Statistical Methods

Lecture 4 – Probability

Luke Dickens

Autumn 2020

Experiments, outcomes and events

Definition: An **experiment** is any process that produces an observation or **outcome**.

Definition: The set of all possible outcomes of an experiment is the **sample space**, denoted S.

Definition: An **event** is any set of outcomes of the experiment. An event is a subset of the sample space, denoted $A, B, C, \ldots$

We say event **A occurs** when the outcome is contained in A.

An event can include some, all or none of the outcomes in S.

The empty event is denoted $\emptyset$.

Example events

Ross [Ros17] gives a number of examples, one is the finishing positions of horses in a seven horse race:

- $S = \{\text{all orderings of } 1, 2, 3, 4, 5, 6, 7\}$
- outcome $(4, 1, 6, 7, 5, 3, 2)$ means horse 4 comes first, horse 1 comes second, ...
- If A is the event that horse 4 comes first, then:

$$A = \{(h_1, h_2, \ldots, h_7) \in S \mid h_1 = 4\}$$

- ... if outcome is $(4, 1, 6, 7, 5, 3, 2)$ then event A occurs
- ... if outcome is $(6, 2, 5, 7, 4, 3, 1)$ then event A does not occur

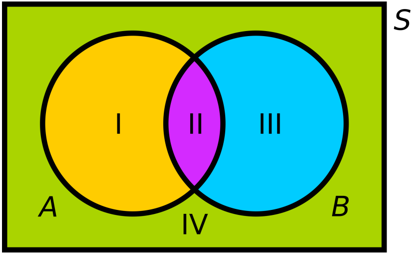
Union, intersection and complement

- The **union of events A and B**, $A \cup B$ comprises all outcomes that are in A or in B or in both.
- The **intersection of events A and B**, $A \cap B$ comprises all outcomes that are both in A and in B.
- The **complement of event A**, A^c comprises all outcomes in the sample space that are not in A.
- Unions or intersections can be over 3 or more events, e.g. $A \cup B \cup C$ comprises any event in at least one of A, B or C.
- Brackets used to specify order of operations, e.g. $(A \cup B) \cap C$ is read as:

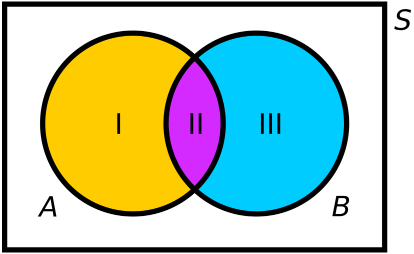
  Outcomes in C and also in either A or B.

How would you read $A \cup (B \cap C)$?

Illustrating union of events

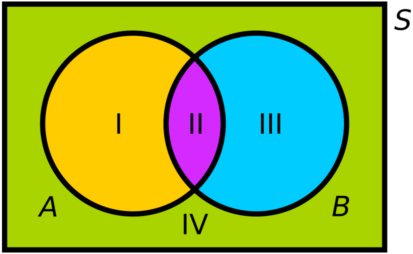


Venn diagram showing sample space, S, and events A and B.

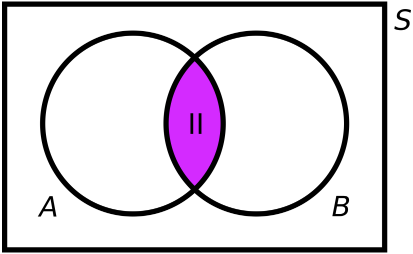


Venn diagram with event $A \cup B$ coloured.

Illustrating intersection of events

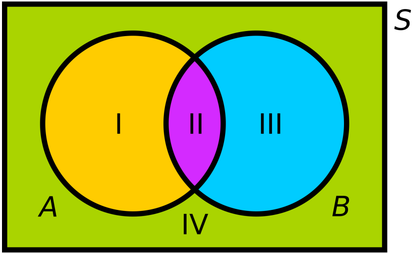


Venn diagram showing sample space, S, and events A and B.

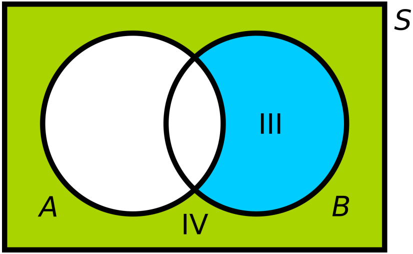


Venn diagram with event $A \cap B$ coloured.

