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: $\Omega = \{$*all people in class*$\}$
Event A: $A = \{$*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:** $\mathbb{P}(A \mid B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.
- **Multiplication rule:** $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A)$.
- **The Partition Theorem:** if $B_1, B_2, \ldots, B_m$ form a partition of Ω, then

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

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

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

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

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

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

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

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

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?

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

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

•

or:

or:

We could write the above as:

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:

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

so:

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

and:

$$\mathbb{P}(B \,|\, \text{female}) =$$

Likewise, we could replace 'female' with any attribute A:

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

This is how we get the definition of conditional probability:

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

By conditioning on event A, we have

Definition: Let A and B be events on the same sample space: so
The **conditional probability of event B, given event A**, is

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

Conditioning as 'changing the sample space'

The idea that can be very helpful in understanding how to manipulate conditional probabilities.

Any 'unconditional' probability can be written as a conditional probability:

Writing $\mathbb{P}(B) = \mathbb{P}(B \mid \Omega)$ 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

Then

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

So A is just another sample space. Thus

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

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:

(Proof: Exercise).

The rules:

$$\boxed{\begin{array}{c}\mathbb{P}(\,\cdot \mid A) = \mathbb{P}_A(\,\cdot\,) \\ \mathbb{P}_A(B \mid C) = \mathbb{P}(B \mid C \cap A) \text{ for any } A,\, B,\, C.\end{array}}$$

Examples:

1. Probability of a union. In general,

$$\mathbb{P}(B \cup C) =$$

So,

Thus,

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

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

(b) $\dfrac{\mathbb{P}(B \mid C)}{\mathbb{P}(A)}$.

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

(d) $\mathbb{P}(B \mid C)\mathbb{P}(C \mid A)$.

Solution:

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:

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:

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

   Answer:

Exercise: if we wish to express $\mathbb{P}(B \,|\, A)$ in terms of only B and $\overline{A}$, show that $\mathbb{P}(B \,|\, A) = \dfrac{\mathbb{P}(B) - \mathbb{P}(B \,|\, \overline{A})\mathbb{P}(\overline{A})}{1 - \mathbb{P}(\overline{A})}$. 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

If A and B are mutually exclusive,
For all other A and B,

Definition: Any number of events $B_1, B_2, \ldots, B_k$ are **mutually exclusive** if every pair of the events is mutually exclusive: ie.

Definition: A **partition** of Ω is a

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

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

$$\textbf{and} \quad \bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots \cup B_k = \Omega.$$

$B_1, \ldots, B_k$ form a partition of Ω if they

Examples:

Partitioning an event A

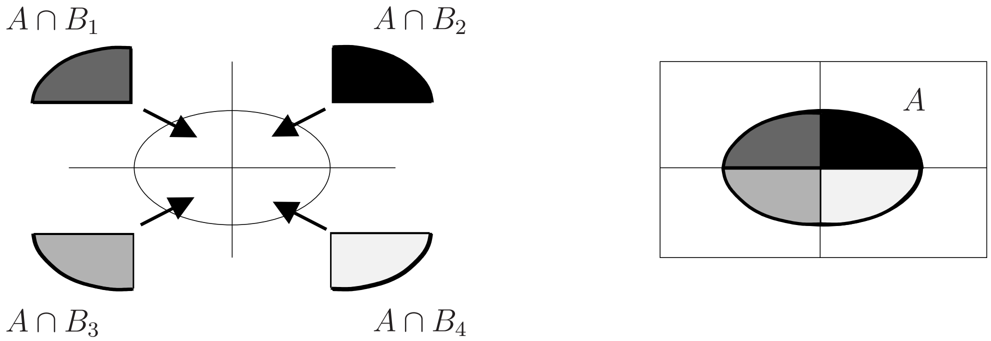
Any set A can be partitioned: it doesn't have to be Ω.
In particular, if $B_1, \ldots, B_k$ form a partition of Ω, then $(A \cap B_1), \ldots, (A \cap B_k)$ form a partition of A.

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

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

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.

Theorem 2.5: Bayes' Theorem

For any events A and B:

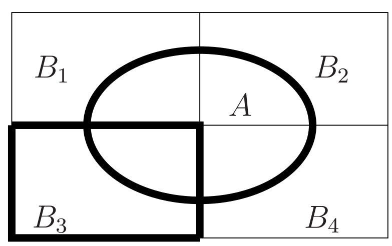
Proof:

$$\begin{aligned} \mathbb{P}(B \cap A) &= \mathbb{P}(A \cap B) \\ \mathbb{P}(B \mid A)\mathbb{P}(A) &= \mathbb{P}(A \mid B)\mathbb{P}(B) \qquad \text{(multiplication rule)} \\ \therefore \qquad \mathbb{P}(B \mid A) &= \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}. \qquad \square \end{aligned}$$

Extension of Bayes' Theorem

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

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



Special case: $m = 2$

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

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?

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

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 $S_1, \ldots, S_k$ 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:

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

Example 1: Tennis game at Deuce.

