

Chapter 2: Probability

The aim of this chapter is to revise the basic rules of probability. By the end of this chapter, you should be comfortable with:

- conditional probability, and what you can and can't do with conditional expressions;
 - the Partition Theorem and Bayes' Theorem;
 - First-Step Analysis for finding the probability that a process reaches some state, by conditioning on the outcome of the first step;
 - calculating probabilities for continuous and discrete random variables.
-

2.1 Sample spaces and events

Definition: A sample space, Ω , is a *set of possible outcomes of a random experiment*.

Definition: An event, A , is a *subset of the sample space*.

This means that event A is simply *a collection of outcomes*.

Example:

Random experiment: Pick a person in this class at random.

Sample space: $\Omega = \{\text{all people in class}\}$

Event A : $A = \{\text{all males in class}\}$.

Definition: Event A occurs if *the outcome of the random experiment is a member of the set A* .

In the example above, event A occurs if *the person we pick is male*.

2.2 Probability Reference List

The following properties hold for all events A, B .

- $\mathbb{P}(\emptyset) = 0$.
- $0 \leq \mathbb{P}(A) \leq 1$.
- **Complement**: $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.
- **Probability of a union**: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

For three events A, B, C :

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C).$$

If A and B are **mutually exclusive**, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

- **Conditional probability**: $\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- **Multiplication rule**: $\mathbb{P}(A \cap B) = \mathbb{P}(A | B)\mathbb{P}(B) = \mathbb{P}(B | A)\mathbb{P}(A)$.
- **The Partition Theorem**: if B_1, B_2, \dots, B_m form a partition of Ω , then

$$\mathbb{P}(A) = \sum_{i=1}^m \mathbb{P}(A \cap B_i) = \sum_{i=1}^m \mathbb{P}(A | B_i)\mathbb{P}(B_i) \quad \text{for any event } A.$$

As a special case, B and \overline{B} partition Ω , so:

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \\ &= \mathbb{P}(A | B)\mathbb{P}(B) + \mathbb{P}(A | \overline{B})\mathbb{P}(\overline{B}) \quad \text{for any } A, B. \end{aligned}$$

- **Bayes' Theorem**: $\mathbb{P}(B | A) = \frac{\mathbb{P}(A | B)\mathbb{P}(B)}{\mathbb{P}(A)}$.

More generally, if B_1, B_2, \dots, B_m form a partition of Ω , then

$$\mathbb{P}(B_j | A) = \frac{\mathbb{P}(A | B_j)\mathbb{P}(B_j)}{\sum_{i=1}^m \mathbb{P}(A | B_i)\mathbb{P}(B_i)} \quad \text{for any } j.$$

- **Chains of events**: for any events A_1, A_2, \dots, A_n ,

$$\mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 | A_1)\mathbb{P}(A_3 | A_2 \cap A_1) \dots \mathbb{P}(A_n | A_{n-1} \cap \dots \cap A_1).$$



2.3 Conditional Probability

Suppose we are working with sample space $\Omega = \{\text{people in class}\}$. I want to find the proportion of people in the class who ski. What do I do?

Count up the number of people in the class who ski, and divide by the total number of people in the class.

$$\mathbb{P}(\text{person skis}) = \frac{\text{number of skiers in class}}{\text{total number of people in class}}.$$

Now suppose I want to find the proportion of *females* in the class who ski. What do I do?

Count up the number of females in the class who ski, and divide by the total number of females in the class.

$$\mathbb{P}(\text{female skis}) = \frac{\text{number of female skiers in class}}{\text{total number of females in class}}.$$

By changing from asking about everyone to asking about females only, we have:

- *restricted attention to the set of females only,*
- or: reduced the sample space from the set of everyone to the set of females,*
- or: conditioned on the event $\{\text{females}\}$.*

We could write the above as:

$$\mathbb{P}(\text{skis} \mid \text{female}) = \frac{\text{number of female skiers in class}}{\text{total number of females in class}}.$$

Conditioning is like changing the sample space: we are now working in a new sample space of females in class.

In the above example, we could replace ‘skiing’ with *any* attribute B . We have:

$$\mathbb{P}(\text{skis}) = \frac{\# \text{ skiers in class}}{\# \text{ class}}; \quad \mathbb{P}(\text{skis} | \text{female}) = \frac{\# \text{ female skiers in class}}{\# \text{ females in class}};$$

so:

$$\mathbb{P}(B) = \frac{\# B\text{'s in class}}{\text{total } \# \text{ people in class}},$$

and:

$$\begin{aligned} \mathbb{P}(B | \text{female}) &= \frac{\# \text{ female } B\text{'s in class}}{\text{total } \# \text{ females in class}} \\ &= \frac{\# \text{ in class who are } B \text{ and female}}{\# \text{ in class who are female}}. \end{aligned}$$

Likewise, we could replace ‘female’ with any attribute A :

$$\mathbb{P}(B | A) = \frac{\text{number in class who are } B \text{ and } A}{\text{number in class who are } A}.$$

This is how we get the definition of conditional probability:

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

By conditioning on event A , we have *changed the sample space to the set of A ’s only*.

Definition: Let A and B be events on the same sample space: so $A \subseteq \Omega$ and $B \subseteq \Omega$.

