NYU Math Camp Lecture Notes

Patrick Le Bihan

February 16, 2012

Contents

1 Notation, Functions, and Calculus

1.1 Elements of Set Theory

1.1.1 Basic Definitions

Definition 1.1. *A **set** is a collection of distinguishable objects. An object can be anything: a number, the name of a city, a book, etc. An object that is contained in a set S is called an **element** of S. We require that for any possible object, we can determine whether the object is an element of S or not. In other words, it is not possible for an object to be only partly an element of a set.*

We write $x \in S$ (or $S \ni x$) to say that an object x is an element of S and $x \notin S$ (or $S \not\ni x$) when we want to signify that x does not belong to S. Similarly, $x, y, z \in S$ means that x, y, and z are in S. We can define a set either by listing its elements or by characterizing the property (or properties) that an object needs to satisfy to be part of the set. Here are some examples:
$S := \{1, 2, 3\}$ which reads as 'S is the set containing the numbers 1,2,3', where := means 'is defined to be equal' or 'is equal by definition'. Some authors use $\equiv$ instead.
$S := \{blue, green, black\}$ 'S is the set of colors blue, green and black'.
$S := \{x : 1 < x < 3\}$ which reads as 'S is the set of x's such that x is strictly greater than 1 and strictly lower than 3.'
$S := \{x : x \text{ is a member state of the European Union}\}$

Exercise 1.1. *Express the following sets in terms of the property or properties of its elements.*

(a) $S :=$ {*Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune*}.

(b) $S :=$ {*Brown, Columbia, Cornell, Dartmouth, Harvard, Princeton, University of Pennsylvania, Yale*}

(c) $S :=$ {*Alabama, Alaska, Arizona, ..., Wisconsin, Wyoming*}.

(d) $S :=$ {*ManU, FC Arsenal, FC Chelsea, FC Liverpool, Tottenham hotspurs, ...*}.

(e) $S := \{1, 2, 4, 8, 16, 32, 64, \ldots\}$

(f) $S := \{10, 11, 12, 13, 14, 15, 16, 17, 18, 19\}$.

We usually denote a set by a capital letter A, B, etc. and elements using lower-case letters a, b, etc.

Definition 1.2. *We say that two sets A and B are **equal** and write $A = B$ when A and B have exactly the same elements. In other words $A = B$ iff (iff is short for if and only if) for all x, $x \in A$ implies $x \in B$ and $x \in B$ implies $x \in A$. If A and B are not equal, we write $A \neq B$.*

Example 1.2. $\{1, 2, 3\} = \{3, 1, 2\}$, $\{x, x\} = \{x\}$, *but* $\{1, 2\} \neq \{1, 2, 3\}$ *and* $\{\{x\}\} \neq \{x\}$.

Definition 1.3. *We say that A is a* ***subset*** *of B and write $A \subseteq B$ if for all x, $x \in A$ implies $x \in B$. We say that B is a* ***superset*** *of A and write $B \supseteq A$ if A is a subset of B. We say that A is a* ***proper subset*** *of B and write $A \subset B$ if A is a subset of B and B is not a subset of A.*

Example 1.3. $\{1, 2\} \subset \{1, 2, 3\}$, $\{1, 2, 3\} \subseteq \{1, 2, 3\}$.

Remark 1.4. *$A = B$ iff $A \subseteq B$ and $B \subseteq A$.*

Definition 1.4. *We say that a set is empty if it does not contain any element. The* ***empty set*** *is denoted by $\emptyset$. Formally, it can be defined as $\{x : x \neq x\}$. If $A \neq \emptyset$, then we say that A is* ***nonempty****.*

Remark 1.5. *(a) $\emptyset \subseteq A$ for any set A.*

(b) The empty set is unique. Quick proof: *Let X and Y be two empty sets. Then $X \subseteq Y$ and $Y \subseteq X$, hence $X = Y$.*

(c) $\{\emptyset\} \neq \emptyset$. Something that contains something that does not contain anything is not the same thing as something that contains nothing.

Definition 1.5. *We say that a set A is* ***finite*** *if it has finitely many elements. The total number of elements that A contains is denoted by $|A|$. This number is called the* ***cardinality*** *of A. If $|A| = 1$, then we say that A is a* ***singleton****. A set A is* ***infinite****, if A contains infinitely many elements, in which case we write $|A| = \infty$.*

Definition 1.6. *The* ***power set*** *of a set A, denoted 2^A, is the collection of all subsets of A. Formally, $2^A := \{B : B \subseteq A\}$. Note that the power set is a set that contains sets.*

Example 1.6. *Let $A := \{1, 2, 3\}$. Then $2^A = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Note that 2^A has 2^3 elements. In fact, for any finite set A, the number of elements of the power set is $2^{|A|}$. This fact motivates the notation 2^A for the power set.*

Exercise 1.7. *(a) What is the power set of $\{4, 2, 5\}$?*

(b) Describe $2^\emptyset$, $2^{2^\emptyset}$, $2^{2^{2^\emptyset}}$.

1.1.2 Set Operations

Definition 1.7. *The* ***intersection*** *between A and B, denoted $A \cap B$, is defined as the set $\{x : x \in A$ and $x \in B\}$. We say that two sets A and B are* ***disjoint*** *if $A \cap B = \emptyset$.*

Example 1.8. *(a) $\{1, 2, 3\} \cap \{1, 2, 4\} = \{1, 2\}$.*

(b) $\{1, 2\} \cap \{4, 6\} = \emptyset$. Hence, $\{1, 2\}$ and $\{4, 6\}$ are disjoint sets.

Definition 1.8. *The* ***union*** *of A and B, denoted $A \cup B$, is defined as the set $\{x : x \in A$ or $x \in B\}$.*

Example 1.9. $\{1,2,3\} \cup \{1,2,4\} = \{1,2,3,4\}$.

Remark 1.10. *Note that in mathematics or is an inclusive or. Hence, $x \in A$ or $x \in B$ corresponds to any of the following three cases: (i) $x \in A$ and $x \notin B$, (ii) $x \in B$ and $x \notin A$, and (iii) $x \in A$ and $x \in B$. An exclusive or would rule out the third possibility. It follows from this remark that $A \cap B \subseteq A \cup B$.*

Definition 1.9. *The **difference** between A and B, denoted $A \setminus B$, is defined as the set $\{x : x \in A \text{ and } x \notin B\}$. If $A \subseteq B$, then $B \setminus A$ is often called the **complement** of A with respect to B. If the set with respect to which the complement is taken is clear we also denote the complement of A simply as A^c.*

Example 1.11. *(a)* $\{1,2,3,4\} \setminus \{1,2,3\} = \{4\}$.

(b) $\{4,5,8\} \setminus \{3,4,6\} = \{5,8\}$

Note that we can generalize some of these set operations to more than just two sets. To see how, let I be a so-called index set. For example $I := \{1,\dots,n\}$ and for each $i \in I$ let A_i be a set. We can then define the family or class of sets $\mathcal{A} := \{A_i : i \in I\}$. The union of all members of this class, denoted $\cup\mathcal{A}$, is then defined to be equal to $\{x : \text{ there exists } i \in I \text{ with } x \in A_i\}$. Similarly $\cap\mathcal{A}$ is defined by $\{x : \text{ for all } i \in I, x \in A_i\}$. Note that we sometimes write $\cup_{i\in I} A_i$ and $\cap_{i\in I} A_i$ instead of $\cup\mathcal{A}$ and $\cap\mathcal{A}$ respectively.

Exercise 1.12. *Let* $P :=$ {J. S. Bach, Goethe, David Hume, Mozart, Newton, George Washington} *and let* $Y := \{1750, 1751, 1752, \dots, 1759\}$. *For any* $y \in Y$, *let* $A_y := \{p \in P : \textit{ the person } p \textit{ was alive at some time during the year } y\}$. *Compute* $\cup_{y\in Y} A_y$ *and* $\cap_{y\in Y} A_y$.