Illustrating complement of event



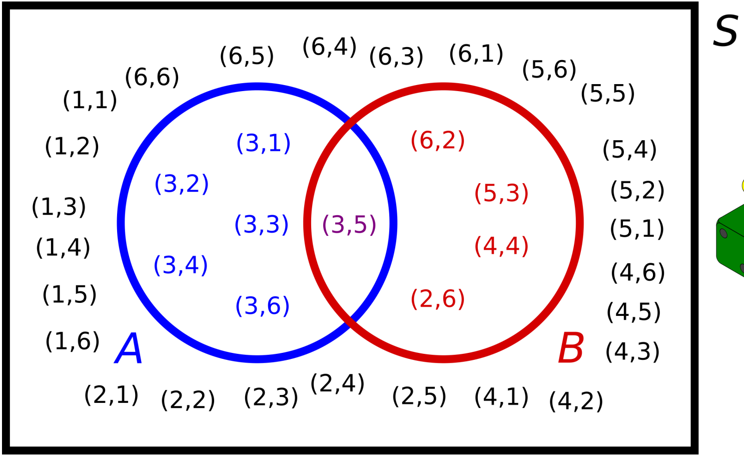
Venn diagram showing sample space, S, and events A and B.

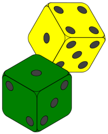


Venn diagram with event A^c coloured.

Example: 2 Dice Experiment

Outcomes (i, j) (green die shows i and yellow shows j).
Event A: green die shows 3. Event B: values sum to 8.





Dice image sourced from here.

The long run frequency of an event

In (Ross, 2017, Sec. 4.3) it says:

> *It is an empirical fact that if an experiment is continually repeated under the same conditions, then, for any event A, the proportion of times that the outcome is contained in A approaches some value as the number of repetitions increases. For example, if a coin is continually flipped, then the proportion of flips landing on tails will approach some value as the number of flips increases. It is this long-run proportion, or relative frequency, that we often have in mind when we speak of the probability of an event.*

What does this really mean? What is an "empirical fact"?

The (Weak) Law of Large numbers

To say event A has probability p of occuring (of success), means that for large enough number of trials n, the fraction of successes will very likely be close to 1.

More precisely: The chance that the fraction of successes in n independent trials is close to p approaches 1 as n grows.

Yet more precisely:
For any real number $e > 0$ and as n increases – for n trials

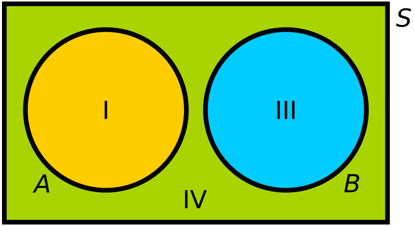
$$P\left(\left|\frac{\text{number of outcomes in A}}{n} - p\right| < e\right) \text{ approaches } 1$$

Try reading first just the blue. Then read the blue and green together. Finally read the whole statement.

For an experiment with sample space S, we suppose that for each event A there is a number, $P(A)$ called the probability of event A with the following properties:

1. $0 \leq P(A) \leq 1$
2. $P(S) = 1$
3. If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

Property 3 for non-overlapping (mutually exclusive) events.

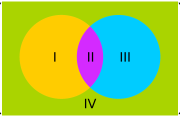


Interpreting Property 3

Property 3: If $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$

In terms of more general diagrams, this says that the combined probability of two (or more) non-overlapping regions is the sum of the probabilities of each region individually.

For example (also shows property 2):

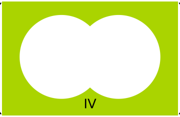
$P(S) = P$ (

)

$= P$ (

) $+ P$ (

) $+ P$ (

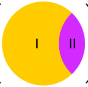
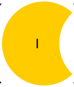
) $+ P$ (

)

$= 1$

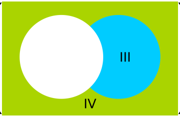
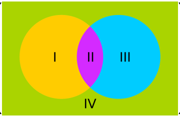
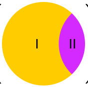
Other examples of these properties

The probability of event A:

$P(A) = P\Bigg($  $\Bigg) = P\Bigg($  $\Bigg) + P\Bigg($  $\Bigg)$

$$= P(A \cap B^c) + P(A \cap B)$$

The probability of the complement of A:

$P(A^c) = P\Bigg($  $\Bigg) = P\Bigg($  $\Bigg) - P\Bigg($  $\Bigg)$

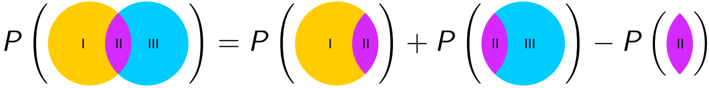
$$= P(S) - P(A) = 1 - P(A)$$

The addition rule

Definition: The addition rule for probability:

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

Graphical Proof:



The odds

The odds of event A, denoted $o(A)$ is given by:

$$o(A) = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

And indicates how many times more likely it is that A occurs than that it does not occur.