The conditional probability of event B , given event A , is

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

Multiplication Rule: (Immediate from above). For any events A and B ,

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

Conditioning as ‘changing the sample space’

The idea that “*conditioning*” = “*changing the sample space*” can be very helpful in understanding how to manipulate conditional probabilities.

Any ‘unconditional’ probability can be written as a conditional probability:

$$\mathbb{P}(B) = \mathbb{P}(B | \Omega).$$

Writing $\mathbb{P}(B) = \mathbb{P}(B | \Omega)$ just means that we are looking for the probability of event B , out of all possible outcomes in the set Ω .

In fact, the symbol \mathbb{P} *belongs* to the set Ω : it has *no meaning without* Ω . To remind ourselves of this, we can write

$$\mathbb{P} = \mathbb{P}_\Omega.$$

Then
$$\mathbb{P}(B) = \mathbb{P}(B | \Omega) = \mathbb{P}_\Omega(B).$$

Similarly, $\mathbb{P}(B | A)$ means that we are looking for the probability of event B , out of all possible outcomes in the set A .

So A is just another sample space. Thus *we can manipulate conditional probabilities* $\mathbb{P}(\cdot | A)$ *just like any other probabilities, as long as we always stay inside the same sample space A .*

The trick: Because we can think of A as just another sample space, let’s write

$$\mathbb{P}(\cdot | A) = \mathbb{P}_A(\cdot)$$

Note: NOT standard notation!

Then *we can use* \mathbb{P}_A *just like* \mathbb{P} , *as long as we remember to keep the A subscript on EVERY \mathbb{P} that we write.*

This helps us to make quite complex manipulations of conditional probabilities without thinking too hard or making mistakes. There is only one rule you need to learn to use this tool effectively:

$$\mathbb{P}_A(B | C) = \mathbb{P}(B | C \cap A) \text{ for any } A, B, C.$$

(Proof: Exercise).

The rules:

$$\begin{aligned} \mathbb{P}(\cdot | A) &= \mathbb{P}_A(\cdot) \\ \mathbb{P}_A(B | C) &= \mathbb{P}(B | C \cap A) \text{ for any } A, B, C. \end{aligned}$$

Examples:

1. Probability of a union. In general,

$$\mathbb{P}(B \cup C) = \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B \cap C).$$

So, $\mathbb{P}_A(B \cup C) = \mathbb{P}_A(B) + \mathbb{P}_A(C) - \mathbb{P}_A(B \cap C).$

Thus, $\mathbb{P}(B \cup C | A) = \mathbb{P}(B | A) + \mathbb{P}(C | A) - \mathbb{P}(B \cap C | A).$

2. Which of the following is equal to $\mathbb{P}(B \cap C | A)$?

(a) $\mathbb{P}(B | C \cap A).$ (c) $\mathbb{P}(B | C \cap A)\mathbb{P}(C | A).$

(b) $\frac{\mathbb{P}(B | C)}{\mathbb{P}(A)}.$ (d) $\mathbb{P}(B | C)\mathbb{P}(C | A).$

Solution:

$$\begin{aligned} \mathbb{P}(B \cap C | A) &= \mathbb{P}_A(B \cap C) \\ &= \mathbb{P}_A(B | C) \mathbb{P}_A(C) \\ &= \mathbb{P}(B | C \cap A) \mathbb{P}(C | A). \end{aligned}$$

Thus the correct answer is (c).

3. Which of the following is true?

(a) $\mathbb{P}(\overline{B} | A) = 1 - \mathbb{P}(B | A)$.

(b) $\mathbb{P}(\overline{B} | A) = \mathbb{P}(B) - \mathbb{P}(B | A)$.

Solution:

$$\mathbb{P}(\overline{B} | A) = \mathbb{P}_A(\overline{B}) = 1 - \mathbb{P}_A(B) = 1 - \mathbb{P}(B | A).$$

Thus the correct answer is (a).

4. Which of the following is true?

(a) $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(A) - \mathbb{P}(B \cap A)$.

(b) $\mathbb{P}(\overline{B} \cap A) = \mathbb{P}(B) - \mathbb{P}(B \cap A)$.

Solution:

$$\begin{aligned} \mathbb{P}(\overline{B} \cap A) &= \mathbb{P}(\overline{B} | A) \mathbb{P}(A) = \mathbb{P}_A(\overline{B}) \mathbb{P}(A) \\ &= (1 - \mathbb{P}_A(B)) \mathbb{P}(A) \\ &= \mathbb{P}(A) - \mathbb{P}(B | A) \mathbb{P}(A) \\ &= \mathbb{P}(A) - \mathbb{P}(B \cap A). \end{aligned}$$

Thus the correct answer is (a).

5. True or false: $\mathbb{P}(B | A) = 1 - \mathbb{P}(B | \overline{A})$?

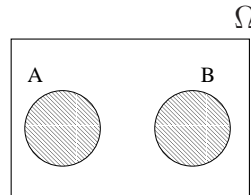
Answer: False. $\mathbb{P}(B | A) = \mathbb{P}_A(B)$. Once we have \mathbb{P}_A , we are stuck with it! There is no easy way of converting from \mathbb{P}_A to $\mathbb{P}_{\overline{A}}$: or anything else. Probabilities in one sample space (\mathbb{P}_A) cannot tell us anything about probabilities in a different sample space ($\mathbb{P}_{\overline{A}}$).

