Introduction to Probability

MSCI521. Statistics and Descriptive Analytics

Alisa Yusupova
LUMS, Room A48
a.yusupova@lancaster.ac.uk

11/10/2021

Marketing Analytics & Forecasting | Lancaster University Management School

Plan for the day

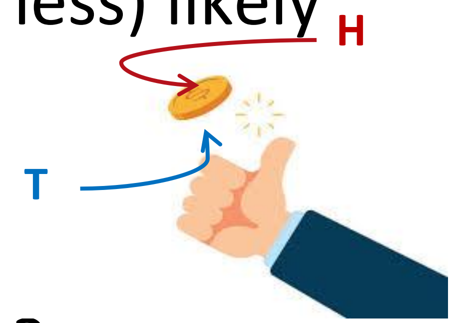
1. Probability: The Basics.
2. Combined Events:
   1. Addition of Probabilities.
   2. Multiplication of Probabilities.
   3. Conditional Probability and Independence.
3. Bayes’ Theorem.

PROBABILITY: THE BASICS

Chapter 3 of the Newbold et al. textbook.

Probability: The Basics

Consider a **fair coin** (when flipped it's not more (or less) likely to land on one side or the other):



Random Experiment

Flip the coin once. What are the possible outcomes?

Sample Space

$$\{H,T\}$$

Event

What is the probability of landing on tails? $(P(T) = ?)$

$$P(T) = \frac{N\ of\ outcomes\ that\ satisfy\ the\ condition\ T}{N\ of\ equally\ likely\ outcomes}$$

$$P(T) = \frac{1}{2} = 0.5$$

Probability: The Basics

Consider a **fair dice:** Random Experiment

Roll the dice once. What are the possible outcomes?

Sample Space

{1,2,3,4,5,6}

What is the probability of rolling a number ≤ 2?

Event

$$P(\text{rolli} \quad \leq 2) = \frac{2}{6} = \frac{1}{3}$$

What is the probability of getting an odd number?

Event

$$P(\text{odd}) = \frac{3}{6} = \frac{1}{2}$$

What is the probability of getting a 2 and a 6?

Event

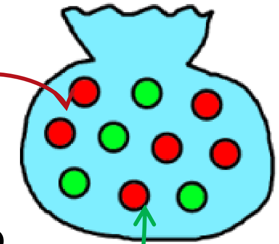
$$P(\square\ \text{AND}\ \square) = \frac{0}{6} = 0$$

Probability: The Basics

- **Experiment:** an action whose outcome is uncertain   (e.g. roll a dice)
- **Sample Space:** set of all possible outcomes of an experiment (S = {1, 2, 3, 4, 5, 6})
- **Event:** a subset of outcomes that is of interest to us (e.g. rolling an even number)
- **Probability:** measure of how likely an event is to occur (between 0 and 1)

Theoretical and Experimental Probability

Consider a bag with **50 red** and **50 green** marbles:



Theoretically, what is the probability of picking a green marble?

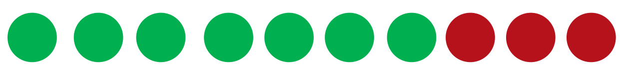
50 red

50 green

$$P(pick^{i}\quad gre^{e}\ ) = \frac{50}{10^{0}} = \frac{1}{2} = 50\%$$

'True', theoretical probability

Now lets do an experiment: pick a marble, record the colour, place it back in the bag.

After 10 experiments: 