Equally likely outcomes

Consider an experiment where, it is reasonable to assume that each outcome S is equally likely to occur. Without loss of generality, let's enumerate the outcomes, so $S = \{1, 2, \ldots, N\}$, in which case:

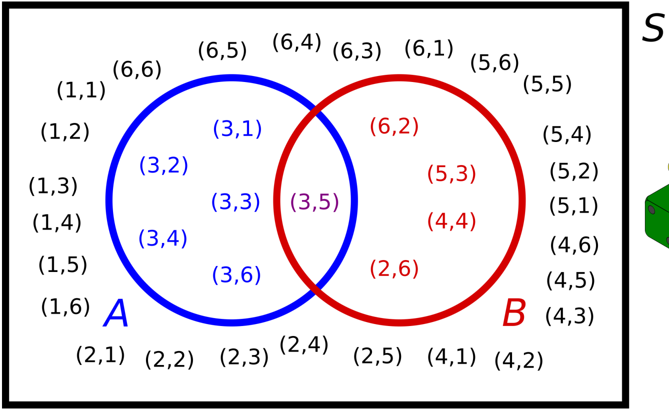
$$P(\{1\}) = P(\{2\}) = \ldots = P(\{N\})$$

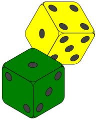
$$\text{and} \qquad P(\{i\}) = \frac{1}{N} \qquad \text{for each } i = 1, \ldots, N$$

Examples:

- The face of a fair tossed coin
- The face of a rolled die
- The ball number drawn first in a lottery
- The last digit (1000th of a second) on a paused stopwatch

Example: Two Dice again





Dice image sourced from here.

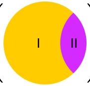
Assuming that each outcome (i, j) is equally likely, we can compute: $P(A) = \frac{1}{6}$, $P(B) = \frac{5}{36}$ and $P(A \cap B) = \frac{1}{36}$.

Conditional Probabilities

Conditional probabilities are probabilities when some partial information concerning the outcome of the experiment is available.

The **conditional probability** of event B given that event A has already been observed is denoted: $P(B|A)$, and is the **relative probability of event $A \cap B$ with respect to A**:

$P(B|A) = P\Big($ 

$\Big) \Big/ P\Big($ 

$\Big)$

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

Take Care! $P(B|A)$ is only defined if $P(A) > 0$. Why?

Long-run relative frequencies

Ross [Ros17] gives an explanation of conditional probabilities in terms of the long-run frequencies roughly as follows.

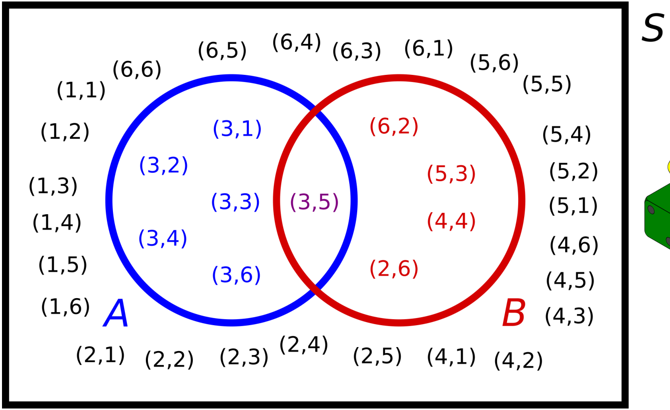
If we conduct $n \gg 1$ trials then:

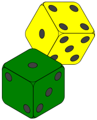
- approximately $nP(A)$ outcomes will be in A
- approximately $nP(A \cap B)$ outcomes will be in $A \cap B$ (and by definition also in A)
- call $f_{B|A}$ the fraction of outcomes in A that are also in $A \cap B$:

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

Conditional probability example

Returning to our two dice example, what is the conditional probability $P(A|B)$. (A: green die shows 3. B dice sum to 8.)