Exercise: if we wish to express $\mathbb{P}(B | A)$ in terms of only B and \overline{A} , show that
$$\mathbb{P}(B | A) = \frac{\mathbb{P}(B) - \mathbb{P}(B | \overline{A}) \mathbb{P}(\overline{A})}{1 - \mathbb{P}(\overline{A})}.$$
 Note that this does not simplify nicely!

2.4 The Partition Theorem (Law of Total Probability)

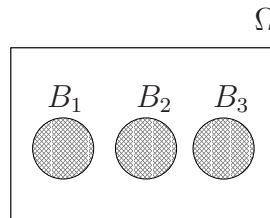
Definition: Events A and B are mutually exclusive, or disjoint, if $A \cap B = \emptyset$.

This means events A and B cannot happen together. If A happens, it excludes B from happening, and vice-versa.



If A and B are mutually exclusive, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
For all other A and B , $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Definition: Any number of events B_1, B_2, \dots, B_k are mutually exclusive if every pair of the events is mutually exclusive: ie. $B_i \cap B_j = \emptyset$ for all i, j with $i \neq j$.



Definition: A partition of Ω is a collection of mutually exclusive events whose union is Ω .

That is, sets B_1, B_2, \dots, B_k form a partition of Ω if

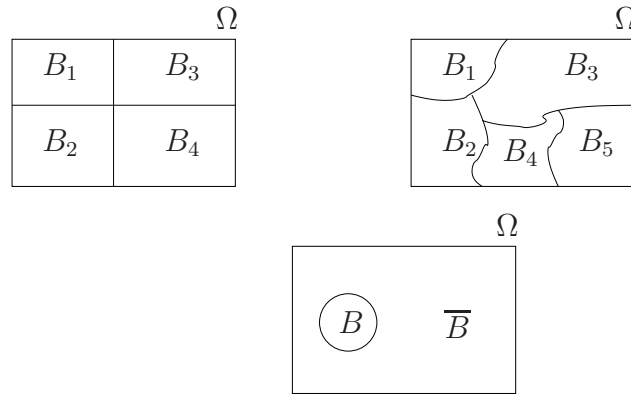
$$B_i \cap B_j = \emptyset \text{ for all } i, j \text{ with } i \neq j,$$

and

$$\bigcup_{i=1}^k B_i = B_1 \cup B_2 \cup \dots \cup B_k = \Omega.$$

B_1, \dots, B_k form a partition of Ω if they *have no overlap*
and collectively cover all possible outcomes.

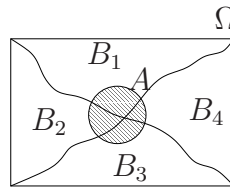
Examples:



Partitioning an event A

Any set A can be partitioned: it doesn't have to be Ω .

In particular, if B_1, \dots, B_k form a partition of Ω , then $(A \cap B_1), \dots, (A \cap B_k)$ form a partition of A.



Theorem 2.4: The Partition Theorem (Law of Total Probability)

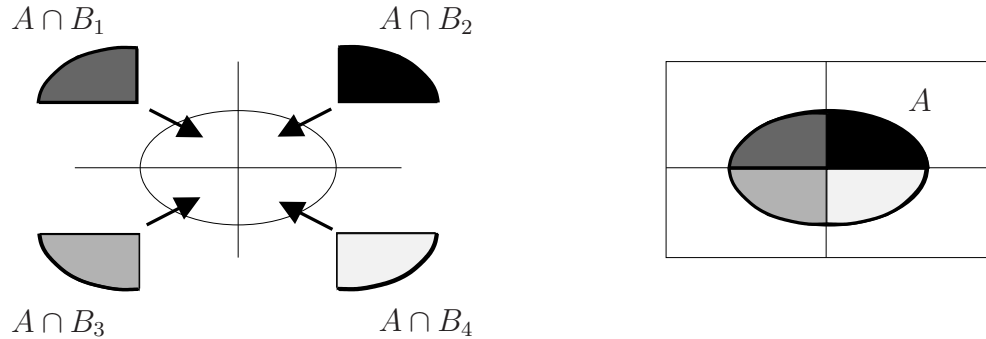
Let B_1, \dots, B_m form a partition of Ω . Then for any event A,

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

Both formulations of the Partition Theorem are very widely used, but especially the conditional formulation $\sum_{i=1}^m \mathbb{P}(A | B_i) \mathbb{P}(B_i)$.

Intuition behind the Partition Theorem:

The Partition Theorem is easy to understand because it simply states that “the whole is the sum of its parts.”



$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \mathbb{P}(A \cap B_3) + \mathbb{P}(A \cap B_4).$$

2.5 Bayes' Theorem: inverting conditional probabilities

Bayes' Theorem allows us to “invert” a conditional statement, ie. *to express* $\mathbb{P}(B | A)$ *in terms of* $\mathbb{P}(A | B)$.