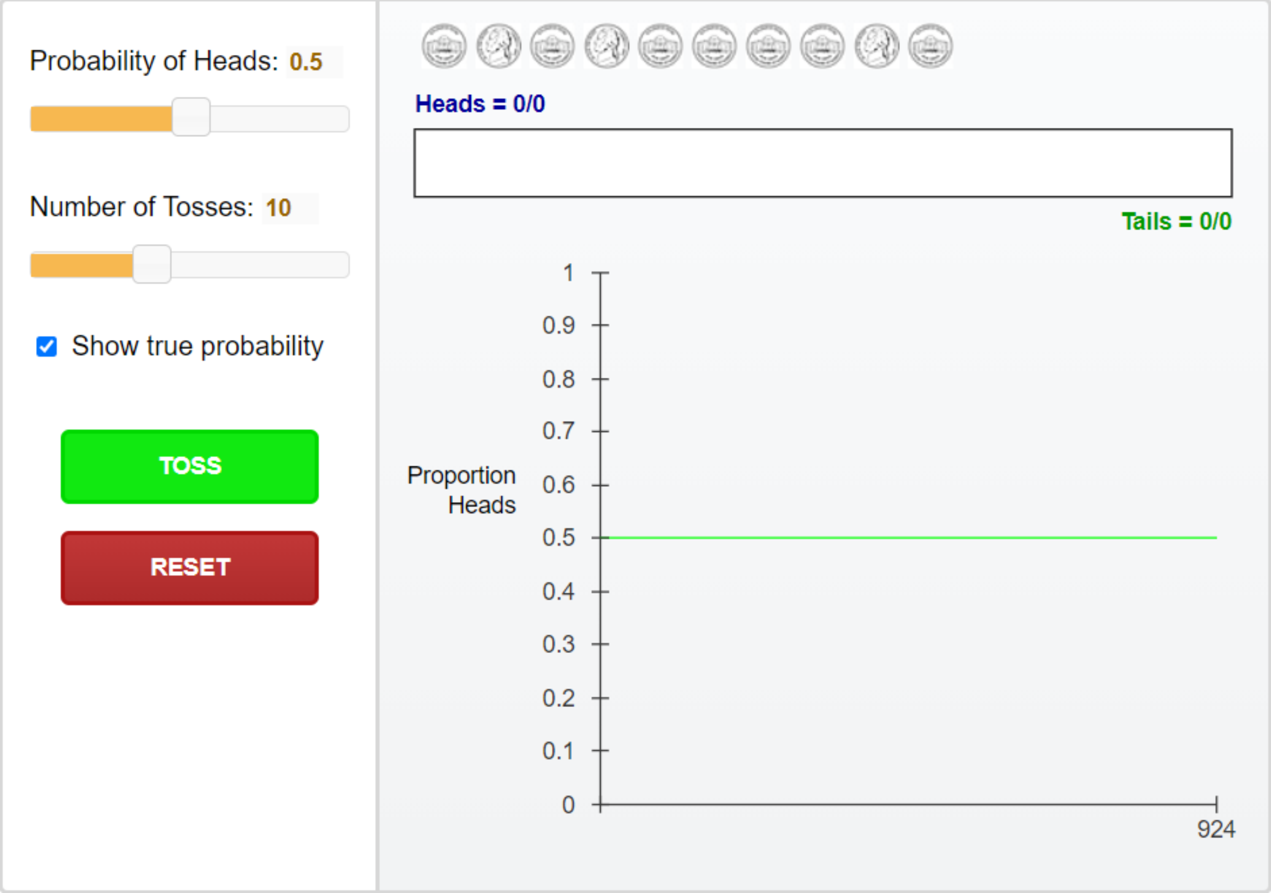
$$P(pick^{i}\quad gre^{e}\ ) = \frac{7}{10} = 0.7 = 70\%$$

Experimental probability

Over a large number of experiments, though, the percentage of red and green marbles will come to approximate the true probability of each outcome.

Theoretical and Experimental Probability

Click here to see how the experimental probability converges to the true probability as the number of experiments increases.



COMBINED EVENTS:
ADDITION OF PROBABILITIES

Chapter 3 of the Newbold et al. textbook.

Addition of Probabilities

Consider a standard deck of playing cards:

4 suits: ♠♥♣♦

Each suit contains: A,2,3,4,5,6,7,8,9,10,J,Q,K

What is the probability of picking a face card (J,Q,K)?

$$P(F) = \frac{12}{52} = \frac{3}{13}$$

What is the probability of picking a ♠?

$$P(S) = \frac{13}{52} = \frac{1}{4}$$

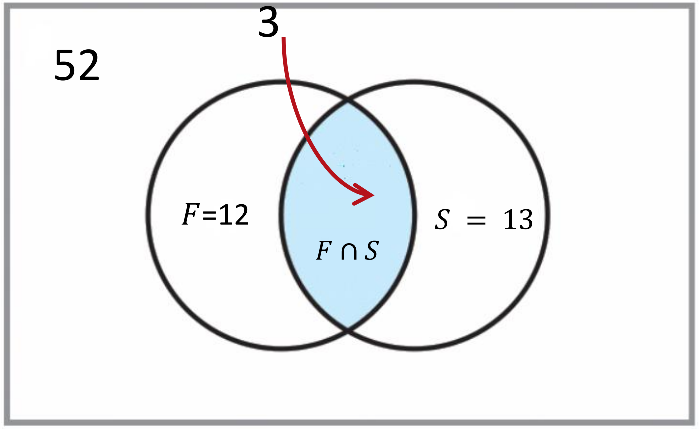
What is the probability of picking a card that is both F and S?