Proposition 1.13 (Properties of Set operations)**.** *For any set A, B, and C we have the following properties:*

(a) Commutativity:
$A \cap B = B \cap A$ *and* $A \cup B = B \cup A$

(b) Associativity:
$A \cap (B \cap C) = (A \cap B) \cap C$ *and* $A \cup (B \cup C) = (A \cup B) \cup C$.

(c) Distributivity:
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ *and* $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

1. *(d) De Morgan Laws:*
$A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$ *and* $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

Remark 1.14. *If $A \subseteq B$, then $A \cap B = A$ and $A \cup B = B$.*

Exercise 1.15. *Let* $S := \{x : 1 < x < 3\}$ *and* $T := \{x : 2 \le x \le 4\}$, $U := \{x : 0 \le x < 2\}$

(a) Compute $S \cap T$, $S \cup T$, $S \setminus T$, $T \setminus S$, $S \setminus (T \cup U)$, $S \setminus (T \cap U)$.

(b) *Verify that* $S \cap T \subseteq S \cup T$.

Definition 1.10. *An* ***ordered pair*** *is an ordered list* (a, b) *consisting of two objects* a *and* b. *The pair is ordered in the sense that for any two ordered pairs* (a, b) *and* (a', b'), *we have* $(a, b) = (a', b')$ *iff* $a = a'$ *and* $b = b'$. *In fact it is not in general the case that* (a, b) *is equal to* (b, a). *The only case where* $(a, b) = (b, a)$ *is when* $b = a$

Definition 1.11. *The* ***Cartesian product*** *between* A *and* B, *denoted* $A \times B$, *is the set of all ordered pairs* (a, b) *where* a *comes from* A *and* b *comes from* B. *Or formally*

$$A \times B := \{(a, b) : a \in A \text{ and } b \in B\}.$$

Example 1.16. $\{1, 2\} \times \{2, 3, 4\} = \{(1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4)\}$.

Proposition 1.17. *Suppose* A, B, *and* C *are sets.*

(a) $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

(b) $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

(c) $A \times \emptyset = \emptyset \times A = \emptyset$.

Remark 1.18. *Note that the Cartesian product operation is not commutative, i.e.* $A \times B$ *need not be equal to* $B \times A$. *In fact,* $A \times B = B \times A$ *iff either* $A = \emptyset$, $B = \emptyset$, *or* $A = B$.

Exercise 1.19. *Let* $A := \{3, 5, 6\}$, $B := \{2, 4, 8\}$, $C := \{1\}$.

(a) *Compute* $A \times B$, $B \times A$, $A \times C$, $C \times A$.

(b) *Compute* $A \times (B \cap C)$ *and* $(A \times B) \cap (A \times C)$ *and check that these sets are equal.*

(c) *Compute* $A \times (B \cup C)$ *and* $(A \times B) \cup (A \times C)$ *and check that these sets are equal.*

Exercise 1.20.

(a) *Let* $A := \{x \in \mathbb{N} : 1 \leq x \leq 7\}$, $B := \{y : y \text{ is a prime number}\}$, *and* $C := \{z : z = 2k, k \in \mathbb{N}\}$. *Compute* $A \cap B$, $A \cap C$, $B \cap C$, $A \cap B \cap C$, $A \cup B$, $A \setminus B$ *and* $(A \setminus C) \cap B$.

(b) *Prove that* $(A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$.

(c) *Let* $M := \{a, b, c\}$. *Compute* 2^M.

(d) *Let* $A := \{0, 1\}$. *Compute* 2^A, 2^{A^A}, $2^A \setminus A$. *Explain the difference between* 0, $\{0\}$, $\emptyset$, *and* $\{\emptyset\}$.

(e) *Let* $M_1 := \{a, b, c\}$ *and* $M_2 := \{1, 2, 3\}$. *Compute* $M_1 \times M_2$ *and* $M_2 \times M_1$.

(f) *In a class 13 students do not live in the dorms and 8 students are from New Jersey. 17 students are from New Jersey or are not living in the dorms. How many students from New Jersey are living in the dorms?*

(g) *Let* A, B, *and* C *be subsets of the reals with* $A := \{x : 1 \leq x \leq 5\}$, $B := \{x : 3 < x \leq 7\}$, $C := \{x : x \leq 0\}$. *Compute the following sets* A^c, $A \cup B$, $B \cap C^c$, $A^c \cap B^c \cap C^c$, $(A \cup B) \cap C$.

1.2 The Number System

1.2.1 Natural Numbers, Integers, Rationals, and Real Numbers

Definition 1.12. *The numbers $1, 2, 3, 4, 5, 6, ...$ are called* ***natural numbers****. The set of all natural numbers is denoted $\mathbb{N}$. Formally,*

$$\mathbb{N} := \{1, 2, 3, 4, 5, 6, ...\}.$$

Note that for any numbers $a, b \in \mathbb{N}$, we have $a + b$ and $a \cdot b \in \mathbb{N}$. However, $a - b$ need not be in $\mathbb{N}$. For example, $3 - 5$ is not a natural number.

Definition 1.13. *If the set $\mathbb{N}$ is augmented by the number zero and by the negatives of the natural numbers, then we get the set of all* ***integers****:*

$$\mathbb{Z} := \{..., -3, -2, -1, 0, 1, 2, 3, ...\}.$$

Note that if $a, b \in \mathbb{Z}$, then $a + b$, $a \cdot b$ and $a - b$ are in $\mathbb{Z}$ as well. However, a/b need not be in $\mathbb{Z}$. For example, $1/2$ is not an integer.

Here are two particular kinds of natural numbers that are worth knowing.

Definition 1.14. *An integer n is said to be an* ***even number*** *if there exists an integer m such that $n = 2m$. If n is not even, then we say that it is an* ***odd*** *number.*

Definition 1.15. *A natural number m is said to be a* ***prime number*** *if for any $a, b \in \mathbb{N}$ such that $m = a \cdot b$ we have $a = 1$ or $b = 1$. The first six prime numbers are $1, 2, 3, 5, 7, 11$.*

Remark 1.21. *In some texts, 0 is considered to be a natural number as well. In these notes, however, the symbol $\mathbb{N}$ will always exclude 0. If we want to denote the set $\{0, 1, 2, 3, 4, ...\}$ we will use the notation $\mathbb{Z}_+$. (Some authors use $\mathbb{N}_0$ instead of $\mathbb{Z}_+$).*

Definition 1.16. *To make sure that a/b is well-defined we extend the set of integers to the set of all quotients of integers:*

$$\mathbb{Q} := \{\frac{a}{b} : a, b \in \mathbb{Z} \text{ and } b \neq 0\}.$$

This set $\mathbb{Q}$ is called the set of ***rational numbers****. Note that all the usual arithmetic operations are well-defined on $\mathbb{Q}$, i.e. for all $a, b \in \mathbb{Q}$, we have $a + b$, $a - b$, $a \cdot b$ and $\frac{a}{b}$ in $\mathbb{Q}$. Division of course is only defined if $b \neq 0$.*

Unfortunately (or maybe not), not all numbers can be written as quotients of integers. For example, $\sqrt{2}$ is not a rational number. In other words, if we were to put all the rational numbers on a line there would be holes between the rational numbers. The numbers that close these holes are called the **irrational numbers**. The irrational numbers include all the numbers that are not rational. One way to identify irrational numbers is by their decimal expansions. Numbers whose decimal expansions stops after a finite number of digits (like 0.23 or -2,347261) or repeat the same pattern indefinitely from some point on (like $1.3333333\cdots$ or $3.247247247\cdots$) are rational numbers. Numbers however whose decimal expansions never end and have no repeating pattern are irrational numbers.