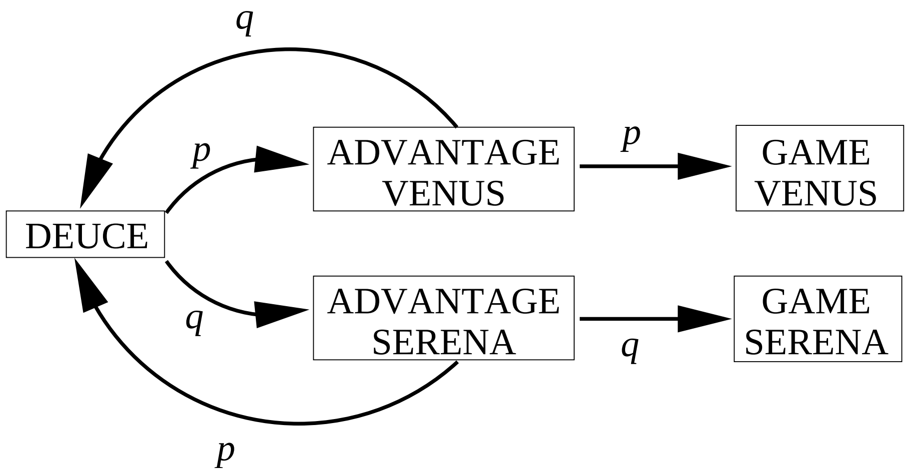
Venus and Serena are playing tennis, and have reached the score Deuce (40-40). (*Deuce* comes from the French word *Deux* for 'two', meaning that each player needs to win two consecutive points to win the game.)

For each point, let:

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

Assume that all points are independent.

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


q
p
ADVANTAGE
VENUS
p
GAME
VENUS
DEUCE
ADVANTAGE
SERENA
q
GAME
SERENA
q
p

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

So the final probability that Venus wins the game is:

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

Pair of steps	Probability	Outcome

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

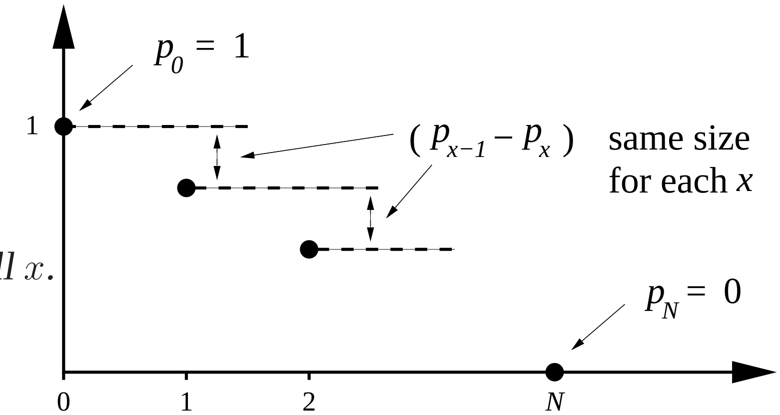
Example 2: **Gambler's Ruin.**
This is a famous problem in probability. A gambler starts with $\$x$. She tosses a fair coin repeatedly.

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)?

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
$(p_{x-1} - p_x) = (p_x - p_{x+1})$ *for all* x.
So each step has size $1/N$,
$\Rightarrow \quad p_0 = 1,\ p_1 = 1 - 1/N,$
$p_2 = 1 - 2/N$, *etc.*

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

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:

$$\begin{aligned} p_0 &= A + B \times 0 = 1 \quad \Rightarrow \quad A = 1 \\ p_N &= A + B \times N = 1 + BN = 0 \quad \Rightarrow \quad B = \frac{-1}{N}. \end{aligned}$$

So $\boxed{p_x = A + B\,x = 1 - \frac{x}{N}}$ *as before.*

2.7 Independence

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

This implies that A and B are statistically independent if and only if

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

For interest: more than two events

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

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

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, $A_1, A_2, \ldots, A_n$ are **pairwise independent** if

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

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 $A_1, A_2, \ldots$ be an *increasing sequence of events*: i.e.

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

Then

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

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

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

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

Then

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

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

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

Define $C_1 = A_1$, and $C_i = A_i \backslash A_{i-1}$ for $i = 2, 3, \ldots$. Then $C_1, C_2, \ldots$ 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

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

2.9 Random Variables

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

A random variable therefore

A random variable is essentially

The Distribution Function

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

$F_X(x)$ 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* $F_X(x_1) \leq F_X(x_2)$.

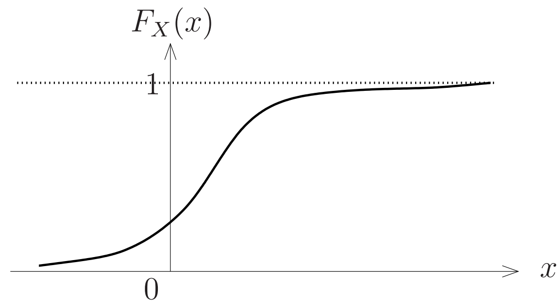
3) If $b > a$, then $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$.

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.* $X : \Omega \to [0, 1]$ *or* $X : \Omega \to [0, \infty)$.



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:

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

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***

or

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

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

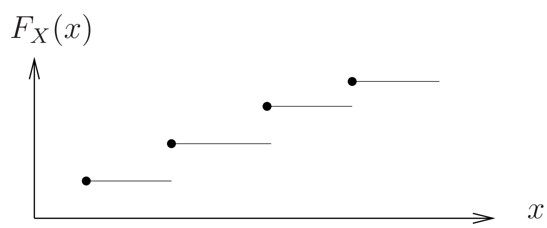
(a) Find the cumulative distribution function, $F_X(x)$.

(b) Find $\mathbb{P}\left(X \leq 1.5\right)$.

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,* $X : \Omega \to \{x_1, x_2, \ldots\}$.

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 $0, 1, 2, \ldots$, 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 $B_1, B_2, \ldots$ that form a ***partition*** of Ω,

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

-
-
-

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

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

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

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

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: