Introductory Mathematics for Computer Science (COMP0011)

Raphael Li

Year 1 Term 2, 2024–25

Contents

1 Polynomials

Given a field $\mathbb{K}$ and a variable x, a *polynomial* P of $\mathbb{K}[x]$ is defined as a linear combination of powers of x.

$$P = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2x^2 + a_1x + a_0 \qquad \text{(assuming } a_n \neq 0\text{)}$$

Note that

- The numbers a_i should all belong to the field $\mathbb{K}$. They are called *coefficients*.
- The products $a_i x^i$ are called *terms*.
- Here, n is the highest power in P. This is known as the *degree* of the polynomial.

Examples of polynomials include

$$x^2 \qquad \text{(degree } 2\text{)}$$

$$y - 1.5y^3 + 2 \qquad \text{(degree } 3\text{)}$$

$$z + 3z^4 \qquad \text{(degree } 4\text{)}$$

Although any symbol can be used to denote the variable of a polynomial, we will most use the letter x in this section.

A polynomial of degree 0 is simply a constant. On the other hand, a polynomial of degree 1, such as $2x + 1$, is said to be *linear*.

1.1 Polynomials are not functions

One important distinction to make is that technically speaking, polynomials are not functions:

- A *polynomial* is a strictly algebraic object. It is sometimes represented as a vector in a vector space. (We will further explore this idea below.)
- A *function* is a mapping — a rule that maps elements from a set to elements of another set.
- A *polynomial function* is a specific type of function. While there is a one-to-one correspondence between polynomials and polynomial functions, they do not refer to the same idea.

Despite this, polynomials can be evaluated by substituting the variable with a value. For example, given the polynomial $P = x^2 - 5x$, we can substitute $x = 7$ to get

$$P(7) = 7^2 - 5 \times 7 = 14.$$

1.2 Addition, multiplication and division of polynomials

Polynomials can be added by summing up their like terms.

$$\begin{aligned}(7x^3 + 8x^2 + 3) + (2x^2 + 9x - 4) &= 7x^3 + (8x^2 + 2x^2) + 9x + (3 - 4)\\ &= 7x^3 + 10x^2 + 9x - 1\end{aligned}$$

Polynomials can also be multiplied by an element of $\mathbb{K}$.

$$2(7x^3 + 8x^2 + 3) = 14x^3 + 16x^2 + 6$$

Let $\mathbb{K}_n[x]$ be the set of all polynomials with coefficients in K and of degree at most n. Since this set is closed under addition and scaling, it is a vector space.

We can multiply two polynomials by using expansion via distributivity.

$$
\begin{aligned}
(8x+3)(9x-4) &= (8x)(9x) + (8x)(-4) + (3)(9x) + (3)(-4) \\
&= 72x^2 - 32x + 27x - 12 \\
&= 72x^2 - 5x - 12
\end{aligned}
$$

If we denote the degree of a polynomial P as $\deg(P)$, then:

$$
\begin{aligned}
&\deg(P+Q) \leq \max(\deg(P) + \deg(Q)) && \text{(highest-degree terms may cancel out)} \\
&\deg(P \times Q) \leq \deg(P) + \deg(Q)
\end{aligned}
$$

The fact that we can multiply polynomials implies that polynomials can have *factors* — for instance, we say that the polynomial $72x^2 - 5x - 12$ has factors $8x + 3$ and $9x - 4$.

A polynomial with no non-constant factors is said to be *irreducable*[1]. For example, $7x+4$ is irreducable, but the following polynomials are not.

$$
\begin{aligned}
x^2 + 7x + 12 &= (x+4)(x+3) \\
x^3 + x^2 + 2x + 2 &= (x^2+2)(x+1)
\end{aligned}
$$

A polynomial is said to *split* if it has only linear factors. Therefore, of the two polynomials listed above, the first one splits but the second one doesn't.

Lastly, we can perform division on polynomials, as explained below.

Euclidean division of polynomials.

Given two polynomials A and B, there exists polynomials Q and R such that

$$A = QB + R$$

where R has a lower degree than B. Here,

- A is the dividend and B is the divisor.
- Q is the quotient and R is the remainder.