Definition 1.17. *The set of all rational and irrational numbers is called the set of* ***real numbers*** *and is denoted* $\mathbb{R}$*. Note that all the usual arithmetic operations are well-defined on* $\mathbb{R}$ *as well.*

Remark 1.22. $\mathbb{N} \subset \mathbb{Z}_+ \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

There are many properties that $\mathbb{R}$ satisfies but $\mathbb{Q}$ does not. The one that really differentiates $\mathbb{R}$ from $\mathbb{Q}$ is called the **completeness axiom**. To be able to state this axiom correctly we first need some more vocabulary.

Definition 1.18. *A nonempty set* $S \subseteq \mathbb{R}$ *is said to be* ***bounded from above*** *if there exists a real number* a *such that for all* $s \in S$*, we have* $s \leq a$*. The real number* a *is then called an* ***upper bound*** *of* S*. A nonempty set* $S \subseteq \mathbb{R}$ *is said to be* ***bounded from below*** *if there exists a real number* a *such that for all* $s \in S$*, we have* $a \leq s$*. The real number* a *is called a* ***lower bound*** *of* S*. A nonempty set* $S \subseteq \mathbb{R}$ *is said to be* ***bounded*** *if it is bounded from above and from below.*

Definition 1.19. *The real number* s^* *is said to be the* ***supremum*** *of* $\emptyset \neq S \subseteq \mathbb{R}$ *if it satisfies the following two conditions:*

i. For all $s \in S$*, we have* $s \leq s^*$*.*

ii. For any $b \in \mathbb{R}$ *such that* $s \leq b$ *for all* $s \in S$*, we have* $s^* \leq b$*.*

We denote the supremum of S *by* $\sup S$*. The number* $\sup S$ *is also called the* ***least upper bound*** *of* S*.*
The real number s^* *is said to be the* ***infimum*** *(or least upper bound) of* $\emptyset \neq S \subseteq \mathbb{R}$ *if it satisfies the following two conditions:*

i. For all $s \in S$*, we have* $s^* \leq s$*.*

ii. For any $b \in \mathbb{R}$ *such that* $s \geq b$ *for all* $s \in S$*, we have* $s^* \geq b$*.*

We denote it $\inf S$*. The number* $\inf S$ *is also called the* ***greatest lower bound*** *of* S*.*

Note that by requirement (i.) the supremum is an upper bound of S. Moreover, by (ii.) it is the smallest such upper bound, in the sense that for any other upper bound b, we have $\sup S \leq b$. A similar remark holds for the infimum.

Example 1.23. *Consider the set* $S := \{1/n : n \in \mathbb{N}\}$*. Note that* $1 \geq 1/n$ *for all* $n \in \mathbb{N}$*. Hence,* 1 *is an upper bound of* S*. Now as* $1 \in S$*,* $t < 1$ *cannot be an upper bound of* S*. It follows that* $\sup S = 1$*. Similarly,* $0 \leq 1/n$ *for all* $n \in \mathbb{N}$ *and thus* 0 *is a lower bound of* S*. Now for any* $t > 0$*, there exists* $n \in \mathbb{N}$ *such that* $1/n < t$*. Hence,* t *cannot be a lower bound of* S*. It follows that* $\inf S = 0$*.*

Two concepts that are closely related to the notions of supremum and infimum are the minimum and the maximum of a set.

Definition 1.20. *Let S be a nonempty subset of $\mathbb{R}$. The* ***maximum*** *of S, if it exists, is a number $\bar{s} \in S$ such that $\bar{s} \geq s$ for all $s \in S$. The* ***minimum*** *of S, if it exists, is a number $\underline{s} \in S$ such that $\underline{s} \leq s$ for all $s \in S$.*

Remark 1.24. *(a) The maximum and the minimum need not exist, even if the set is bounded. For example the set $S := \{1/n : n \in \mathbb{N}\}$ doesn't have a minimum.*

(b) Note that if $\sup S \in S$, then it is also the maximum of S. Similarly, if S has a maximum then $\sup S \in S$ and the maximum is equal to the supremum. Similar statements hold for the infimum and the minimum.

Exercise 1.25. *Find the supremum, infimum, minimum, and maximum (if they exist) of the following sets.*

(a) $\mathbb{N}$,

(b) $\{1/2, -1/2, 2/3, -2/3, 3/4, \dots\}$,

(c) $\{x \in \mathbb{R} : 0 < x < 1\}$,

(d) $\{x \in \mathbb{R} : 0 \leq x < 1\}$,

(e) $\{x \in \mathbb{R} : 0 \leq x \leq 1\}$.

We are now able to state the completeness axiom.

Definition 1.21 (The Completeness Axiom (or least upper bound property)**).** *Every nonempty subset of $\mathbb{R}$ that is bounded from above has a supremum in $\mathbb{R}$.*

Quite a lot of properties and theorems that you will meet in this semester will ultimately rest on the completeness axiom. Hence, it is a good idea to carry it with you at all times. Two important consequences of the completeness axiom are stated in the following proposition.

Proposition 1.26. *(a) (The Archimedean Property). For any real number $a > 0$ and any real number b, there exists an $m \in \mathbb{N}$ such that $b < ma$.*

(b) For any real numbers a and b such that $a < b$, there exists a $q \in \mathbb{Q}$ such that $a < q < b$. In other words, between any two real numbers, there is a rational number.

Proof. (a) Assume not. Then there exists $a > 0$ and $b \in \mathbb{R}$ such that $ma \leq b$ for all $m \in \mathbb{N}$. In other words the set $S := \{ma : m \in \mathbb{N}\}$ is bounded from above. By the completeness axiom there then exists a number $s \in \mathbb{R}$ such that $s = \sup\{ma : m \in \mathbb{N}\}$. As $a > 0$, $s - a < s$. As s is the least upper bound of S, this implies that $s - a$ is not an upper bound of S. Hence, there exists an $m \in \mathbb{N}$, let's call it m^*, such that $m^*a > s - a$. It follows that $(m^* - 1)a > s$. Note that as m^* is a natural number $m^* + 1$ is a natural number as well. But then $(m^* - 1)a > s$ implies that s is not an upper bound of S, which contradicts the fact that s was defined as the supremum of S. From this contradiction we deduce that S does not have a supremum. It thus follows from the completeness axiom that S is not bounded.

(b) Let a and b be two real numbers such that $b-a>0$. By the Archimedean property there then exists a natural number m such that $m(b-a)>1$. It follows that $mb>ma+1$. Now let $n:=\min\{k\in\mathbb{Z}:k>ma\}$. Then $ma<n\leq ma+1<mb$. If we divide this last expression by m, we get $a<n/m<b$. By construction n/m is a rational number. We're done.

□

We are now in the position to show that $\mathbb{Q}$ does not satisfy the completeness axiom. Consider for example the set $S:=\{q\in\mathbb{Q}:q^2<2\}$. Obviously, this set is bounded from above. Hence, by the completeness axiom S has a supremum in $\mathbb{R}$, which is easily found to be $\sqrt{2}$ (Check this!). However, as was noted earlier $\sqrt{2}$ is not a rational number. It follows that S does not have a supremum in $\mathbb{Q}$.

One important kind of subsets of $\mathbb{R}$ are intervals.

Definition 1.22. *For any real numbers a and b with $a<b$, we define the **open interval** (a,b) as $(a,b):=\{t\in\mathbb{R}:a<t<b\}$, the **semiopen intervals** $(a,b]$ and $[a,b)$ as $(a,b]:=\{t\in\mathbb{R}:a<t\leq b\}$ and $[a,b):=\{t\in\mathbb{R}:a\leq t<b\}$ respectively, and the **closed interval** $[a,b]$ as $[a,b]:=\{t\in\mathbb{R}:a\leq t\leq b\}$. Note that any of these intervals is bounded and of length $b-a$.*

1.2.2 Properties of addition and multiplication

Remark 1.27. *For any $a,b,c\in\mathbb{R}$, addition and multiplication satisfy the following properties:*

i. (Commutativity) $a+b=b+a$, $a\cdot b=b\cdot a$.

ii. (Associativity) $(a+b)+c=a+(b+c)$, $(a\cdot b)\cdot c=a\cdot(b\cdot c)$.

iii. (Distributivity) $a\cdot(b+c)=a\cdot b+a\cdot c$.