Theorem 2.5: Bayes' Theorem

For any events A and B:

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

Proof:

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

$$\mathbb{P}(B | A)\mathbb{P}(A) = \mathbb{P}(A | B)\mathbb{P}(B) \quad (\text{multiplication rule})$$

$$\therefore \mathbb{P}(B | A) = \frac{\mathbb{P}(A | B)\mathbb{P}(B)}{\mathbb{P}(A)}. \quad \square$$

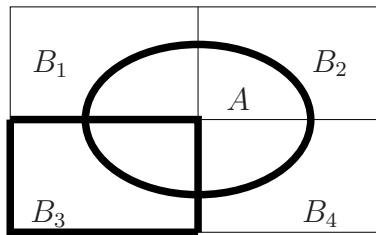
Extension of Bayes' Theorem

Suppose that B_1, B_2, \dots, B_m form a partition of Ω . By the Partition Theorem,

$$\mathbb{P}(A) = \sum_{i=1}^m \mathbb{P}(A | B_i) \mathbb{P}(B_i).$$

Thus, for *any single partition member* B_j , put $B = B_j$ in Bayes' Theorem to obtain:

$$\mathbb{P}(B_j | A) = \frac{\mathbb{P}(A | B_j) \mathbb{P}(B_j)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A | B_j) \mathbb{P}(B_j)}{\sum_{i=1}^m \mathbb{P}(A | B_i) \mathbb{P}(B_i)}.$$



Special case: $m = 2$

Given any event B , the events B and \overline{B} form a partition of Ω . Thus:

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

Example: In screening for a certain disease, the probability that a healthy person wrongly gets a positive result is 0.05. The probability that a diseased person wrongly gets a negative result is 0.002. The overall rate of the disease in the population being screened is 1%. If my test gives a positive result, what is the probability I actually have the disease?

1. Define events:

$$D = \{\text{have disease}\} \quad \overline{D} = \{\text{do not have the disease}\}$$

$$P = \{\text{positive test}\} \quad N = \overline{P} = \{\text{negative test}\}$$

2. Information given:

$$\text{False positive rate is } 0.05 \Rightarrow \mathbb{P}(P | \overline{D}) = 0.05$$

$$\text{False negative rate is } 0.002 \Rightarrow \mathbb{P}(N | D) = 0.002$$

$$\text{Disease rate is } 1\% \Rightarrow \mathbb{P}(D) = 0.01.$$

3. Looking for $\mathbb{P}(D | P)$:

$$\text{We have} \quad \mathbb{P}(D | P) = \frac{\mathbb{P}(P | D)\mathbb{P}(D)}{\mathbb{P}(P)}.$$

$$\begin{aligned} \text{Now} \quad \mathbb{P}(P | D) &= 1 - \mathbb{P}(\overline{P} | D) \\ &= 1 - \mathbb{P}(N | D) \\ &= 1 - 0.002 \\ &= 0.998. \end{aligned}$$

$$\begin{aligned} \text{Also} \quad \mathbb{P}(P) &= \mathbb{P}(P | D)\mathbb{P}(D) + \mathbb{P}(P | \overline{D})\mathbb{P}(\overline{D}) \\ &= 0.998 \times 0.01 + 0.05 \times (1 - 0.01) \\ &= 0.05948. \end{aligned}$$

Thus

$$\mathbb{P}(D | P) = \frac{0.998 \times 0.01}{0.05948} = 0.168.$$

Given a positive test, my chance of having the disease is only 16.8%.

2.6 First-Step Analysis for calculating probabilities in a process

In a stochastic process, what happens at the next step depends upon the current state of the process. We are often interested to know the probability of *eventually* reaching some particular state, given our current position.

Throughout this course, we will tackle this sort of problem using a technique called *First-Step Analysis*.

The idea is to consider all possible first steps away from the current state. We derive a system of equations that specify the probability of the eventual outcome given each of the possible first steps. We then try to solve these equations for the probability of interest.

First-Step Analysis depends upon ***conditional probability and the Partition Theorem***. Let S_1, \dots, S_k be the k possible first steps we can take away from our current state. We wish to find the probability that event E happens eventually. First-Step Analysis calculates $\mathbb{P}(E)$ as follows:

$$\mathbb{P}(E) = \mathbb{P}(E|S_1)\mathbb{P}(S_1) + \dots + \mathbb{P}(E|S_k)\mathbb{P}(S_k).$$

Here, $\mathbb{P}(S_1), \dots, \mathbb{P}(S_k)$ give the probabilities of taking the different first steps $1, 2, \dots, k$.

Example 1: Tennis game at Deuce.

Venus and Serena are playing tennis, and have reached the score Deuce (40-40). (*Deuce* comes from the French word *Deux* for ‘two’, meaning that each player needs to win two consecutive points to win the game.)

