metacognition/

Metacognition

Metacognition is thinking about one's thinking. More precisely, it refers to the processes used to plan, monitor, and assess one's understanding and performance. Metacognition includes a critical awareness of a) one's thinking and learning and b) oneself as a thinker and learner.

https://cft.vanderbilt.edu/guides-sub-pages/metacognition/

- What do I find difficult?
- What are the barriers to my understanding?
- Do I have bad learning habits?
- What strategies work for me?
- What do I need to change?

Conditional Probability

and Bayes' Theorem

Conditional Probability: P(A | B)

- Two events are not independent if the result of one changes the probability of the other
- P(A | B)
  - Probability of ‘A’ given ‘B’
  - Probability of ‘A’ is conditional on the outcome of ‘B’

Conditional Probability: P(A | B)

- Two events are <u>not independent</u> if the result of one changes the probability of the other
- P(A | B)
  - Probability of 'A' given 'B'
  - Probability of 'A' is conditional on the outcome of 'B'
- P(Statistics passed | Programming passed)

Conditional probability distribution

Statistics Result	Programming Result	
	Passed	Failed
True (i.e. pass)	90%	60%
False (i.e. fail)	10%	40%
Total	**100%**	**100%**

Rules: Non-Independent Events

- Probability of A and B uses conditional probability

Rule 3: $P(A \text{ and } B) = P(A).P(B \mid A)$

Rule 4: $P(A \text{ or } B) = P(A) + P(B) – P(A \text{ and } B)$

No change

- As before, ‘something must happen’ in all cases

Rule 2”:

$P(A \mid B) + P(\text{not } A \mid B) = 1$

$P(A \mid \text{not } B) + P(\text{not } A \mid \text{not } B) = 1$

Conditional Probability Example

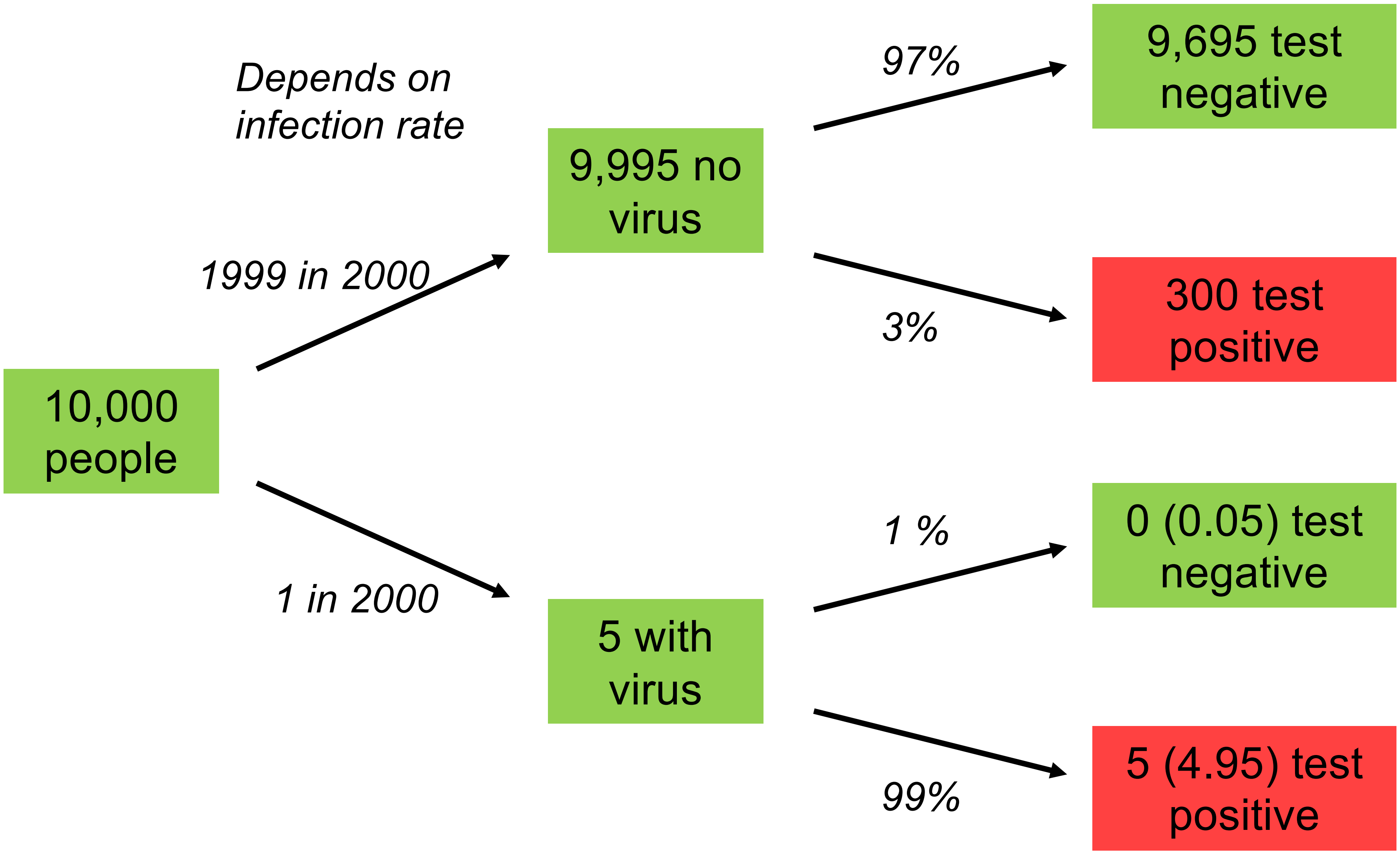
- Suppose I have a test for a virus
  - Let T stand for ‘test positive’ and V for ‘virus present’
  - The idea of the test is that T depends on V
  - However, it is not perfect …

- There are four cases

Case	Known As	Probability	Value
Test positive and virus present	True positive	P(T | V)	99%
Test negative, but virus present	False negative	P(not T | V)	1%
Test positive but virus absent	False positive	P(T | not V)	3%
Test negative and virus absent	True negative	P(not T | not V)	97%

Characteristics of test

Testing the Population



Altogether 305 test positive, but only 5 really have the virus

Bayes Theorem (Inverse Probability)

- 'Testing the Population' is an example of Bayes Theorem
  - We know P(T$_{\text{est result}}$ | V$_{\text{irus}}$), what is P (V | T) ?
- Follows from Rule 3:

$$P(V \text{ and } T) = P(V).P(T \mid V) = P(T).P(V \mid T)$$

Bayes Theorem (Inverse Probability)

- 'Testing the Population' is an example of Bayes Theorem
  - We know P($T_{\text{est result}}$ | V_{irus}), what is P (V | T) ?
- Follows from Rule 3:

$$P(V \text{ and } T) = \boxed{P(V).P(T \mid V) = P(T).P(V \mid T)}$$

Bayes Theorem

Bayes Theorem (Inverse Probability)

- 'Testing the Population' is an example of Bayes Theorem
  - We know P($T_{\text{est result}}$ | V_{irus}), what is P (V | T) ?
- Follows from Rule 3:

$$P(V \text{ and } T) = \boxed{P(V).P(T \mid V) = P(T).P(V \mid T)}$$

$$\Big\downarrow \textit{ Bayes Theorem}$$

$$\boxed{P(V \mid T) \;\; \text{is proportional to} \;\; P(V).P(T \mid V)}$$

Bayes Theorem (Inverse Probability)

- 'Testing the Population' is an example of Bayes Theorem
  - We know P(Test result | Virus), what is P (V | T) ?
- Follows from Rule 3:

$$P(V \text{ and } T) = \boxed{P(V).P(T \mid V) = P(T).P(V \mid T)}$$

Bayes Theorem

$$\boxed{P(V \mid T) \text{ is proportional to } P(V).\underbrace{P(T \mid V)}_{\text{Test accuracy}}}$$

Bayes Theorem (Inverse Probability)

- 'Testing the Population' is an example of Bayes Theorem
  - We know P(T$_{\text{est result}}$ | V$_{\text{irus}}$), what is P (V | T) ?
- Follows from Rule 3:

$$P(V \text{ and } T) = \boxed{P(V).P(T \mid V) = P(T).P(V \mid T)}$$

$$\Big\downarrow \textit{ Bayes Theorem}$$

$$\boxed{P(V \mid T) \text{ is proportional to } \underbrace{P(V)}_{\text{Prevalence of virus}}.\underbrace{P(T \mid V)}_{\text{Test accuracy}}}$$

Bayes Theorem (Inverse Probability)

- 'Testing the Population' is an example of Bayes Theorem
  - We know $P(T_{est\ result} \mid V_{irus})$, what is $P(V \mid T)$?
- Follows from Rule 3:

$$P(V \text{ and } T) = \boxed{P(V).P(T \mid V) = P(T).P(V \mid T)}$$

Bayes Theorem

$$\boxed{P(V \mid T) \text{ is proportional to } P(V).P(T \mid V)}$$

Virus status after test

Prevalence of virus

Test accuracy

Quiz

Summary

- Probability: between 0 and 1
- Probability rules
  - Mostly common sense!
  - Think about meaning
- Brief introduction to Bayes' theorem
  - Increasing important in statistics
- Binomial distribution
  - First example of a parametric distribution
  - Result of trials