Given a dividend and a divisor, we can use long division to identify the corresponding quotient and remainder. An example of this is given in 1, with the division

$$x^3 - 2x^2 - 4 = (x^2 + x + 3)(x - 3) + 5.$$

Note how the remainder ($R = 5$) has a lower degree than the divisor $B = x - 3$.

$$
\begin{array}{r r}
 & x^2 + x + 3 \\
x - 3 \,) & \overline{x^3 - 2x^2 + 0x - 4} \\
 & \underline{x^3 - 3x^2 \phantom{{}+0x-4}} \\
 & +x^2 + 0x \phantom{{}-4} \\
 & \underline{+x^2 - 3x \phantom{{}-4}} \\
 & +3x - 4 \\
 & \underline{+3x - 9} \\
 & +5
\end{array}
$$

Figure 1: An example of performing long division on polynomials.

[1]This is analogous to how prime numbers work.

1.3 Roots of polynomials

A number a in $\mathbb{K}$ is said to be a root of a polynomial P if $P(a) = 0$.

For example,

- The linear polynomial $P = 5x + 2$ has the root $x = -2/5$ because $P(-2/5) = 5(-2/5) + 2 = 0$.
- The polynomial $Q = x^2 + 3x + 2$ has a root $x = -2$ because $Q(-2) = (-2)^2 + 3(-2) + 2 = 0$.

We now introduce the *factor theorem*, which is illustrated below.

Factor theorem. A number a is a root of a polynomial P if and only if $(x - a)$ is a factor of P.

Proof. We prove this statement in two directions.

($\Leftarrow$):

$$\begin{aligned} (x-a) \text{ is a factor of } P &\implies P = (x-a)Q \qquad\qquad \text{for some polynomial } Q \\ &\implies P(a) = (a-a)Q \\ &\implies P(a) = 0 \\ &\implies a \text{ is a root of } P \end{aligned}$$

($\Rightarrow$): Assume a is a root of P, so $P(a) = 0$.

We divide P by $x - a$. By Euclidean division, there exists a polynomial Q and a constant r such that

$$P = Q \cdot (x - a) + r.$$

(Recall that the remainder must have a lower degree than the divisor $x - a$. Since the divisor $(x - a)$ has degree 1, the remainder must be a constant with degree 0.)

We evaluate both sides of the equation with $x = a$.

$$\begin{aligned} P(a) &= Q(a) \cdot (a - a) + r \\ P(a) &= r \\ r &= 0 \end{aligned}$$

Hence $P = Q \cdot (x - a)$, so $x - a$ is a factor of P.

This theorem is extremely useful for finding the roots of a polynomial. For instance, we can factor the polynomial

$$\begin{aligned} 2x^3 - x^2 - 8x + 4 &= x^2(2x - 1) - 4(2x - 1) \\ &= (x^2 - 4)(2x - 1) \\ &= (x + 2)(x - 2)(2x - 1) \\ &= 2(x + 2)(x - 2)\left(x - \frac{1}{2}\right) \end{aligned}$$

to show that it has the roots -2, 2 and $1/2$. But are these the only roots?

Yes, they are. We can prove this using the following theorem.

Theorem. The number of roots[2] of a polynomial in $\mathbb{K}$ cannot exceed its degree.

[2] In this case, identical roots are counted separately. For example, the polynomial $x^2 - 2x + 1 = (x - 1)^2$ is treated as having two roots, both of value 2. We will dive deeper into this technicality later in this section when we talk about the multiplicity of a root.

Proof. We label the roots of a polynomial P as $r_1,\ r_2,\ r_3,\ \cdots,\ r_n$. Hence, by the factor theorem, we have

$$\begin{aligned} P &= (x-r_1)(x-r_2)(x-r_3)\cdots(x-r_n)Q && \text{(for some polynomial } Q\text{)} \\ \deg(P) &= \deg((x-r_1)(x-r_2)(x-r_3)\cdots(x-r_n)Q) \\ &= n + \deg(Q) \\ &> n \end{aligned}$$

Therefore $n < \deg(P)$.

Note that the theorem above implies that the number of roots of a polynomial in $\mathbb{K}$ may not necessarily equal its degree. To see why this is, we will have to take a closer look at polynomials of degree 2.