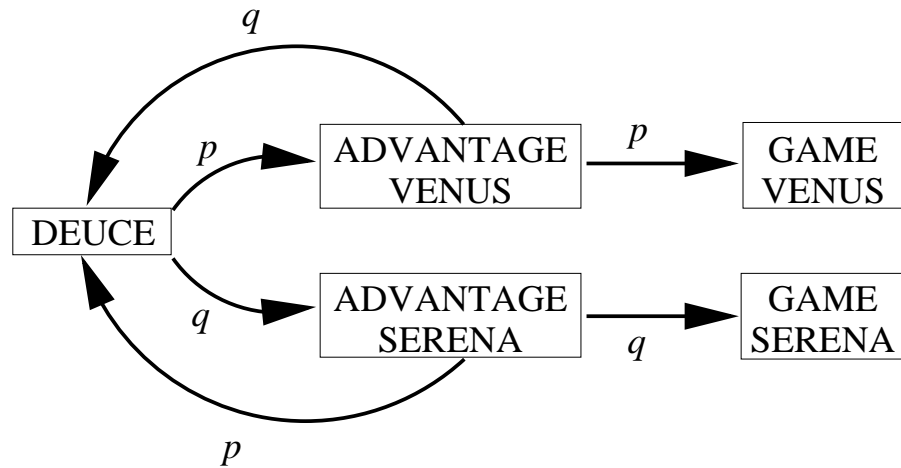


For each point, let:

$$p = \mathbb{P}(\text{Venus wins point}), \quad q = 1 - p = \mathbb{P}(\text{Serena wins point}).$$

Assume that all points are independent.

Let v be the probability that Venus wins the game eventually, starting from Deuce. Find v .



Use First-step analysis. The possible first steps starting from Deuce are:

1. Venus wins the next point (probability p): state AV;
2. Serena wins the next point (probability q): state AS.

Thus:

$$\begin{aligned}
 v = \mathbb{P}(\text{Venus wins} \mid \text{Deuce}) &= \mathbb{P}_D(VW \mid AV)\mathbb{P}_D(AV) + \mathbb{P}_D(VW \mid AS)\mathbb{P}_D(AS) \\
 &= \mathbb{P}(VW \mid AV)p + \mathbb{P}(VW \mid AS)q. \quad (\star)
 \end{aligned}$$

Now we need to find $\mathbb{P}(VW \mid AV)$, and $\mathbb{P}(VW \mid AS)$, again using First-step analysis:

$$\begin{aligned}
 \mathbb{P}(VW \mid AV) &= \mathbb{P}(VW \mid \text{Game Venus})p + \mathbb{P}(VW \mid \text{Deuce})q \\
 &= 1 \times p + v \times q \\
 &= p + qv. \quad (a)
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 \mathbb{P}(VW \mid AS) &= \mathbb{P}(VW \mid \text{Game Serena})q + \mathbb{P}(VW \mid \text{Deuce})p \\
 &= 0 \times q + v \times p \\
 &= pv. \quad (b)
 \end{aligned}$$

Substituting (a) and (b) into (\star) ,

$$\begin{aligned}v &= (p + qv)p + (pv)q \\v &= p^2 + 2pqv \\v(1 - 2pq) &= p^2 \\v &= \frac{p^2}{1 - 2pq}.\end{aligned}$$

Note: Because $p + q = 1$, we have:

$$1 = (p + q)^2 = p^2 + q^2 + 2pq.$$

So the final probability that Venus wins the game is:

$$v = \frac{p^2}{1 - 2pq} = \frac{p^2}{p^2 + q^2}.$$

Note how this result makes intuitive sense. For the game to finish from Deuce, either Venus has to win two points in a row (probability p^2), or Serena does (probability q^2). The ratio $p^2/(p^2 + q^2)$ describes Venus's 'share' of the winning probability.

Alternative approach: two-step analysis

The diagram above is a simple one, and we can see that the *only* way of winning from Deuce is to take two steps away from Deuce. In situations like these, it is worth considering whether a two-step approach is possible.

For a two-step approach, we look at the *possible pairs of steps we can take out of the state Deuce*.

Pair of steps	Probability	Outcome
$AV \rightarrow GV$	p^2	<i>Venus wins</i>
$AS \rightarrow GS$	q^2	<i>Serena wins</i>
$AV \rightarrow D$	pq	<i>back to start (Deuce)</i>
$AS \rightarrow D$	qp	<i>back to start (Deuce)</i>

Our approach to finding $v = \mathbb{P}(\text{Venus wins})$ can be summarized as:

$$\mathbb{P}(\text{Venus wins}) = v = \sum_{\text{options}} \mathbb{P}(\text{Venus wins} \mid \text{option}) \mathbb{P}(\text{option})$$

$$v = 1 \times p^2 + 0 \times q^2 + v \times pq + v \times qp$$

$$v = p^2 + 2pqv$$

$$v(1 - 2pq) = p^2$$

$$v = \frac{p^2}{1 - 2pq} \quad \text{as before.}$$



Example 2: Gambler's Ruin.

This is a famous problem in probability. A gambler starts with $\$x$. She tosses a fair coin repeatedly.

If she gets a Head, she wins $\$1$. If she gets a Tail, she loses $\$1$.

The coin tossing is repeated until the gambler has either $\$0$ or $\$N$. What is the probability of the Gambler's Ruin (i.e. that the gambler ends up with $\$0$)?