Remark 1.28. *For any $a,b,c,d\in\mathbb{R}$, we have the following:*

i. $\dfrac{a}{b}+\dfrac{c}{d}=\dfrac{a\cdot d+c\cdot b}{b\cdot d}$.

ii. $\dfrac{a}{b}-\dfrac{c}{d}=\dfrac{a\cdot d-c\cdot b}{b\cdot d}$.

iii. $\dfrac{a}{b}\cdot\dfrac{c}{d}=\dfrac{a\cdot c}{b\cdot d}$.

iv. $\dfrac{a}{b}:\dfrac{c}{d}=\dfrac{\frac{a}{b}}{\frac{c}{d}}=\dfrac{a}{b}\cdot\dfrac{d}{c}=\dfrac{a\cdot d}{c\cdot b}$.

Definition 1.23. *Let $x_1,x_2,\ldots,x_n$ be real numbers. We define $\sum_{i=1}^n x_i:=x_1+x_2+\ldots+x_n$ and $\prod_{i=1}^n x_i:=x_1\cdot x_2\cdot\ldots\cdot x_n$.*

Definition 1.24. *For all $x\in\mathbb{R}$ and for any positive integer n, we define $x^0:=1$, $x^n:=x^{n-1}\cdot x$ and $x^{-n}:=\frac{1}{x^n}$. Note that x^{-n} is only defined if $x\neq 0$.*

Proposition 1.29 (Rules of exponentiation)**.** *For any $a, b \in \mathbb{R}$ and any $p, q \in \mathbb{Z}$, we have*

i. $a^p \cdot a^q = a^{p+q}$.

ii. $(a^p)^q = a^{p \cdot q}$.

iii. $a^p \cdot b^p = (a \cdot b)^p$.

iv. $\left(\dfrac{a}{b}\right)^p = \dfrac{a^p}{b^p}$.

v. $\dfrac{a^p}{a^q} = a^p \cdot a^{-q} = a^{p-q}$.

Definition 1.25. *Let $a \in \mathbb{R}_+$ and $n \in \mathbb{N}$. The nonnegative real number x such that*

$$x^n = a$$

is called the n^{th} root of a and is denoted by $x = \sqrt[n]{a} = a^{\frac{1}{n}}$.

Proposition 1.30. *For any $a, b \in \mathbb{R}_+$ and any $n, m \in \mathbb{N}$, we have:*

i. $\sqrt[m]{a} \cdot \sqrt[n]{a} = a^{\frac{1}{m}} \cdot a^{\frac{1}{n}} = a^{\frac{1}{m}+\frac{1}{n}} = a^{\frac{n+m}{n \cdot m}} = \sqrt[n \cdot m]{a^{n+m}}$.

ii. $\sqrt[m]{\sqrt[n]{a}} = (a^{\frac{1}{n}})^{\frac{1}{m}} = a^{\frac{1}{n \cdot m}} = \sqrt[n \cdot m]{a}$.

iii. $\sqrt[n]{a} \cdot \sqrt[n]{b} = a^{\frac{1}{n}} \cdot b^{\frac{1}{n}} = (a \cdot b)^{\frac{1}{n}} = \sqrt[n]{a \cdot b}$.

iv. $\sqrt[m]{a^n} = (a^n)^{\frac{1}{m}} = a^{\frac{n}{m}} = (a^{\frac{1}{m}})^n = (\sqrt[m]{a})^n$.

v. $\sqrt[n]{\frac{a}{b}} = \frac{\sqrt[n]{a}}{\sqrt[n]{b}}$.

1.3 Functions and correspondences.

1.3.1 Functions

Intuitively, we think of a function from a set X to a set Y as a 'rule' that assigns to any element $x \in X$ a unique element $y \in Y$. Although this is probably the way functions were introduced to you in previous classes this is not a proper definition of a function as it is unclear what a rule is. To give a formal definition of a function we first need to introduce the concept of a relation from A to B, which we now do.

Definition 1.26. *For any two nonempty sets A and B, a (binary)* ***relation*** *R from A to B is a subset of the Cartesian product between A and B. Formally, $R \subseteq A \times B$.*

A function in turn is nothing else than a special kind of relation.

Definition 1.27. *A **function** f from X to Y, denoted $f : X \to Y$, is a relation from X to Y such that for any $x \in X$ there exists a unique $y \in Y$ such that $(x, y) \in f$. If $(x, y) \in f$, then we write $f(x) = y$. y is referred to as the **image** (or **value**) **of** x **under** f. The set X is called the **domain** of f, while Y is called the **codomain** of f. The **range** of f is defined as*

$$f(X) := \{y \in Y : f(x) = y \text{ for some } x \in X\}.$$

Two functions f and g from X to Y are said to be equal iff $f(x) = g(x)$ for all $x \in X$.

Example 1.31. *(a) Let $S := \{1, 2, 3\}$ and $T := \{4, 5, 6\}$. Then $f := \{(1, 5), (2, 4), (3, 6)\}$ is a function from S to T. However, $g := \{(1, 5), (2, 4), (1, 6)\}$ is not.*

(b) Similarly, $f := \{(x, y) \in \mathbb{R} : y = x^2\}$ is a function. Obviously, we usually write $f(x) = x^2$ instead.

Exercise 1.32. *Suppose $f : A \to C$ and $g : B \to C$. Prove that if A and B are disjoint then $f \cup g : A \cup B \to C$.*

Exercise 1.33. *Let $E := \{(p, q) \in P \times P :$ the person p is the enemy of the person $q\}$, and $F := \{(p, q) \in P \times P :$ the person p is a friend of the person $q\}$, where P is the set of all people. What does saying 'an enemy of my enemy is one's friend' mean about the relations E and F?*

Exercise 1.34. *(a) Let $A := \{1, 2, 3\}$, $B := \{4\}$, and $f := \{(1, 4), (2, 4), (3, 4)\}$. Is f a function from A to B.*

(b) Let $A := \{1\}$, $B := \{2, 3, 4\}$, and $f := \{(1, 2), (1, 3), (1, 4)\}$. Is f a function from A to B?

(c) Let C be the set of all cars registered in your state, and let S be the set of all finite sequences of letters and digits. Let $L := \{(c, s) \in C \times S :$ the license plate number of the car c is $s\}$. Is L a function from C to S?

Exercise 1.35. *(a) Let N be the set of all countries and C the set of all cities. Let $H : N \to C$ be the function defined such that for any country $n \in N$, $H(n)$ is the capital of that country. What is $H(Italy)$?*

(b) Let $A := \{1, 2, 3\}$ and $B := 2^A$. Let $f : B \to B$ be the function defined by $F(X) = A \setminus X$. What is $f(\{1, 3\})$?

(c) Let $f : \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ be defined by $f(x) = (x + 1, x - 1)$. What is $f(2)$?

Definition 1.28. *If $f(X) = Y$, that is if the range and the codomain are equal, then f is said to map X **onto** Y. f is then called a **surjection** (or **surjective map**). If for any $x, y \in X$ such that $x \neq y$, we have $f(x) \neq f(y)$, then we say that f is an **injection** (or a **one-to-one** or **injective** function/map). If a function f is both injective and surjective then it is called a **bijection** (or **bijective** function/map).*

Here are other characterizations of injection and surjection that are sometimes useful.

Proposition 1.36. *Suppose $f : X \to Y$. Then,*

(a) f is injective iff for all $x_1, x_2 \in X$, $f(x_1) = f(x_2)$ implies $x_1 = x_2$.

(b) f is surjective iff for all $y \in Y$ there exists an $x \in X$ such that $f(x) = y$.

Example 1.37. *Let $f : \mathbb{R} \setminus \{-1\} \to \mathbb{R}$ be defined by $f(x) = \frac{2x}{x+1}$. To show that f is one-to-one let us choose x_1 and x_2 such that $f(x_1) = f(x_2)$. Note that $\frac{2x_1}{x_1+1} = \frac{2x_2}{x_2+1}$ implies $2x_2(x_1+1) = 2x_1(x_2+1)$ which implies that $x_1 = x_2$ as we sought. To see that f is not onto, let us try to see if there exists an x such that $f(x) = 2$. If there were such an x, we would have $\frac{2x}{x+1} = 2$ which implies that $2x = 2x + 2$ which is obviously impossible.*

Remark 1.38. *A fact worth remembering is that any injection can be made a bijection by restricting the codomain of the function to the range of the function. Formally, if $f : X \to Y$ is an injection, then $f : X \to Z$ is a bijection where $Z := f(X)$.*

Exercise 1.39. *Let $A := \mathbb{R} \setminus \{1\}$, and let $f : A \to A$ be defined by*

$$f(x) := \frac{x+1}{x-1}.$$

Show that f is injective and surjective.

Exercise 1.40. *Let $A := 2^{\mathbb{R}}$. Define $f : \mathbb{R} \to A$ by $f(x) := \{y \in \mathbb{R} : y^2 < x\}$. What is $f(2)$? Is f one-to-one? Is f onto?*

Example 1.41. *(a) A **constant function** assigns the same value $y \in Y$ to each x in its domain X, i.e. $f(x) = y$ for all $x \in X$.*

*(b) If the domain and the codomain of a function are identical, then the function is called a **self-map** on the domain X.*

*(c) The **identity function** is a particular type of self-map on X that assigns the value x to each $x \in X$. We denote the identity function id_X. Obviously, it is defined by $id_X(x) := x$ for all $x \in X$. Note that the identity function is a bijection.*

*(d) Another important function is the indicator function. Here is its definition. Let $S \subseteq X$. The **indicator function of** S, denoted $\mathbf{1}_S$, then assigns 0 to each $x \in X$ such that $x \notin S$ and 1 to each $x \in S$. Formally, $\mathbf{1}_S(x) := \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{if } x \in X \setminus S \end{cases}$.*

*(e) A function $f : X \times Y \to X$ defined by $f(x, y) := x$ is called the **projection of** $X \times Y$ **onto** X. Note that a projection is necessarily surjective.*

We now discuss ways to get new functions from existing ones.

Definition 1.29. *(a) Let $Z \subseteq X \subseteq W$ and $f : X \to Y$. The **restriction** of f to Z, denoted $f|_Z$, is the function $f|_Z : Z \to Y$ defined by $f|_Z(z) := f(z)$ for all $z \in Z$. An **extension** of f to W is a function $f^* : W \to Y$ with $f^*|_X = f$, i.e. $f^*(x) = f(x)$ for all $x \in X$.*

*(b) Let $f : X \to Y$ and $g : Y \to Z$. The **composition** of f and g, denoted $g \circ f$, is a function from X to Z defined by $g \circ f(x) = g(f(x))$.*

Example 1.42. *Let $g : \mathbb{Z} \to \mathbb{R}$ defined by $g(x) = 2x + 3$ and let $f : \mathbb{Z} \to \mathbb{Z}$ be defined by $f(n) = n^2 - n - 1$. Then $g \circ f : \mathbb{Z} \to \mathbb{R}$ with $(g \circ f)(n) = 2(n^2 - n - 1) + 3 = 2n^2 - 2n - 1$.*

Exercise 1.43. *Let f and g be functions from $\mathbb{R}$ to $\mathbb{R}$ defined by $f(x) := \frac{1}{x^2+2}$ and $g(x) := 2x - 1$ respectively. What are the formulas of $(f \circ g)(x)$ and $(g \circ f)(x)$?*

Exercise 1.44. *For each of the following pairs of functions g and h, compute the formulas of $g \circ h$ and $h \circ g$. What is the domain of each of these composite functions.*

(a) $g(x) = x^2 + 4$, $h(z) = 5z - 1$;

(b) $g(x) = x^3$, $h(z) = (z - 1)(z + 1)$;

(c) $g(x) = (x - 1)/(x + 1)$, $h(z) = (z + 1)/(1 - z)$;

(d) $g(x) = 4x + 2$, $h(z) = \frac{1}{4}(z - 2)$;

(e) $g(x) = 1/x$, $h(z) = z^2 + 1$.

We now consider the question of when we can *invert* a function. Let us consider the function $f : X \to Y$. As you remember a function is nothing else than a special kind of relation, i.e. $f \subseteq X \times Y$. Now by the **inverse** of a relation $R \subseteq X \times Y$, denoted R^{-1}, we mean the following set:

$$R^{-1} := \{(y, x) \in Y \times X : (x, y) \in R\}.$$

Note that R^{-1} is a relation from Y to X that contains exactly the same ordered pairs as the original relation R only with the order being reversed. Obviously in the case of the function $f : X \to Y$, f^{-1} is also a relation from Y to X. The question we now ask is the following: under what conditions is f^{-1} a function as well, i.e. under what conditions is it the case that for all $y \in Y$ there exists a unique $x \in X$ such that $(y, x) \in f^{-1}$? Note that if f^{-1} is a function then we say that f is **invertible.**

Let us first show that even if f is a function from X to Y it need not be the case that f^{-1} is a function from Y to X. Let $X := \{1, 2, 3\}$ and $Y := \{8, 9, 10\}$. Now let $f(1) = 8$, $f(2) = 10$ and $f(3) = 8$. It follows that $f^{-1} = \{(8, 1), (10, 2), (8, 3)\}$. For f^{-1} to be a function it has to be the case that for all $y \in Y$ we have a unique $x \in X$ such that $(y, x) \in f^{-1}$. Note however that this is not the case for 9 as there is no x such that $(9, x) \in f^{-1}$. Moreover there are two x's (1 and 3) such that $(8, x) \in f^{-1}$ instead of a unique one as is required of a function. The problem that arises with 9 is that f is not onto. The problem that arises with 8 is that f is not one-to-one. In fact, this is all we need to characterize the existence of an inverse function:

Proposition 1.45. *Let X and Y be two nonempty sets and $f : X \to Y$. Then f^{-1} is a function from Y to X iff f is one-to-one and onto.*

And here is another characterization of invertible functions.

Proposition 1.46. *Let X and Y be two nonempty sets. Then $f : X \to Y$ is invertible iff there exists a function $g : Y \to X$ such that $g \circ f = id_X$ and $f \circ g = id_Y$.*

Example 1.47. *Let $f : \mathbb{R} \to \mathbb{R}_+$ be defined by $f(x) = x^2$. Obviously, f is not invertible as $(1, 1)$ and $(1, -1)$ are both in f^{-1}. However, $f|_{\mathbb{R}_+}$ is invertible with $f|_{\mathbb{R}_+}^{-1}(y) = \sqrt{x}$ for all $y \in \mathbb{R}_+$.*

Exercise 1.48. *Compute an expression of the inverse of each of the following functions. What is the domain of each of the inverse functions?*

(a) $f(x) = 3x + 6$,

(b) $f(x) = 1/(x + 1)$,

(c) $f(x) = x^{2/3}$,

(d) $f(x) = x^2 + x + 2$.

1.3.2 Sequences

By a **sequence** in a nonempty set X we mean a function $f : \mathbb{N} \to X$. It is customary to think of a sequence as an ordered array of the form $(x_1, x_2, \ldots)$ with $x_i := f(i)$ for each $i \in \mathbb{N}$. A specific sequence is usually denoted by (x_m) (other used notations are $(x_m)_{m=1}^{\infty}$ or $\{x_m\}_{m=1}^{\infty}$). The set of all sequences in X is denoted X^∞ (some texts use the notation $X^{\mathbb{N}}$ instead).