$$P(F \text{ AND } S) = \frac{3}{52}$$

Addition of Probabilities

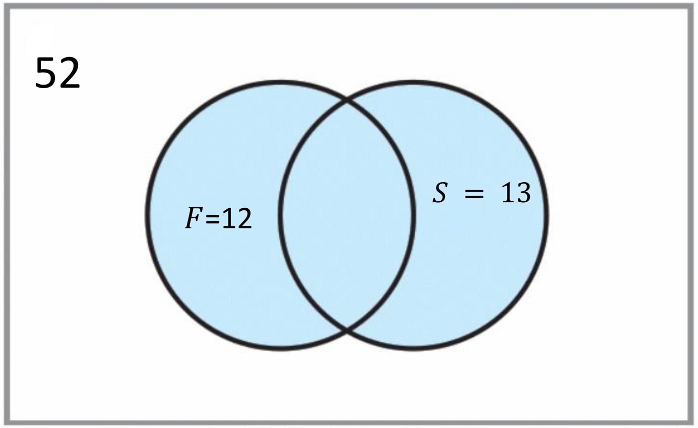
We can visualise probabilities using Venn diagrams:

$$P(F \text{ AND } S) = P(F \cap S) = \frac{3}{52}$$



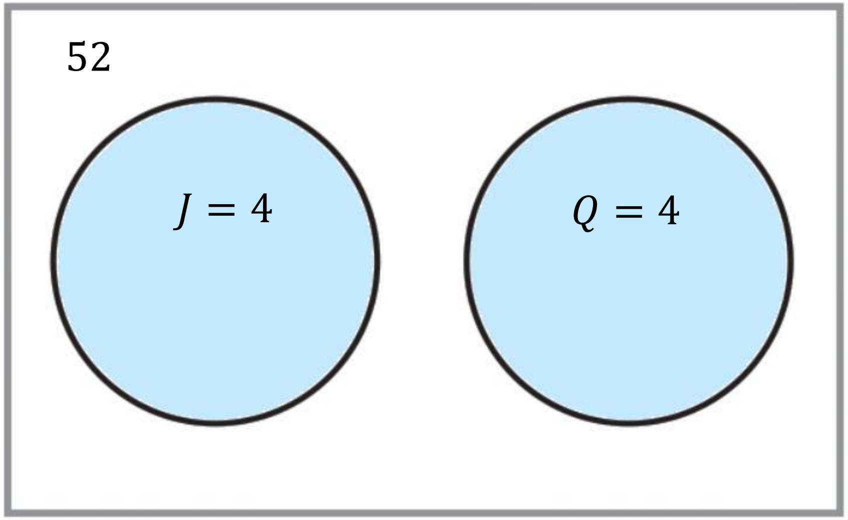
Then, what the probability of picking either a F or a S?

$$P(F \text{ OR } S) = P(F \cup S) = \frac{12}{52} + \frac{13}{52} - \frac{3}{52} =$$
$$= \frac{22}{52} = \frac{11}{26}$$



Addition of Probabilities

Now, what is the probability of picking either a Jack or a Queen (of any suit)?



$$P(J \cup Q) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$$

Mutually exclusive events

Addition Law for Probabilities (in general):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

A **special case** is when Events A and B are **mutually exclusive**, i.e. they cannot both happen, in which case the addition law:

$$P(A \cup B) = P(A) + P(B)$$

COMBINED EVENTS: MULTIPLICATION OF PROBABILITIES

Chapter 3 of the Newbold et al. textbook.

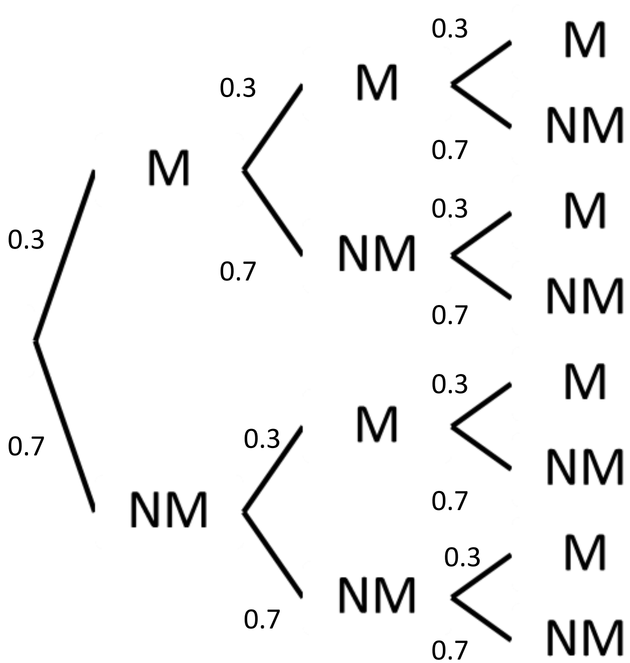
CMAF

Multiplication for Independent Events

When writing a coursework Jane has a 0.7 probability that there will be no spelling mistakes on a page. One day Jane writes a coursework that is 3 pages long.

Assuming that Jane is equally likely to have a spelling mistake on each of the 3 pages, what is the probability that she will have no spelling mistakes on at least one of them?

What happens on the first page in no way affects what happens on the other pages. The events are independent.



$$P(NM\ on\ at\ least\ 1\ page) = P(M_1 \cap M_2 \cap NM_3)$$
$$+P(M_1 \cap NM_2 \cap M_3) + P(M_1 \cap NM_2 \cap NM_3)$$
$$+P(NM_1 \cap M_2 \cap M_3) + P(NM_1 \cap M_2 \cap NM_3)$$
$$+P(NM_1 \cap NM_2 \cap M_3) + P(NM_1 \cap NM_2 \cap NM_3)$$

$$P(NM\ on\ at\ least\ 1\ page) = 1 - P(M\ on\ all\ 3\ pages)$$
$$= 1 - P(M_1 \cap M_2 \cap M_3)$$

$$P(M_1 \cap M_2 \cap M_3) = P(M_1) \times P(M_2) \times P(M_3)$$
$$= (0.3)^3 = 0.027$$

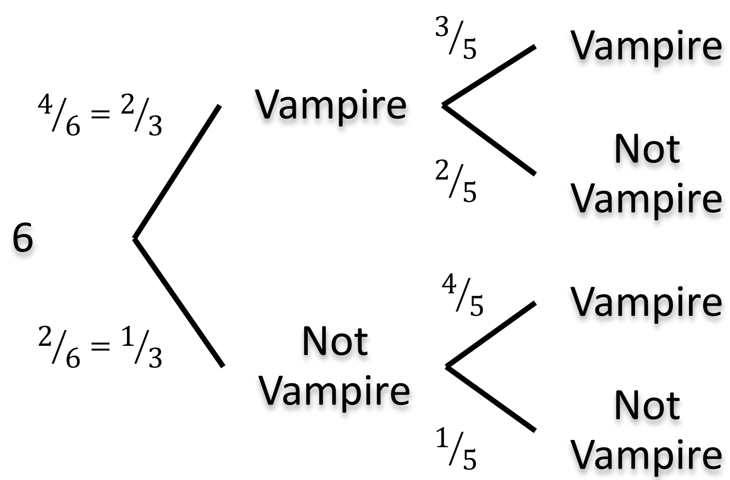
$$P(NM\ on\ at\ least\ 1\ page) = 1 - 0.027 = 0.973 \approx 97\%$$

Tree Diagrams for Dependent Events

In a class of 6, there are 4 students who are secretly vampires.

If the teacher chooses 2 students, what is the probability that neither of them are vampires?

The 1st Event (choosing a student who is not a vampire) impacts the probability of the 2nd Event (choosing a second not vampire student), i.e. the two Events are **dependent**.



$$P(NV_1 \cap NV_2) = P(NV_1) \times P(NV_2|NV_1)$$

Conditional Probability: Probability of the 2nd student being NV given that the 1st student is a NV

$$P(NV_1 \cap NV_2) = \frac{1}{3} \times \frac{1}{5} = \frac{1}{15} \approx 0.06^7 \approx 6.7\%$$

Multiplication Law for Probabilities

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

We can rearrange the above to compute conditional probabilities:

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

A **special case** is when Events A and B are **independent**, i.e. the occurrence of A has no influence on the probability of B (and vice versa), i.e.:

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

In this case the multiplication law simplifies to:

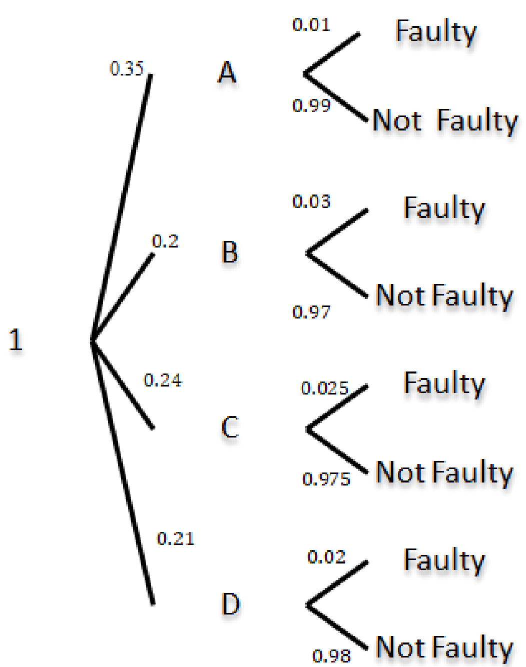
$$P(A \cap B) = P(A) \times P(B)$$

One More Example

In a factory, a brand of chocolates is packed into boxes on 4 production lines A, B, C, D. Records show that a small percentage of boxes are not packed properly for sale as follows:

Line	**A**	**B**	**C**	**D**
% Faulty	1	3	2.5	2
% Output	35	20	24	21

What is the probability that a box chosen at random from the factory's output is faulty?



$$P(F) = P(F \cap A) + P(F \cap B) + P(F \cap C) + P(F \cap D)$$

Remember, that for dependent events A and B:
$P(A \cap B) = P(A|B) \times P(B)$, hence

$$P(F) = P(F|A) \times P(A) + P(F|B) \times P(B) + P(F|C) \times P(C) + P(F|D) \times P(D)$$

$$= 0.01 \times 0.35 + 0.03 \times 0.2 + 0.025 \times 0.24 + 0.02 \times 0.21 =$$
$$0.0197 = 1.97\%$$

Law of Total Probability

Suppose events $\boldsymbol{A_1}, \boldsymbol{A_2}, \boldsymbol{A_3}, \ldots, \boldsymbol{A_n}$ are **mutually exclusive** (in our example chocolates can only be packed on one of the 4 production lines A, B, C or D) and **collectively exhaustive** ($\boldsymbol{A_1} \cup \boldsymbol{A_2} \cup \cdots \cup \boldsymbol{A_n} = \boldsymbol{S}$), then the probability of another event $\boldsymbol{B}$ (in our example, the probability of the faulty packaging) can be calculated by weighting the conditional probabilities of $\boldsymbol{B}$:

$$\boldsymbol{P}(\boldsymbol{B}) = \boldsymbol{P}(\boldsymbol{B}|\boldsymbol{A_1}) \times \boldsymbol{P}(\boldsymbol{A_1}) + \boldsymbol{P}(\boldsymbol{B}|\boldsymbol{A_2}) \times \boldsymbol{P}(\boldsymbol{A_2}) + \cdots + \boldsymbol{P}(\boldsymbol{B}|\boldsymbol{A_n}) \times \boldsymbol{P}(\boldsymbol{A_n})$$

BAYES' THEOREM

Chapter 3 of the Newbold et al. textbook.

Bayes' Theorem

Thomas Bayes, 1764 "An Essay Toward Solving a Problem in the Doctrine of Chances"

Consider two events A and B. According to the multiplication law:

$$
\begin{aligned}
P(A \cap B) &= P(A|B) \times P(B) \\
&= P(B|A) \times P(A)
\end{aligned}
$$

And thus

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

Bayes' Theorem

Bayes' Theorem (Alternative Statement)

Suppose events $\boldsymbol{A_1, A_2, A_3, \dots, A_n}$ are **mutually exclusive** and **collectively exhaustive**, and that $\boldsymbol{B}$ is another event.

Then according to the Bayes' theorem, conditional probability of $\boldsymbol{A_k}$ given $\boldsymbol{B}$:

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

Using the **Law of Total Probability**, we can rewrite the above:

$$P(A_k|B) = \frac{P(B|A_k)\ \times P(A_k)}{P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + \cdots + P(B|A_n) \times P(A_n)},$$

where $P(B) = P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + \cdots + P(B|A_n) \times P(A_n)$

Bayes' Theorem Application. Diagnostic Testing

Manchester airport is testing for COVID-19 those who arrived from abroad. If a person has COVID, the test is designed to return a "positive" result. If a person does not have the disease, the test should return a "negative" result. No test is perfect though.

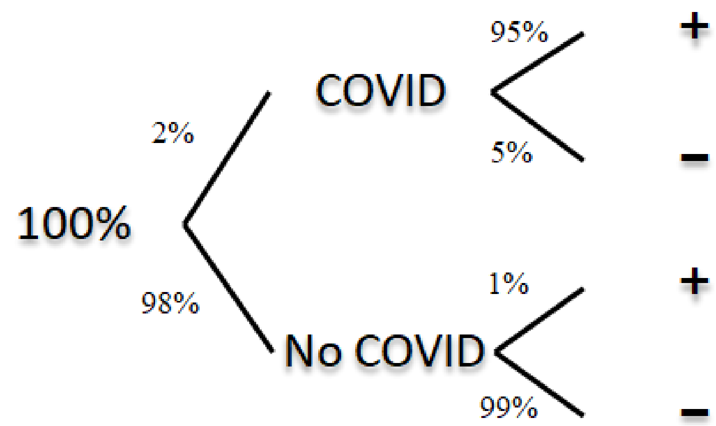
According to the ONS:

- Sensitivity of the PCR test is 95%, i.e. $P(+|COV^{I}\ \ ) = 0.95$
- Specificity of the PCR test is 99%, i.e. $P(-|NO\ COV^{I}\ \ ) = 0.99$
- 2% of all arrivals have the disease.

If a random person who arrived from abroad tests positive, what is the probability that they have the disease, i.e. $P(COV^{I}\ \ |+)$?