Define events:

$$S_x = \{\text{starts with } \$x\} \quad R = \{\text{ends with ruin}\}$$

Looking for:

$$\mathbb{P}(\text{ruin} \mid \text{starts with } x) = \mathbb{P}(R \mid S_x)$$

Define:

$$p_x = \mathbb{P}(R \mid S_x).$$

What we know:

$$p_0 = \mathbb{P}(\text{ruin} \mid \text{starts with } \$0) = 1,$$

$$p_N = \mathbb{P}(\text{ruin} \mid \text{starts with } \$N) = 0.$$

<i>Possible steps from state S_x</i>	<i>Probability</i>	<i>Outcome</i>
<i>Throw Head (event H)</i>	<i>0.5</i>	<i>Win \$1, start again from $\\$(x+1)$: State S_{x+1}</i>
<i>Throw Tail (event T)</i>	<i>0.5</i>	<i>Lose \$1, start again from $\\$(x-1)$: State S_{x-1}</i>

First-step analysis:

$$\begin{aligned}
 p_x &= \mathbb{P}(R | S_x) \\
 &= \mathbb{P}_{S_x}(R) \quad \text{in notation of Section 2.3} \\
 &= \mathbb{P}_{S_x}(R | H)\mathbb{P}_{S_x}(H) + \mathbb{P}_{S_x}(R | T)\mathbb{P}_{S_x}(T)
 \end{aligned}$$

So:

$$\begin{aligned}
 p_x &= \mathbb{P}_{S_x}(R | H) \times \frac{1}{2} + \mathbb{P}_{S_x}(R | T) \times \frac{1}{2} \\
 &= \mathbb{P}(R | S_{x+1}) \times \frac{1}{2} + \mathbb{P}(R | S_{x-1}) \times \frac{1}{2} \\
 p_x &= \frac{1}{2}p_{x+1} + \frac{1}{2}p_{x-1}. \quad (\star)
 \end{aligned}$$

True for $1 \leq x \leq N - 1$ with boundary conditions $p_0 = 1, p_N = 0$.

Solution of difference equation (\star) :

We have:

$$\begin{aligned}
 p_x &= \frac{1}{2}p_{x+1} + \frac{1}{2}p_{x-1} \\
 \frac{1}{2}p_x + \frac{1}{2}p_x &= \frac{1}{2}p_{x+1} + \frac{1}{2}p_{x-1} \\
 \text{Rearranging:} \quad p_{x-1} - p_x &= p_x - p_{x+1}.
 \end{aligned}$$

Boundaries: $p_0 = 1, p_N = 0$.

There are several ways of solving this equation.

1. By inspection

There are N steps to go down
from $p_0 = 1$ to $p_N = 0$.

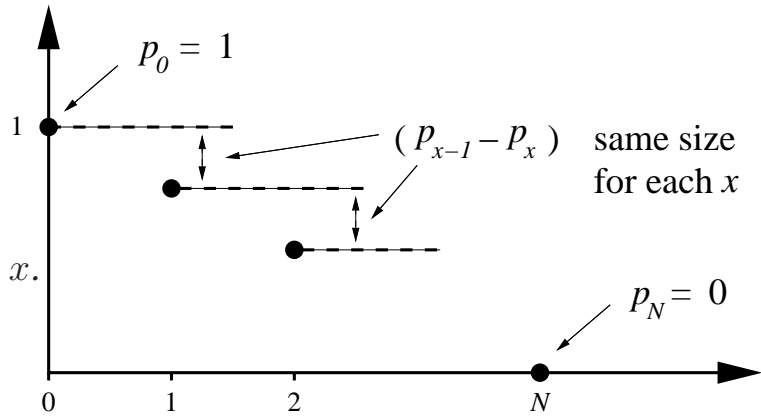
Each step is the same size,
because

$$(p_{x-1} - p_x) = (p_x - p_{x+1}) \text{ for all } x.$$

So each step has size $1/N$,

$$\Rightarrow p_0 = 1, p_1 = 1 - 1/N, \\ p_2 = 1 - 2/N, \text{ etc.}$$

So
$$p_x = 1 - \frac{x}{N}.$$



2. Theory of linear 2nd order difference equations

Theory tells us that the general solution of (\star) is $p_x = A + Bx$ for some constants A, B .

Boundary conditions:

$$p_0 = A + B \times 0 = 1 \Rightarrow A = 1$$

$$p_N = A + B \times N = 1 + BN = 0 \Rightarrow B = \frac{-1}{N}.$$

So
$$p_x = A + Bx = 1 - \frac{x}{N}$$
 as before.

3. Repeated substitution: rearrange (\star) to give:

$$p_{x+1} = 2p_x - p_{x-1}$$

$$\Rightarrow (x = 1) \quad p_2 = 2p_1 - 1 \quad (\text{recall } p_0 = 1)$$

$$(x = 2) \quad p_3 = 2p_2 - p_1 = 2(2p_1 - 1) - p_1 = 3p_1 - 2$$

$$(x = 3) \quad p_4 = 2p_3 - p_2 = 2(3p_1 - 2) - (2p_1 - 1) = 4p_1 - 3 \quad \text{etc}$$

$$\begin{array}{ll} \text{giving} & p_x = xp_1 - (x-1) \quad \text{in general,} \quad (\star\star) \\ \text{likewise} & p_N = Np_1 - (N-1) \quad \text{at endpoint.} \end{array}$$

$$\text{Boundary condition: } p_N = 0 \Rightarrow Np_1 - (N-1) = 0 \Rightarrow p_1 = 1 - 1/N.$$

Substitute in $(\star\star)$:

$$\begin{aligned} p_x &= xp_1 - (x-1) \\ &= x\left(1 - \frac{1}{N}\right) - (x-1) \\ &= x - \frac{x}{N} - x + 1 \\ p_x &= 1 - \frac{x}{N} \quad \text{as before.} \quad \square \end{aligned}$$

2.7 Independence

Definition: Events A and B are statistically independent if and only if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

This implies that A and B are statistically independent if and only if $\mathbb{P}(A | B) = \mathbb{P}(A)$.