1.4 On the real roots of polynomials of degree 2

The simplest polynomial of degree 2 takes the form $P = x^2 - a$. Finding its roots is equivalent to solving the equation

$$x^2 = a$$

which has the solutions $\sqrt{a}$ and $-\sqrt{a}$.

In general, however, a polynomial of degree 2 takes the form $ax^2 + bx + c$. The corresponding function $f(x) = ax^2 + bx + c$ is quadratic — its graph is a parabola. The roots of the polynomial correspond to the points where the graph crosses the x-axis. See figures 2, 3 and 4.

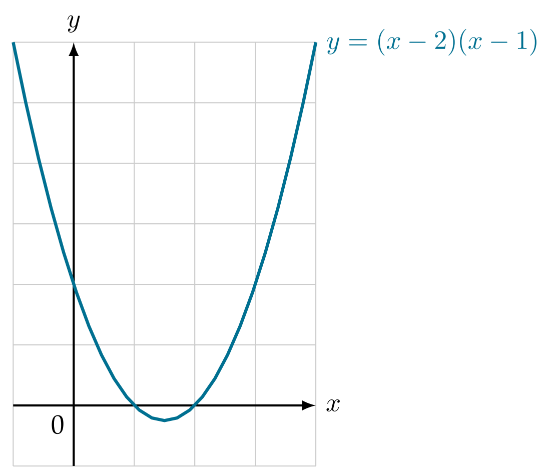

Figure 2: An example of a polynomial of degree 2 with 2 distinct roots. The corresponding function has a graph has two x-intercepts.

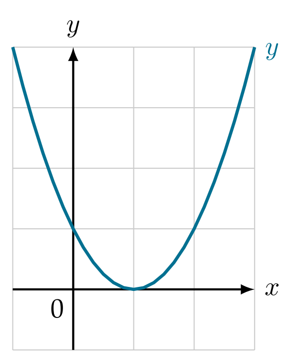

Figure 3: An example of a polynomial of degree 2 with 1 root. The corresponding function has a graph has one x-intercept.

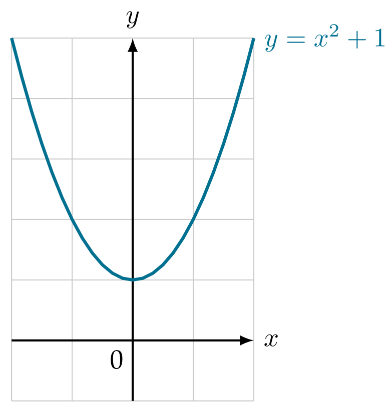

Figure 4: An example of a polynomial of degree 2 with no roots in $\mathbb{R}$. The corresponding function has a graph has no x-intercepts.

Usually, we want to work out the number of roots of a polynomial of degree 2 without plotting its graph. This can be done by analysing its *discriminant*. For a polynomial $P = ax^2 + bx + c$, its determinant is defined as $\Delta = b^2 - 4ac$.

- If $\Delta > 0$, then P has two distinct roots in $\mathbb{R}$ and can be factored into the form $P = a(x-r_1)(x-r_2)$.
- If $\Delta = 0$, then P has one root[3] in $\mathbb{R}$ and can be factored into the form $P = a(x-r)^2$.
- If $\Delta < 0$, then P has no roots in $\mathbb{R}$ and is irreducable.

In the first two cases, the root(s) of P are given by the quadratic formula, as shown below.

$$x = \frac{-b \pm \sqrt{\Delta}}{2a} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

1.5 On the complex roots of polynomials of degree 2

If we allow complex roots, then a polynomial of degree 2 always have two (not necessarily distinct) roots r_1 and r_2 in $\mathbb{C}$, as given by the quadratic formula. In other words, every polynomial of degree 2 splits in $\mathbb{C}$ and can be factored into the form $a(x - r_1)(x - r_2)$.

The two roots of a degree 2 polynomial

$$\begin{aligned} r_1 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \\ r_2 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

are complex conjugates. This is proved below.

Theorem. The two roots of a degree 2 polynomial must be complex conjugates.

Proof. If $\Delta \geq 0$, then both roots are real and therefore must be conjugates in $\mathbb{C}$.