In turn, a **real sequence** is nothing else but a function $f : \mathbb{N} \to \mathbb{R}$. Following our previous remark the set of all real sequences is denoted $\mathbb{R}^\infty$. We will learn more about real sequences during the regular math class. For the purposes of the prefresher the only additional notion that we need to introduce is the notion of convergence. Intuitively we say that a real sequence (x_m) is **convergent** if there exists a real number x such that later terms of the sequence get arbitrarily close to x. Formally, (x_m) converges to x if, for any $\epsilon > 0$, there exists an $M \in \mathbb{N}$ such that for all $m \geq M$ we have $|x_m - x| < \epsilon$. If (x_m) converges to x we write $\lim_{m\to\infty} x_m = x$, or $\lim x_m = x$, or simply, $x_m \to x$. If $x_m \to x$, then x is called the **limit** of (x_m).

Remark 1.49. *If $x_m \to x$ then all but finitely many terms of the sequence (x_m) are contained in $(x - \epsilon, x + \epsilon)$, no matter how small $\epsilon > 0$.*

Example 1.50. *The sequence $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots)$ converges to 0.*

1.3.3 Correspondences

A **correspondence** is a function (let's call it Γ) from a nonempty set X into the set of nonempty subsets of a set Y, i.e. $\Gamma : X \to 2^Y \setminus \emptyset$. Hence, for any $x \in X$, $\Gamma(x)$ is a nonempty subset of Y. For this reason correspondences are often called set-valued functions. Instead of the notation $\Gamma : X \to 2^Y \setminus \emptyset$, we usually denote a correspondence from X to Y by $\Gamma : X \rightrightarrows Y$ (some textbooks use $\Gamma : X \to\to Y$). The **domain** of the correspondence is X while the **codomain** is Y. For any $S \subseteq X$, we let

$$\Gamma(S) := \bigcup\{\Gamma(x) : x \in S\}.$$

The range of the correspondence is then $\Gamma(X)$. A correspondence is **surjective** if $\Gamma(X) = Y$ and it is called a **self-correspondence** if $\Gamma(X) \subseteq X$.

Remark that every function $f : X \to Y$ can be viewed as correspondence $\Gamma : X \rightrightarrows Y$ defined by $\Gamma(x) := \{f(x)\}$. Similarly, if $|\Gamma(x)| = 1$ for all $x \in X$, then Γ can be viewed as a function from X into Y. In this latest case we say that Γ is a single-valued correspondence.

Example 1.51. *Let* $\Gamma : \mathbb{R} \rightrightarrows \mathbb{R}$ *be defined by* $\Gamma(x) = \begin{cases} 0 & \textit{if } x \leq 1 \\ [1, 2] & \textit{if } x > 1 \end{cases}$.

1.4 Real Functions

1.4.1 Basic Definitions

Let T be a nonempty set. If f is a function from T into the reals, i.e. $f : T \to \mathbb{R}$, then f is called a **real function** (or **real-valued function**) on T. We define the addition, subtraction, multiplication and division of two real functions f and g on T as follows:

$$(f + g)(t) := f(t) + g(t),$$

$$(f - g)(t) := f(t) - g(t),$$

$$(fg)(t) := f(t)g(t),$$

and

$$\left(\frac{f}{g}\right)(t) := \frac{f(t)}{g(t)}$$

for all $t \in T$. Of course $\left(\frac{f}{g}\right)(t)$ is only well-defined if $g(t) \neq 0$ for all $t \in T$. Similarly for any real number a, $(af)(t) := af(t)$.

Now let T be a nonempty subset of $\mathbb{R}$ and f be a real function on T. We say that f is **increasing** if for all $x, y \in T$, if $x \geq y$ then $f(x) \geq f(y)$ and **strictly increasing** if for all $x, y \in T$, if $x > y$, then $f(x) > f(y)$. Similarly, f is **decreasing** if for all $x, y \in T$, if $x \geq y$ then $f(x) \leq f(y)$ and **strictly decreasing** if for all $x, y \in T$, if $x > y$, then $f(x) < f(y)$. Finally, a function is **monotonic** if it is increasing or decreasing.

Finally, we say that a function $f : T \to \mathbb{R}$ is **bounded** if there exists a real number K such that $|f(t)| \leq K$ for all $t \in T$. Here are a few facts worth knowing.

Remark 1.52. *1. Any monotonic real function f on an interval $[a, b]$ is bounded.*

2. Any strictly increasing (decreasing) real function f on an interval $[a, b]$ is invertible.

1.4.2 Limits, Continuity and Differentiation

We now introduce the notion of continuity of a real-valued function. Although, intuitively relatively easy to grasp the formalization of this notion is somewhat more complicated. Intuitively, we say that a function is continuous if points that are close to each other are mapped to points that are close to each other. Here is the formalization.

Definition 1.30. *Let T be a nonempty subset of $\mathbb{R}$ and $f : T \to \mathbb{R}$. We say that f is* ***continuous*** *at x if for any $\epsilon > 0$ there exists a $\delta > 0$ such that $|f(x) - f(t)| < \epsilon$ for all $t \in T \setminus \{x\}$ with $|t - x| < \delta$. If f is continuous at any $x \in S \subseteq T$, then we say that it is continuous on S. If it is continuous at every $x \in T$, we simply say that f is continuous. Note that if f is a continuous real function on T, we write $f \in \mathbf{C}(T)$.*

Here is an equivalent definition of continuity that relies on sequences.

Proposition 1.53. *Let T be a nonempty subset of $\mathbb{R}$ and $f : T \to \mathbb{R}$. f is continuous at x iff for any sequence $(x_m) \in T \setminus \{x\}$ such that $x_m \to x$, we have $f(x_m) \to f(x)$, in which case we write $\lim_{t\to x} f(t) = f(x)$ or $\lim_{k\to\infty} f(x_k) = f(x)$.*

Remark 1.54. *The notion of continuity is a so-called* ***local property*** *in the sense that if the function f is continuous at a particular $x_0 \in T$, it is not necessarily the case that f is continuous at some $y \neq x_0$.*

Exercise 1.55. *For each of the following functions, sketch its graph and determine whether the function is continuous at the point of transition of its two formulas.*

(a) $y = \begin{cases} +x^2 \text{ if } x \geq 0, \\ -x^2 \text{ if } x < 0; \end{cases}$

(b) $y = \begin{cases} +x^2 + 1 \text{ if } x \geq 0, \\ -x^2 - 1 \text{ if } x < 0; \end{cases}$

(c) $y = \begin{cases} x^3 \text{ if } x \leq 1, \\ x \text{ if } x > 1; \end{cases}$

(d) $y = \begin{cases} x^3 \text{ if } x < 1, \\ 3x - 2 \text{ if } x \geq 1. \end{cases}$

In proposition 1.53 we gave an idea of the notion of a limit somewhat in passing by. Let me give a more precise definition.

Definition 1.31. *Let T be a nonempty subset of $\mathbb{R}$ and $f : T \to \mathbb{R}$. If there exists at least one sequence in $T \setminus \{x\}$ that converges to $x \in T$, we say that y is the* ***limit*** *of f at x if for any sequence (x_m) in $T \setminus \{x\}$ such that $x_m \to x$, we have $f(x_m) \to y$. If this is the case we write $\lim_{t\to x} f(t) = y$.*

Remark 1.56. *Note that the limit of f at x need not exist. Consider for example the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) := \begin{cases} 0 \text{ if } x \leq 0 \\ 1 \text{ if } x > 0 \end{cases}$. Then, f does not have a limit at 0. To see this remark that for every sequence (x_m) such that $x_m \geq 0$ and $x_m \to 0$, we have $f(x_m) \to 1$, whereas for any sequence (x_m) such that $x_m < 0$ and $x_m \to 0$, we have $f(x_m) \to 0$.*

Proposition 1.57. *Let T be a nonempty subset of $\mathbb{R}$ and f, g be two real real functions on T. Assume that the limits of f and g exist at $x \in T$ and are finite, then the following holds*

1. *$lim_{t\to x} (f(t) + g(t)) = lim_{t\to x} f(t) + lim_{t\to x} g(t)$,*
2. *$lim_{t\to x} (f(t) - g(t)) = lim_{t\to x} f(t) - lim_{t\to x} g(t)$,*
3. *$lim_{t\to x} f(t)g(t) = lim_{t\to x} f(t) lim_{t\to x} g(t)$ and*
4. *$lim_{t\to x} \frac{f(t)}{g(t)} = \frac{lim_{t\to x} f(t)}{lim_{t\to x} g(t)}$. This last statement only holds if $g(t) \neq 0$ for all $t \in T$ and $lim_{t\to x} g(t) \neq 0$ as well.*

Here are a few consequences of continuity that are worth remembering.