Dice image sourced from here.

The product rule

Rearranging the conditional probability formula:

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

We get the product rule:

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

Generalised to K events, A_k for $k = 1, \ldots, K$:

$$\begin{aligned} &P(A_1 \cap A_2 \cap \ldots A_K) \\ &\quad = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_K|A_1 \cap \ldots \cap A_{K-1}) \end{aligned}$$

Independence

In general conditional probability $P(B|A)$ is not equal to **marginal probability** $P(B)$.

When $P(B|A) = P(B)$, we say that **B is independent of A** and

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

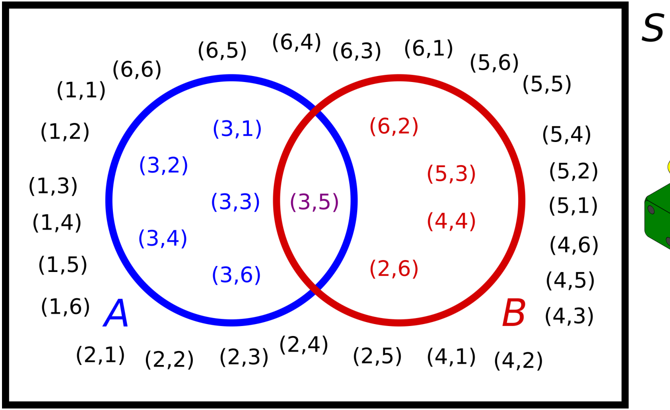
As a consequence:

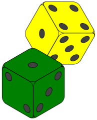
- $P(B|A) = P(B) \implies P(A|B) = P(A)$
  (independence is a **symmetric property**.)
- $P(B|A) = P(B) \implies P(B|A^c) = P(B)$
  (occurrence or non-occurrence of A tells us nothing about B.)

More generally, if $A_1, \ldots, A_K$ are independent, then

$$P(A_1 \cap A_2 \cap \ldots \cap A_K) = P(A_1)P(A_2)\ldots P(A_K)$$

Are our two dice events indepedent?





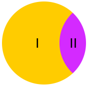
Dice image sourced from here.

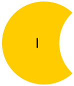
No. $P(B) = \frac{5}{36} \neq \frac{1}{6} = P(B|A)$.

$P(A) = \frac{1}{6} \neq \frac{1}{5} = P(A|B)$.

Mutually exclusive parts of A

Given event B, can partition event A into mutually exclusive parts:

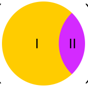
$A =$ 

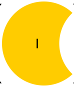
$=$ 

$\cup$ 

$$= (A \cap B^c) \cup (A \cap B)$$

And for probabilities:

$P(A) = P\Big($ 

$\Big) = P\Big($ 

$\Big) + P\Big($ 

$\Big)$

$$= P(A \cap B^c) + P(A \cap B)$$

Finally using product rule:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

Interpreting $P(A)$

Consider: $P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$

Reads as: the probability of event A is weighted average of the conditional probability of A given that B occurs, and the conditional probability of A given that B does not occur. (Recalling $P(B) + P(B^c) = 1$.)

A useful result, it allows us to construct a probability for A from conditionals, $P(A|B)$ & $P(A|B^c)$ and marginal, $P(B)$.

Bayes Theorem

Consider the problem of reevaluating an initial probability in light of additional evidence. We have a hypothesis under consideration:

- H is event that hypothesis is true.
- $P(H)$ is initial probability of H.
- We have new evidence E concerning H.

We can update our understanding of the chance of H:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|H^c)P(H^c)}$$

Requires that: If H were true, we would know how likely E is, $P(E|H)$. Similarly for $P(E|H^c)$.