If $\Delta < 0$, then $\sqrt{\Delta}$ is purely imaginary and can be expressed as di for some $d \in \mathbb{R}$. Hence,

$$\begin{aligned} r_1 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b + di}{2a} = \frac{-b}{2a} + \frac{d}{2a}i \\ r_2 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b - di}{2a} = \frac{-b}{2a} - \frac{d}{2a}i \end{aligned}$$

[3] This is technically two non-distinct roots.

are complex conjugates.

For example, the polynomial x^2+1 has no real roots but can be factored in $\mathbb{C}$ as $(x-i)(x+i)$. Its roots, i and $-i$, are complex conjugates.

Below shows three useful identities for factoring polynomials of degree 2.

$$\begin{aligned}(a+b)^2 &= a^2+b^2+2ab \\ (a-b)^2 &= a^2+b^2-2ab \\ (a+b)(a-b) &= a^2-b^2\end{aligned}$$

1.6 On the roots of polynomials of arbitrary degree

We introduce the following theorems for polynomials of abritrary degree.

- **Theorem on factorisations in $\mathbb{R}$.**

  In $\mathbb{R}$, only polynomials of degree 2 are irreducable[4]. Therefore, every polynomial $P \in \mathbb{R}[x]$ of degree $n>0$ has a unique factorisation in $\mathbb{R}$ of the form

  $$P = c\underbrace{(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)}_{\text{linear factors}}\underbrace{(x^2+a_1x+b_1)(x^2+a_2x+b_2)\cdots(x^2+a_kx+b_k)}_{\text{quadratic factors}}$$

  where

  - the constants $c,\ \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_m,\ a_1,\ a_2,\ \cdots a_k,\ b_1,\ b_2,\ \cdots b_k$ are real numbers.
  - Each quadratic factor is irreducable with a negative determinant, i.e. $\Delta_i = a_i^2 - 4b_i < 0$ for $1 \le i \le k$.

- **Theorem on factorisations in $\mathbb{C}$.**

  Each polynomial $P \in \mathbb{R}[x]$ of degree $n>0$ has a unique factorisation in $\mathbb{C}$ of the form

  $$P = c(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$$

  where $c, \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n \in \mathbb{C}$.

  In other words, every real polynomial splits in $\mathbb{C}$ with n (not necessarily distinct) complex roots $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$.

When roots are not distinct, a polynomial can be written as

$$P = c(x-\lambda_1)^{k_1}(x-\lambda_2)^{k_2}(x-\lambda_3)^{k_3}\cdots(x-\lambda_j)^{k_j}$$

where $k_1,\ k_2,\ k_3,\ \cdots,\ k_j \ge 1$. We say that k_i is the *multiplicity* of λ_i.

1.7 A theorem on real polynomials of odd degrees

Finally, we introduce an interesting theorem regarding real polynomials of odd degrees.

Theorem.

Any polynomial P of odd degree and with real coefficients has at least one real root.

To prove this, we can either use algebra or calculus.

[4]Note that it is still possible for polynomials of degree greater than 2 to have no real roots. This occurs when they are made entirely of irreducable quadratic factors. For example, the polynomial x^4+1 can be factored into $x^4+2x^2+1-2x^2 = (x^2+1)^2-(\sqrt{2}x)^2 = (x^2+2\sqrt{x}+1)(x^2-2\sqrt{x}+1)$.

- **Proof 1: An algebraic proof.**

  We first prove the following lemma: If z is a complex and non-real root of P, then so is its conjugate $\bar{z}$.

  We write P as follows:

  $$P = c\underbrace{(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)}_{\text{linear factors}}\underbrace{(x^2+a_1x+b_1)(x^2+a_2x+b_2)\cdots(x^2+a_kx+b_k)}_{\text{quadratic factors}}$$

  where the constants $c,\ \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_m,\ a_1,\ a_2,\ \cdots a_k,\ b_1,\ b_2,\ \cdots b_k$ are real numbers; and each quadratic factor is irreducable with a negative determinant.

  We assume z is a root of P, which means that $(x-z)$ is a factor. Since $z \notin \mathbb{R}$, we have $z \neq \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_m$. Hence, $(x-z)$ must be a factor of one of the quadratic factors, i.e. $(x^2+a_jx+b_j)$ where $1 \leq j \leq k$.