Note: If events are *physically* independent, they will also be statistically indept.

For interest: more than two events

Definition: For more than two events, A_1, A_2, \dots, A_n , we say that A_1, A_2, \dots, A_n are mutually independent if

$$\mathbb{P}\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i) \quad \text{for ALL finite subsets } J \subseteq \{1, 2, \dots, n\}.$$

Example: events A_1, A_2, A_3, A_4 are mutually independent if

- i) $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all i, j with $i \neq j$; AND
- ii) $\mathbb{P}(A_i \cap A_j \cap A_k) = \mathbb{P}(A_i)\mathbb{P}(A_j)\mathbb{P}(A_k)$ for all i, j, k that are all different; AND
- iii) $\mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)\mathbb{P}(A_4)$.

Note: For mutual independence, it is **not** enough to check that $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ for all $i \neq j$. Pairwise independence does not imply mutual independence.

Definition: Events, A_1, A_2, \dots, A_n are pairwise independent if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \text{for all } i \neq j.$$

Pairwise independence does NOT imply mutual independence. See example in Stats 210 notes.

2.8 The Continuity Theorem

The Continuity Theorem states that probability is a *continuous set function*:

Theorem 2.8: The Continuity Theorem

- a) Let A_1, A_2, \dots be an *increasing sequence of events*: i.e.

$$A_1 \subseteq A_2 \subseteq \dots \subseteq A_n \subseteq A_{n+1} \subseteq \dots$$

Then

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

Note: because $A_1 \subseteq A_2 \subseteq \dots$, we have: $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

b) Let B_1, B_2, \dots be a *decreasing sequence of events*: i.e.

$$B_1 \supseteq B_2 \supseteq \dots \supseteq B_n \supseteq B_{n+1} \supseteq \dots$$

Then

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} B_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(B_n).$$

Note: because $B_1 \supseteq B_2 \supseteq \dots$, we have: $\lim_{n \rightarrow \infty} B_n = \bigcap_{n=1}^{\infty} B_n$.

Proof (a) only: for (b), take complements and use (a).

Define $C_1 = A_1$, and $C_i = A_i \setminus A_{i-1}$ for $i = 2, 3, \dots$. Then C_1, C_2, \dots are mutually exclusive, and $\bigcup_{i=1}^n C_i = \bigcup_{i=1}^n A_i$, and likewise, $\bigcup_{i=1}^{\infty} C_i = \bigcup_{i=1}^{\infty} A_i$.

Thus

$$\begin{aligned} \mathbb{P}\left(\lim_{n \rightarrow \infty} A_n\right) &= \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(C_i) \quad (C_i \text{ mutually exclusive}) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(C_i) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n C_i\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n). \quad \square \end{aligned}$$

2.9 Random Variables

Definition: A **random variable**, X , is defined as *a function from the sample space to the real numbers*: $X : \Omega \rightarrow \mathbb{R}$.

A random variable therefore *assigns a real number to every possible outcome of a random experiment*.

A random variable is essentially *a rule or mechanism for generating random real numbers*.

The Distribution Function

Definition: The **cumulative distribution function** of a random variable X is given by

$$F_X(x) = \mathbb{P}(X \leq x)$$

$F_X(x)$ is often referred to as simply the **distribution function**.

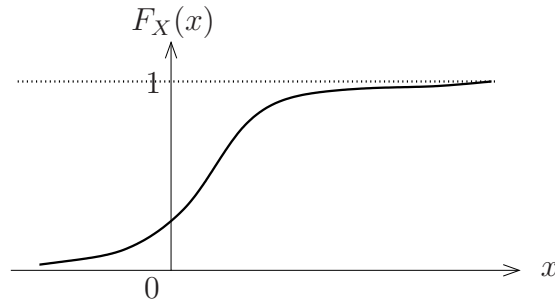
Properties of the distribution function

- 1) $F_X(-\infty) = \mathbb{P}(X \leq -\infty) = 0$.
 $F_X(+\infty) = \mathbb{P}(X \leq \infty) = 1$.
 - 2) $F_X(x)$ is a non-decreasing function of x :
if $x_1 < x_2$, then $F_X(x_1) \leq F_X(x_2)$.
 - 3) If $b > a$, then $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.
 - 4) F_X is right-continuous: i.e. $\lim_{h \downarrow 0} F_X(x + h) = F_X(x)$.
-

2.10 Continuous Random Variables

Definition: The random variable X is continuous if the distribution function $F_X(x)$ is a continuous function.

In practice, this means that a continuous random variable takes values in a continuous subset of \mathbb{R} : e.g. $X : \Omega \rightarrow [0, 1]$ or $X : \Omega \rightarrow [0, \infty)$.