Bayes’ Theorem Application. Diagnostic Testing

We can use a tree diagram to illustrate the problem:



From Bayes’ theorem:

$$P(COV^{I}\ |+) = \frac{P(+|COV^{I}\ ) \times P(COV^{I}\ )}{P(+)}$$

$$= \frac{P(+|COV^{I}\ )\times P(COV^{I}\ )}{P(+|COV^{I}\ )\times P(COV^{I}\ )+P(+|No\ COV^{I}\ )\times P(No\ COV^{I}\ )}$$

$$P(COV^{I}\ |+) = \frac{0.95 \times 0.02}{0.95 \times 0.02 + 0.01 \times 0.98} = \frac{0.019}{0.019 + 0.0098} \approx 0.66$$

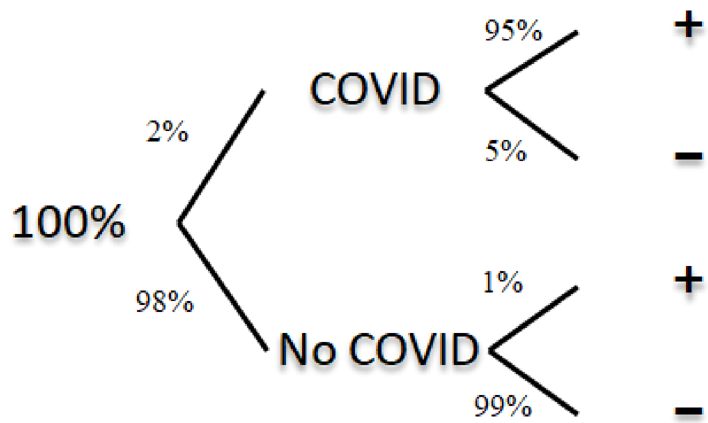
Correct Positive

False Positive:

$$P(No\ COV^{I}\ |+) = 1 - 0.66 = 0.34$$

Bayes' Theorem Application. Diagnostic Testing

Similarly, we can calculate $P(No\ COV^I\ \ |-)$:



$$P(No\ COV^I\ \ |-) = \frac{P(-|No\ COV^I\ \ ) \times P(No\ COV^I\ \ )}{P(-)}$$

$$= \frac{P(-|No\ COV^I\ \ ) \times P(No\ COV^I\ \ )}{P(-|No\ COV^I\ \ ) \times P(No\ COV^I\ \ ) + P(-|COV^I\ \ ) \times P(COV^I\ \ )}$$

$$P(No\ COV^I\ \ |-) = \frac{0.99 \times 0.98}{0.99 \times 0.98 + 0.05 \times 0.02} = \frac{0.9702}{0.001 + 0.9702} \approx 0.99$$

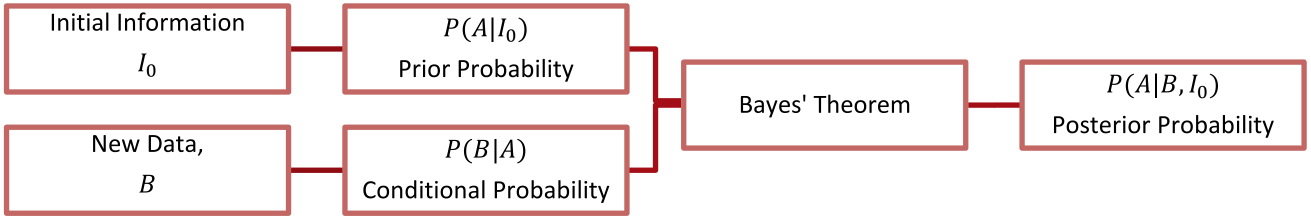
False Negative:

Correct Negative

$$P(COV^I\ \ |-) = 1 - 0.99 = 0.01$$

Bayes' Theorem. Application 2

- Often we begin a probability analysis with initial or **prior probabilities**, e.g. $P(A)$ – subjective view of the probability that $\boldsymbol{A}$ will occur based on some initial information I_0;
- Then, from a sample, a special report, or a product test we obtain some **additional information**, e.g. that event $\boldsymbol{B}$ has occurred, and we know $P(B|A)$;
- Given this information, we can calculate revised or **posterior probabilities**, e.g. $P(A|B, I_0)$;
- **Bayes' theorem** provides the means for doing this.



Bayes' Theorem. Application 2

Recall our chocolate factory example.

In a factory, a brand of chocolates is packed into boxes on 4 production lines A, B, C, D (lets call them A_1, A_2, A_3, A_4 in this case). Records show that a small percentage of boxes are not packed properly for sale as follows:

Line	A_1	A_2	A_3	A_4
% Faulty	1	3	2.5	2
% Output	35	20	24	21

Suppose we randomly pick a box and find that it **is not packed properly**. How does this **extra information** affect the probabilities of the box being packed on the different lines?