  We know that the two roots of a quadratic polynomial are always complex conjugates (This was proved earlier using the quadratic formula.) Therefore, $(x-\bar{z})$ is also a factor of $(x^2+a_jx+b_j)$.

  This means that $(x-\bar{z})$ is a factor of P, so $\bar{z}$ is a root of P. This concludes the proof for the lemma.

  We now prove the required theorem. Since P has an odd degree, it must also have an odd number of roots. By the previously proved lemma, for any complex and non-real root z of P, its conjugate $\bar{z} \neq z$ is also a root. Hence, the number of non-real roots of P must be even. This means that the number of real roots of P must be odd, i.e. at least one. Hence proved.

- **Proof 2: A calculus proof.**

  We write P as follows:
  $$P = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \qquad (a_n \neq 0)$$
  where n is odd. WLOG assume that $a_n > 0$. This means that as x approaches positive infinity, so does P. Hence
  $$\forall d > 0, \exists c > 0, x > c \implies P(x) > d.$$
  Substituting $d = 1$ (or any positive value) tells us that there exists some positive constant c such that $x > c \implies P(x) > 1$. This means that there exists some x_{pos} for which $P(x_{\text{pos}})$ is positive.

  Similarly, notice that as x approaches negative infinity, so does P (since n is odd). This means that
  $$\forall d < 0, \exists c < 0, x > c \implies P(x) < d.$$
  Substituting $d = -1$ (or any negative value) tells us that there exists some negative constant c such that $x < c \implies P(x) < -1$. This means that there exists some x_{neg} for which $P(x_{\text{neg}})$ is negative.

  Combining the two results above with the intermediate value theorem, we see that there must exist some value x_0 such that $x_{\text{neg}} < x_0 < x_{\text{pos}}$ and $P(x_0) = 0$. This x_0 is a real root of P. Hence proved.

2 Probability

Probability is used to model real-life events and their likelihoods with mathematical tools.

When dealing with probabilities, we want to consider a random process and its possible outcomes. For instance, rolling a dice is a random process that produces six outcomes.

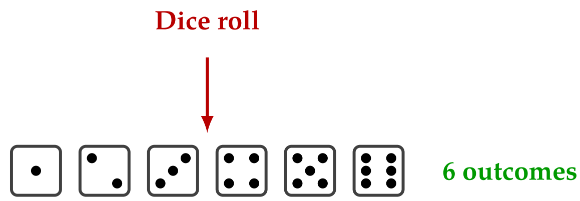

Figure 5: Rolling a dice is a random process that generates six possible outcomes.

The set of all possible outcomes is called the *sample space* or *universe*, which we denote by Ω.

$$\Omega = \{\text{rolling } 1,\ \text{rolling } 2,\ \text{rolling } 3,\ \text{rolling } 4,\ \text{rolling } 5,\ \text{rolling } 6\}$$

Any subset of Ω is called an event. For example, events associated with rolling a dice include the following.

$$A = \{2, 4, 6\} \qquad \text{(rolling an even number)}$$
$$B = \{1, 3, 5\} \qquad \text{(rolling an odd number)}$$
$$C = \{5, 6\} \qquad \text{(rolling a number greater than 4)}$$

This allows us to treat events as sets:

$$\text{“Rolling an even number greater than 4”} = A \cap C \qquad \text{(intersection)}$$
$$\text{“Rolling a number that is odd or greater than 4”} = B \cup C \qquad \text{(union)}$$
$$\text{“Rolling a number not greater than 4”} = \overline{C} \qquad \text{(complement)}$$
$$\text{“Rolling an odd number not greater than 4”} = B \setminus C \qquad \text{(minus)}$$

(The complement of a set S is also sometimes denoted as A^c, but here we will stick to the notation $\overline{S}$.)

We say that two events A and B are *disjoint* if and only if their intersection is the empty set.

$$A \cap B = \emptyset$$

2.1 Probability laws

We represent the likelihood of an event E using its probability $P(E)$. We have the following laws.

$$P(A) \geq 0 \qquad \text{(probabilities are non-negative)}$$
$$P(\Omega) = 1 \qquad \text{(whole sample space has probability 1)}$$
$$P(A \cup B) = P(A) + P(B) \quad \text{if } A, B \text{ disjoint} \qquad \text{(additivity)}$$

This has several consequences, which we will prove below. (Drawing set diagrams are really useful in constructing these proofs.)

Theorem. If $A \subseteq B$, then $P(A) \leq P(B)$.

Intuition. Event B “includes” event A.

Proof. Assume $A \subseteq B$. Let $C = B \setminus A$. Since A and C are disjoint, we have

$$\begin{aligned}
P(B) &= P(A \cup C) && \text{(by definition)} \\
&= P(A) + P(C) && \text{(by additivity)} \\
&\geq P(A). && \text{($P(C)$ must be non-negative)}
\end{aligned}$$

Hence proved.

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

Proof. Let $A' = A \setminus B$ and $B' = B \setminus A$. It follows that

$$\begin{aligned}
\text{RHS} &= P(A) + P(B) - P(A \cap B) \\
&= P(A) + P(B' \cup (A \cap B)) - P(A \cap B) \\
&= P(A) + P(B') + P(A \cap B) - P(A \cap B) && \text{(B' and $A \cap B$ are disjoint)} \\
&= P(A) + P(B') \\
&= P(A \cup B') && \text{(A and B' are disjoint)} \\
&= P(A \cup B) \\
&= \text{LHS}
\end{aligned}$$

Hence proved.

Theorem. $P(\overline{A}) = 1 - P(A)$.

Proof. Since A and $\overline{A}$ are disjoint, we have

$$\begin{aligned}
P(A \cup \overline{A}) &= P(A) + P(\overline{A}) \\
P(\Omega) &= P(A) + P(\overline{A}) \\
1 &= P(A) + P(\overline{A}) \\
P(\overline{A}) &= 1 - P(A)
\end{aligned}$$

Hence proved.

2.2 Discrete probabilities

A set is said to be *countable* if it is finite or in bijection with $\mathbb{N}$.

If the sample space of a random process is countable, then we can measure probabilities in that space as a sum.

$$P(A) = \sum_{a \in A} P(\{a\})$$

This is a direct consequence of the law of additivity for disjoint events.

2.2.1 Example: Finite sample spaces

For example, for our example of dice rolling, let C be the event of rolling a number greater than 4. This event consists of two possible outcomes: rolling a 5 and rolling a 6. Therefore,

$$\begin{aligned} P(C) &= P(\{\text{rolling } 5,\ \text{rolling } 6\}) \\ &= P(\{\text{rolling } 5\}) + P(\{\text{rolling } 6\}) \\ &= \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3} \end{aligned}$$

Furthermore, in the case where Ω has a finite size n with every outcome equally likely (e.g. dice rolling), we can express the probability of any event A as

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

This streamlines the calculation of $P(C)$ above as follows.

$$P(C) = P(\{\text{rolling } 5,\ \text{rolling } 6\}) = \frac{2}{6} = \frac{1}{3}$$

2.2.2 Countably infinite sample spaces

Now consider a different scenario with an infinite, but nevertheless countable sample space.

> A fair coin is tossed repetitively until heads is observed. The number of coin tosses is recorded as the outcome of this experiment. The sample space Ω of this process is thus $\mathbb{N}$, which is countably infinite.
>
> Verify that $P(\Omega) = 1$.

Since the sample space is countably infinite, we can express $P(\Omega)$ as an infinite sum.

$$\begin{aligned} P(\Omega) &= \Sigma_{a\in\Omega}\ P(\{a\}) \\ &= P(\{1\}) + P(\{2\}) + P(\{3\}) + \cdots \\ &= P(\text{First head on toss \#1}) + P(\text{First head on toss \#2}) + P(\text{First head on toss \#3}) + \cdots \\ &= \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} + \cdots \\ &= \sum_{k\geq 1} \frac{1}{2^k} \\ &= 1 && \text{(geometric series)} \end{aligned}$$

2.3 Continuous probabilities

If the sample space is instead *uncountable* (e.g. intervals of $\mathbb{R}$), then we can only measure probabilities as a continuous sum, i.e. an integral.

For example, consider a random number x in the interval $[0, 1]$. The probabiliy that x is strictly higher than 0.7 is given by

$$P(x > 0.7) = \int_{0.7}^{1} dx = 0.3.$$

2.4 Conditional probabilities