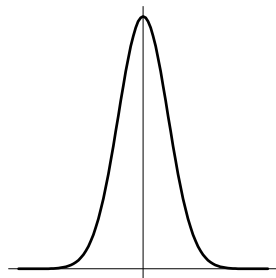


Probability Density Function for continuous random variables

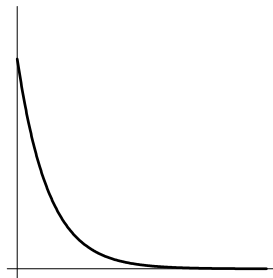
Definition: Let X be a continuous random variable with continuous distribution function $F_X(x)$. The probability density function (p.d.f.) of X is defined as

$$f_X(x) = F'_X(x) = \frac{d}{dx}(F_X(x))$$

The pdf, $f_X(x)$, gives the *shape* of the distribution of X .



Normal distribution



Exponential distribution



Gamma distribution

By the Fundamental Theorem of Calculus, the distribution function $F_X(x)$ can be written in terms of the probability density function, $f_X(x)$, as follows:

$$F_X(x) = \int_{-\infty}^x f_X(u) du$$

Endpoints of intervals

For continuous random variables, every point x has $\mathbb{P}(X = x) = 0$. This means that the endpoints of intervals are not important for continuous random variables.

Thus, $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$.

This is **only** true for **continuous** random variables.

Calculating probabilities for continuous random variables

To calculate $\mathbb{P}(a \leq X \leq b)$, use **either**

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$$

or

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$$

Example: Let X be a continuous random variable with p.d.f.

$$f_X(x) = \begin{cases} 2x^{-2} & \text{for } 1 < x < 2, \\ 0 & \text{otherwise.} \end{cases}$$

- Find the cumulative distribution function, $F_X(x)$.
- Find $\mathbb{P}(X \leq 1.5)$.

$$a) \quad F_X(x) = \int_{-\infty}^x f_X(u) du = \int_1^x 2u^{-2} du = \left[\frac{2u^{-1}}{-1} \right]_1^x = 2 - \frac{2}{x} \quad \text{for } 1 < x < 2.$$

Thus

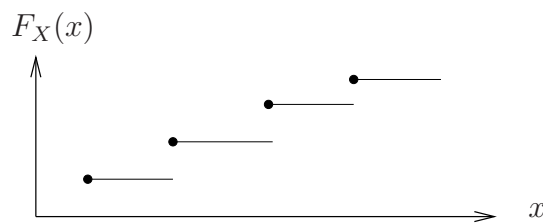
$$F_X(x) = \begin{cases} 0 & \text{for } x \leq 1, \\ 2 - \frac{2}{x} & \text{for } 1 < x < 2, \\ 1 & \text{for } x \geq 2. \end{cases}$$

$$b) \quad \mathbb{P}(X \leq 1.5) = F_X(1.5) = 2 - \frac{2}{1.5} = \frac{2}{3}.$$

2.11 Discrete Random Variables

Definition: The random variable X is **discrete** if X takes values in a finite or countable subset of \mathbb{R} : thus, $X : \Omega \rightarrow \{x_1, x_2, \dots\}$.

When X is a discrete random variable, the distribution function $F_X(x)$ is a *step function*.



Probability function

Definition: Let X be a discrete random variable with distribution function $F_X(x)$.

The **probability function** of X is defined as

$$f_X(x) = \mathbb{P}(X = x).$$

Endpoints of intervals

For discrete random variables, *individual points can have* $\mathbb{P}(X = x) > 0$.

This means that *the endpoints of intervals ARE important for discrete random variables*.

For example, if X takes values $0, 1, 2, \dots$, and a, b are integers with $b > a$, then

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a - 1 < X \leq b) = \mathbb{P}(a \leq X < b + 1) = \mathbb{P}(a - 1 < X < b + 1).$$

Calculating probabilities for discrete random variables

To calculate $\mathbb{P}(X \in A)$ for any countable set A , use

$$\mathbb{P}(X \in A) = \sum_{x \in A} \mathbb{P}(X = x).$$

Partition Theorem for probabilities of discrete random variables

Recall the Partition Theorem: for any event A , and for events B_1, B_2, \dots that form a **partition** of Ω ,

$$\mathbb{P}(A) = \sum_y \mathbb{P}(A \mid B_y) \mathbb{P}(B_y).$$

We can use the Partition Theorem to find probabilities for random variables. Let X and Y be discrete random variables.

- Define event A as $A = \{X = x\}$.
- Define event B_y as $B_y = \{Y = y\}$ for $y = 0, 1, 2, \dots$ (or whatever values Y takes).
- Then, by the Partition Theorem,

$$\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x \mid Y = y) \mathbb{P}(Y = y).$$

2.12 Independent Random Variables

Random variables X and Y are independent if they have no effect on each other. This means that the probability that they both take specified values simultaneously is the product of the individual probabilities.

Definition: Let X and Y be random variables. The joint distribution function of X and Y is given by

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x \text{ and } Y \leq y) = \mathbb{P}(X \leq x, Y \leq y).$$

Definition: Let X and Y be any random variables (continuous or discrete). X and Y are independent if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \text{ for ALL } x, y \in \mathbb{R}.$$

If X and Y are **discrete**, they are independent if and only if their joint probability function is the product of their individual probability functions:

$$\begin{aligned} \text{Discrete } X, Y \text{ are indept} \quad &\Longleftrightarrow \quad \mathbb{P}(X = x \text{ AND } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \\ &\text{for ALL } x, y \\ &\Longleftrightarrow \quad f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for ALL } x, y. \end{aligned}$$