Bayes' Theorem. Application 2

Prior Probabilities of being packaged on lines $A_1, \ldots, A_4$:

$$P(A_1|I_0) = 0.35 \quad P(A_2|I_0) = 0.2 \quad P(A_3|I_0) = 0.24 \quad P(A_4|I_0) = 0.21$$

Conditional Probabilities. Past history with the packaging lines tells us that:

$$P(F|A_1) = 0.01 \quad P(F|A_2) = 0.03 \quad P(F|A_3) = 0.025 \quad P(F|A_4) = 0.02$$

Bayes' Theorem enables us to calculate the **Posterior Probabilities**:

$$P(A_1|F, I_0), P(A_2|F, I_0), P(A_3|F, I_0), P(A_4|F, I_0) = ?$$

Bayes' Theorem. Application 2

Calculate **Posterior Probabilities**:

Given that the box is faulty, the probability that it is packed on line A_1 is:

$$P(A_1|F, I_0) = \frac{P(F|A_1) \times P(A_1|I_0)}{P(F)}$$

$$= \frac{P(F|A_1) \times P(A_1|I_0)}{P(F|A_1) \times P(A_1|I_0) + P(F|A_2) \times P(A_2|I_0) + P(F|A_3) \times P(A_3|I_0) + P(F|A_4) \times P(A_4|I_0)}$$

$$= \frac{0.01 \times 0.35}{0.01 \times 0.35 + 0.03 \times 0.2 + 0.025 \times 0.24 + 0.02 \times 0.21} \approx \mathbf{0.178}$$

Similarly,

$$P(A_2|F, I_0) = \mathbf{0.305}, P(A_3|F, I_0) = \mathbf{0.305}, P(A_4|F, I_0) = \mathbf{0.213}$$

MORE EXAMPLES

Example 1

Alisa's favourite colours a purple and yellow. She has:

1 purple shirt, **1 yellow shirt**, **1 purple hat**, **1 yellow scarf**, **1 purple skirt** and **1 yellow skirt**

Alisa selects one of these garments at random. Let A be the event that she chooses a yellow garment and B that she picks a skirt.

Which of the following statements are true (choose all answers that apply)?

Let's vote: http://etc.ch/vdNo

a) $\boldsymbol{P(A|B) = P(A)}$,

b) $\boldsymbol{P(B|A) = P(B)}$,

c) Events $\boldsymbol{A}$ and $\boldsymbol{B}$ are independent.

d) The outcomes of events $\boldsymbol{A}$ and $\boldsymbol{B}$ are dependent on each other.

e) $\boldsymbol{P(A \cap B) = P(A) \times P(B)}$.

Example 1

Lets solve this problem step by step.

Step 1: calculate P(A), probability of choosing a yellow garment:

$$\boldsymbol{P(A)} = \frac{3}{6} = \frac{1}{2}$$

Step 2: calculate P(B), probability of choosing a skirt:

$$\boldsymbol{P(B)} = \frac{2}{6} = \frac{1}{3}$$

Step 3: calculate $\mathrm{P(A|B)}$, the probability of choosing a yellow garment given that she has chosen a skirt:

$$\boldsymbol{P(A|B)} = \frac{P(A \cap B)}{P(B)} = \frac{1}{6} \times \frac{3}{1} = \frac{1}{2}$$

Example 1

Step 4: calculate $\boldsymbol{P(B|A)}$, the probability of choosing a skirt given that she has chosen a yellow garment:

$$\boldsymbol{P(B|A)} = \frac{P(A \cap B)}{P(A)} = \frac{1}{6} \times \frac{2}{1} = \frac{1}{3}$$

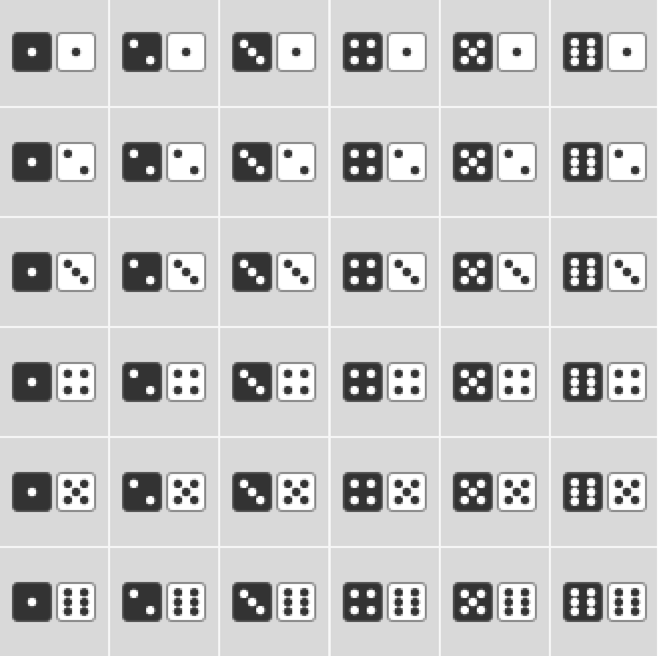
Therefore, statements **a** and **b,** are true, as well as statement

c ($\mathrm{P(A|B) = P(A)}$ holds for independent events) and

e ($\mathrm{P(A \cap B) = P(A) \times P(B)}$ is a multiplication rule for independent events).

Example 2

You roll a pair of fair six-sided dice. The sample space of possible outcomes is shown below. Based on this information, answer the following questions.



a) What is $\boldsymbol{P(A)}$, the probability that the first die is a 3?

The sample space consists of 36 equally likely outcomes and in 6 of them the first die is a 3. Hence,

$$\mathrm{P(A)} = \frac{6}{36} = \frac{1}{6}$$

Example 2

b) What is $\boldsymbol{P(B)}$, the probability that the sum of the dice is 8?

In 5 outcomes the sum of the dice is 8. Hence,

$$\boldsymbol{P(B)} = \frac{5}{36}$$

c) What is the probability that the first die is a 3 and the sum of the dice is 8?

There is only one outcome that satisfies the condition that the first die is a 3 and the sum of the dice is 8. Hence,

$$\boldsymbol{P(A \cap B)} = \frac{1}{36}$$

Example 2

d) Find the conditional probability that the sum of the dice is 8 given that the first die is a 3?

From the multiplication law:

$$\boldsymbol{P}(\boldsymbol{B}|\boldsymbol{A}) = \frac{P(\boldsymbol{A} \cap \boldsymbol{B})}{P(A)} = \frac{1}{36} \times \frac{6}{1} = \frac{1}{6}$$

e) Are the events $\boldsymbol{A}$ and $\boldsymbol{B}$ independent?

The events $\boldsymbol{A}$ and $\boldsymbol{B}$ are independent if

$$\boldsymbol{P}(\boldsymbol{B}|\boldsymbol{A}) = \boldsymbol{P}(\boldsymbol{B}) \text{ (or } \boldsymbol{P}(\boldsymbol{A}|\boldsymbol{B}) = \boldsymbol{P}(\boldsymbol{A})\text{).}$$

In our case this condition is not satisfied, so the events are not independent.

Example 3

Ivan and Alisa are playing football and are having a penalty shootout. They take penalties and defend in turn.

Ivan has probability ½ of netting a penalty. If Alisa scores he becomes very distressed and will always miss afterwards.

Alisa has severe myopia and can only score with probability 2/7. If Ivan scores she cries and her eyesight becomes even worse, so she will always miss afterwards.

If Ivan and Alisa each attempt a penalty once, and Ivan shoots first, what is the probability that Ivan misses and Alisa scores?

Example 3

The probability of event A happening, then event B, is the probability of event A happening times the probability of event B happening given that event A already happened, i.e

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

In this case, event A is Ivan not scoring a penalty and event B is Alisa scoring a penalty.

Ivan attempts first, so the probability of Ivan missing is

$$P(A) = 1 - \frac{1}{2} = 1/2$$

If Ivan didn’t net the penalty, Alisa has a normal chance to score. So, the probability of her scoring given Ivan missed is

$$P(B|A) = \frac{2}{7}$$

Example 3

The probability that Ivan misses the penalty, but Alisa scores is then

$$P(A \cap B) = \frac{1}{2} \times \frac{2}{7} = \frac{1}{7} \approx 14\%$$

…..not very high indeed..:(

CONCLUSIONS

CMAF

Conclusions

- Probability of an event $\boldsymbol{A}$ (e.g. roll an even number) is the proportion of times that $\boldsymbol{A}$ will occur, assuming that all outcomes in a sample space are equally likely to occur.
- When we calculate probabilities involving one event **OR** another event occurring (e.g. pick a Jack or a Queen), we add their probabilities (Venn diagrams are useful);
- Addition rule depends on whether the events are mutually exclusive or not;
- When we calculate probabilities involving one event **AND** another event occurring (e.g. choose 2 students that are both vampires), we multiply their probabilities (Tree diagrams are useful);
- Multiplication rule depends on whether the events are dependent or independent;

Conclusions

- In some cases, the first event happening impacts the probability of the second event. We call these **dependent events**;
- In other cases, the first event happening does not impact the probability of the seconds. We call these **independent events**;
- **Bayes' theorem** is foundational to statistics because it allows us to go from having the probability of an observation given a belief to determining the strength of that belief given the observation;
- It also provides a procedure for determining how probability statements should be adjusted, given additional information.

Thank you for your attention!

Alisa Yusupova

LUMS, Room A48

a.yusupova@lancaster.ac.uk