Example 4.17 from [Ros17]

Insurance company believes:

- New customer is accident-prone (event H) with $P(H) = 0.2$
- Customers will have accident in 1$^{\text{st}}$ year (event A):
  - if accident-prone with $P(A|H) = 0.1$
  - if not with $P(A|H^c) = 0.05$

(a) What is the probability that any customer will have an accident in first year?

$$
\begin{aligned}
P(A) &= P(A|H)P(H) + P(A|H^c)P(H^c) \\
&= (0.1)(0.2) + (0.05)(1 - 0.2) = 0.06
\end{aligned}
$$

(b) If they have an accident in first year, with what probability are they accident prone?

$$
P(H|A) = \frac{P(A|H)P(H)}{P(A)} = \frac{(0.1)(0.2)}{0.06} = \frac{1}{3}
$$

Basic Principle of Counting

The basic principle of counting says:

> *Suppose an experiment consists of two parts. If part 1 can result in any of n possible outcomes and if for each outcome of part 1 there are m possible outcomes of part 2, then there is a total of nm possible outcomes of the experiment. [Ros17]*

If these outcomes are equally likely, we can use this to construct probabilities, e.g.

- Randomly selecting (sampling) one man and one woman from a population
- Tossing two coins
- Rolling two dice
- Taking coloured balls from a bag

Example [Ros17, Ex. 4.20]

One man and one woman are to be selected from a group consisting of 12 women and 8 men. How many different choices are possible?

Solution:

$$12 \text{ (possible women)} \times 8 \text{ (possible men)} = 96 \quad \text{(combinatons)}$$

Two gloves individually randomly selected from collection of 12 pairs (all different). (a) How many selections are possible? (b) What is the probability the two gloves are from the same pair? (Each glove is left or a right and of a specific type.)

(a) First glove is one of 24, second glove is one of 23 (remaining). $24 \times 23 = 552$ combinations.

(b) Either we choose left or right glove first. After which, there is only one way to complete the set – 24×1 ways.

$$P(\text{choosing a pair}) = \frac{24}{552} = \frac{1}{23}$$

Generalised Principle of counting:
Suppose an experiment consists of r parts. There are n_1 possible outcomes of part 1, n_2 possible outcomes of part 2, n_3 possible outcomes of part 3, and so on. Then there is a total of $n_1 \cdot n_2 \ldots n_r$ possible outcomes of the experiment.

Application – Permutations
How many **permutations** (orderings) are there of n objects:

$$n! = n \cdot (n-1) \cdot (n-2) \ldots 3 \cdot 2 \cdot 1$$

There are n ways to choose the first object, $n-1$ ways to choose the second, $n-2$ ways to choose the third and so on.

Example: Ordering three letters

How many ways are there to order the three letters a, b and c?
There are 3 objects ($n = 3$) so,

$$3! = 3 \cdot 2 \cdot 1 = 6$$

We can enumerate them: abc, acb, bac, bca, cab, cba

Example [Ros17, Ex. 4.22]

If four people in a room, what is the probability that no two share a birthday? (Ignore Feb 29$^{\text{th}}$.)

Possible arrangements of birthdays:

$$365^4 = 365 \cdot 365 \cdot 365 \cdot 365$$

Possible arrangements of different birthdays:

$$365 \cdot 364 \cdot 363 \cdot 362$$

If all arrangements are equally likely:

$$P(\text{no shared birthday}) = \frac{365 \cdot 364 \cdot 363 \cdot 362}{365 \cdot 365 \cdot 365 \cdot 365} = 0.934$$

Combinations

How many different ways to choose r objects from n possibilities without replacement (ignoring the order). *We calculate the number of ordered selections, then divide by their permutations.*

$$\begin{pmatrix} n \\ r \end{pmatrix} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-r+1)}{r!}$$
$$= \frac{n!}{(n-r)!r!}$$

By symmetry $\begin{pmatrix} n \\ r \end{pmatrix} = \begin{pmatrix} n \\ n-r \end{pmatrix}$

Example [Ros17, Ex. 4.25]

A committee of 4 people is randomly selected 5 men and 7 women. (a) How many different selections are possible? (b) What is the probability the committee will consist of 2 men and 2 women?

(a) Ways to choose 4 people from 12 is:

$$\binom{12}{4} = \frac{12 \cdot 11 \cdot 10 \cdot 9}{4 \cdot 3 \cdot 2 \cdot 1} = 495$$

(b) Ways of choosing 2 men from 5 and 2 women from 7 is:

$$\binom{5}{2}\binom{7}{2} = \frac{5 \cdot 4}{2 \cdot 1} \cdot \frac{7 \cdot 6}{2 \cdot 1} = 210$$

And so:

$$P(\text{2 men and 2 women}) = \frac{210}{495} = \frac{14}{33} = 0.424$$

References I

[Ros17] Sheldon M. Ross, *Introductory Statistics*, 4 ed., Academic Press, 2017.