Proposition 1.58. *(a) (Weierstrass' Theorem) Take any $a, b \in \mathbb{R}$ with $a \leq b$ and let f be a continuous real function on $[a, b]$. Then, there exist $x, y \in [a, b]$ such that $f(x) \geq f(t) \geq f(y)$ for all $t \in [a, b]$.*

(b) (Intermediate Value Theorem) Take any $a, b \in \mathbb{R}$ with $a \leq b$ and let f be a continuous real function on $[a, b]$. If $f(a) < f(b)$, and if $c \in \mathbb{R}$ satisfies $f(a) < c < f(b)$, then there exists $x \in (a, b)$ such that $f(x) = c$.

Definition 1.32. *Let $I := [a, b]$ with $a < b$ and let $f : I \to \mathbb{R}$. Let x be an arbitrary element of I, then the* ***difference quotient map of f at*** *x, $Q_{f,x} : I \setminus \{x\} \to \mathbb{R}$ is defined by*

$$Q_{f,x}(t) := \frac{f(t) - f(x)}{t - x}.$$

In turn we say that f is ***differentiable at*** *x if*

$$\lim_{t\to x} \frac{f(t) - f(x)}{t - x} \in \mathbb{R}.$$

If f is differentiable at x_0, we let $f'(x_0) = \lim_{t\to x_0} \frac{f(t)-f(x_0)}{t-x_0} \in \mathbb{R}$ and call $f'(x_0)$ the ***derivative of f at*** *x_0. Note that another notation for the derivative of f at x_0 is*

$\frac{df}{dx}(x_0)$. Now let $J \subseteq I$. If f is differentiable at any $x \in J$, then we say that f is differentiable on J. If $J = I$ in this case, we say that f is differentiable.

*If f is differentiable on I, the derivative of f is a function $f' : I \to \mathbb{R}$ that assigns to any $x \in I$ the value of the derivative of f at x. If f' is a continuous function on I then we say that f is **continuously differentiable** and write $f \in \mathbf{C}^1(I)$. If f' is itself differentiable on I, then we say that f is **twice differentiable**. The second derivative of f is then the function $f'' : I \to \mathbb{R}$ that assigns to any $x \in I$ the derivative of f' at x. If f'' is a continuous function on I, then we say that f is **twice continuously differentiable** and write $f \in \mathbf{C}^2(I)$.*

Before proceeding any further, here are a few words about the interpretation of the derivative. The derivative tells us how much $f(x)$ changes if we increase x only by a very small (infinitesimal) amount.

Here is an important fact for you to remember.

Proposition 1.59. *If $f : I \to \mathbb{R}$ is differentiable then it is continuous.*

The converse is not true however. Here is an example.

Example 1.60. *The absolute value function $|\cdot| : \mathbb{R} \to \mathbb{R}$ defined by $|x| := \begin{cases} x \text{ if } x \geq 0 \\ -x \text{ if } x < 0 \end{cases}$ is continuous, however it is not differentiable at 0.*

1.4.3 Rules of Differentiation

We now study how to compute the derivative of some important functions.

Proposition 1.61. *Let f and g be two differentiable real functions on $\mathbb{R}$.*

(a) If $f(x) := cx^n$ then $f'(x) = cnx^{n-1}$.

(b) Let $f : \mathbb{R} \to \mathbb{R}$ and $k \in \mathbb{R}$. Then $(kf)'(x) = kf'(x)$.

(c) Then $(f \pm g)'(x) = f'(x) \pm g'(x)$.

(d) Then $(f \cdot g)'(x) = f'(x)g(x) + f(x)g'(x)$.

(e) Then $\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$.

(f) Then $((f(x))^n)' = nf(x)^{n-1} \cdot f'(x)$.

(g) (Chain Rule) Let $f : I \to \mathbb{R}$ be a differentiable function and $f(I)$ an open interval. If $g : f(I) \to \mathbb{R}$ is differentiable as well then so is $g \circ f$. Moreover, $(g \circ f)' = (g' \circ f)f'$.

(h) (L'Hospital's Rule) Suppose that $\lim_{t\to x} f(x) = \lim_{t\to x} g(x) = 0$. Then $\lim_{t\to x} \frac{f(x)}{g(x)} = \lim_{t\to x} \frac{f'(x)}{g'(x)}$, provided that $\lim_{t\to x} \frac{f'(x)}{g'(x)}$ exists.

Example 1.62. *Let f be a constant function, i.e. $f(x) := c$ for all $x \in \mathbb{R}$. Then for all $x \in \mathbb{R}$,*

$$Q_{f,x} := \frac{f(t) - f(x)}{t - x} = \frac{c - c}{t - x} = 0.$$

Hence,

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} = 0$$

as well and $f'(x) = 0$ for all x.

Example 1.63. *Let $f(x) := x^2$. Then for all $x \in \mathbb{R}$,*

$$Q_{f,x} := \frac{f(t) - f(x)}{t - x} = \frac{t^2 - x^2}{t - x} = \frac{(t - x)(t + x)}{t - x} = t + x.$$

It follows that

$$\lim_{t \to x} \frac{f(t) - f(x)}{t - x} = \lim_{t \to x}(t + x) = \lim_{t \to x} t + \lim_{t \to x} x = 2x,$$

and $f'(x) = 2x$ for all x.

Exercise 1.64. *(a) Prove that the derivative of an arbitrary linear function $f(x) = mx + a$ is $f'(x) = m$.*

(b) Prove that the derivative of $f(x) = x^3$ is $f'(x) = 3x^2$.

Exercise 1.65. *Compute the derivative of the following functions.*

(a) $-7x^3$,

(b) $3x^{-3/2}$,

(c) $3x^2 - 9x + 7x^{2/5} - 3x^{1/2}$,

(d) $(x^2 + 1)(x^2 + 3x + 2)$,

(e) $\frac{x-1}{x+1}$,

(f) $(x^5 - 3x^2)^7$,

(g) $(x^3 + 2x)^3(4x + 5)^2$,

(h) $12x^{-2}$,

(i) $\frac{1}{2}\sqrt{x}$,

(j) $4x^5 - 3x^{1/2}$,

(k) $(x^{1/2} + x^{-1/2})(4x^5 - 3\sqrt{x})$,

(l) $\frac{x}{x^2+1}$,

(m) $5(x^5 - 6x^2 + 3x)^{2/3}$.

Here is an example on how to use the chain rule to compute the derivative of a function.

Example 1.66. *Let $g(x) = x^2$ and $h(z) = z^2 + 4$. Then $(h \circ g)(x) = x^4 + 4$ and $(g \circ h)(z) = (z^2 + 4)^2$. Let us compute the derivative of $h \circ g$. Obviously we can compute this derivate directly. We then get $(h \circ g)'(x) = 4x^3$. Another way would be to use the chain rule. We first compute $h'(z) = 2z$ and $g'(x) = 2x$. Next we compute $(h' \circ g)(x) = 2x^2$ and $(h \circ g)'(x) = (h' \circ g)(x) \cdot g'(x) = 2x^2 \cdot 2x = 4x^3$.*

Exercise 1.67. *Using the chain rule compute all the derivatives of the composite functions you constructed in exercise 1.44.*

So far we have only considered real functions from $\mathbb{R}$ or a subset of $\mathbb{R}$ to $\mathbb{R}$. Now we will start to consider functions from $\mathbb{R}^n$ to $\mathbb{R}$, where $\mathbb{R}^n := \mathbb{R} \times \ldots \times \mathbb{R}$ n times. Here are a few examples of such functions

Example 1.68. *(a) Let $f : \mathbb{R}^2 \to \mathbb{R}$ (equivalently we could write $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$) be defined by $f(x_1, x_2) := 3x_1x_2$.*

(b) Let $f : \mathbb{R}^3 \to \mathbb{R}$ (equivalently we could write $f : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$) be defined by $f(x_1, x_2, x_3) := 3x_1 + x_2^2x_3 + x_1x_3^4$.

We now study how to differentiate such a function with respect to one of its variables while holding the other variables constant. But first we need some additional notation. We denote an arbitrary element of $\mathbb{R}^n$ by $\mathbf{x} := (x_1, \ldots, x_n)$. Moreover, we let e_j be an element of $\mathbb{R}^n$ that takes the value 0 everywhere except at the $j-$th place where it takes the value 1. For example $e_3 \in \mathbb{R}^4$ is equal to $(0, 0, 1, 0)$. Similarly $e_2 \in \mathbb{R}^5$ is equal to $(0, 1, 0, 0, 0)$. We are now able to define the notion of a partial derivative.

Definition 1.33. *Let $f : S \to \mathbb{R}$ with $S \subseteq \mathbb{R}^n$. Then the partial derivative of f with respect to the variable x_j exists at $\mathbf{x}$ if there exists a real number, denoted $\partial f(\mathbf{x})/\partial x_j$ such that*

$$\lim_{t \to 0} \left(\frac{f(\mathbf{x} + te_j) - f(\mathbf{x})}{t} \right) = \frac{\partial f}{\partial x_j}(\mathbf{x}).$$

Note that the derivative of f with respect to $\mathbf{x}$ is a vector that contains all the partial derivatives, i.e. $Df(\mathbf{x}) := [\partial f(\mathbf{x})/\partial x_1, \ldots, \partial f(\mathbf{x})/\partial x_n]$.

Although it may be confusing at times, computing the partial derivatives of a function is a process that is nearly identical to the computation of the derivative of a function from $\mathbb{R}$ to $\mathbb{R}$. All there is to do is to treat all the other variables as constants and apply the rules of differentiation presented earlier. Here are a few examples:

Example 1.69. *(a) Consider $f(x_1, x_2) := 3x_1x_2$. Then*

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 3x_2,$$

and

$$\frac{\partial f(\mathbf{x})}{\partial x_2} = 3x_1.$$

(b) Consider $f(x_1, x_2, x_3) := 3x_1 + x_2^2 x_3 + x_1 x_3^4$. *Then*

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = 3 + x_3^4,$$

$$\frac{\partial f(\mathbf{x})}{\partial x_2} = 2x_2 x_3,$$

and

$$\frac{\partial f(\mathbf{x})}{\partial x_3} = x_2^2 + 4x_1 x_3^3.$$

Exercise 1.70. *Compute all the partial derivatives (i.e. the partial derivative with respect to* x *and the one with respect to* y*) of the following functions:*

(a) $4x^2y - 3xy^3 + 6x$,

(b) xy,

(c) xy^2,

(d) $\frac{x+y}{x-y}$,

(e) $3x^2y - 7x\sqrt{y}$.

1.4.4 Integration and methods of integration

Let a and b be two real numbers such that $a \leq b$. For any $m \in \mathbb{N}$, we let a dissection $[a_0, \ldots, a_m]$ of $[a, b]$ be the set $\{[a_0, a_1], [a_1, a_2], \ldots, [a_{m-1}, a_m]\}$ with $a = a_0 < a_1 < \cdots < a_m = b$. The set of all dissections of $[a, b]$ is denoted $\mathcal{D}[a, b]$. When $a = b$, we let $\mathcal{D}[a, b] := \{\{a\}\}$. Now let $\mathbf{a} := [a_0, \ldots, a_m]$ and $\mathbf{b} := [b_0, \ldots, b_m]$ be two dissections in $\mathcal{D}[a, b]$. By $\mathbf{a} \Cup \mathbf{b}$ we then mean the dissection $[c_0, \ldots, c_m] \in \mathcal{D}[a, b]$ where $\{c_0, \ldots, c_l\} = \{a_0, \ldots, a_m\} \cup \{b_0, \ldots, b_k\}$. Finally, we say that $\mathbf{b}$ is finer than $\mathbf{a}$ if $\{a_0, \ldots, a_m\} \subseteq \{b_0, \ldots, b_k\}$.

Now consider any bounded function $f : [a, b] \to \mathbb{R}$ and let $\mathbf{a} := [a_0, \ldots, a_m]$ be any dissection in $\mathcal{D}[a,]$. We define

$$K_{f,\mathbf{a}}(i) := \sup\{f(t) : a_{i-1} \leq t \leq a_i\}$$

and

$$k_{f,\mathbf{a}}(i) := \inf\{f(t) : a_{i-1} \leq t \leq a_i\}$$

for all $i \in \{1, \ldots, m\}$. We next define the **a-upper Riemann sum** of f by

$$R_{\mathbf{a}}(f) := \sum_{i=1}^{m} K_{f,\mathbf{a}}(i)\,(a_i - a_{i-1})$$

and the **a-lower Riemann sum** of f by

$$r_{\mathbf{a}}(f) := \sum_{i=1}^{m} k_{f,\mathbf{a}}(i)\,(a_i - a_{i-1})\,.$$

As $\mathbf{a}$ gets finer $R_{\mathbf{a}}(f)$ decreases while $r_{\mathbf{a}}(f)$ increases. Moreover, $R_{\mathbf{a}}(f) \geq r_{\mathbf{a}}(f)$ for any $\mathbf{a} \in \mathcal{D}\,[a,b]$. Finally,

$$R(f) := \inf\{R_{\mathbf{a}}(f) : \mathbf{a} \in \mathcal{D}\,[a,b]\} \geq \sup\{r_{\mathbf{a}}(f) : \mathbf{a} \in \mathcal{D}\,[a,b]\} =: r(f).$$

To see that this last statement is correct assume by contradiction that there exists two dissections $\mathbf{a}$ and $\mathbf{b}$ such that $R_{\mathbf{a}}(f) < r_{\mathbf{b}}(f)$. But then $R_{\mathbf{a}\Cup\mathbf{b}}(f) < r_{\mathbf{a}\Cup\mathbf{b}}(f)$ which is impossible by definition.

We are now able to define the integral of a function f.

Definition 1.34. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function. f is said to be* ***Riemann integrable*** *if $R(f) = r(f)$ in which case the real number*

$$\int_a^b f(t)dt := R(f)$$

is called the ***Riemann integral*** *of f.*

Now let $g : [a,\infty) \to \mathbb{R}$ be a bounded function, then the ***improper Riemann integral*** *of g is defined by*

$$\int_a^\infty g(t)dt := \lim_{b\to\infty} \int_a^b g(t)dt,$$

assuming that the Riemann integral of g exists on any interval $[a,b]$ and the limit of $\int_a^b g(t)dt$ exists as b goes to infinity. We define $\int_{-\infty}^a g(t)dt$ similarly for any bounded real function g on $(-\infty,a]$.

Finally, if there exists a real number a such that $\int_a^\infty g(t)dt$ and $\int_{-\infty}^a g(t)dt$ exist and are not both equal to infinity, then we let

$$\int_{-\infty}^\infty f(t)dt := \int_{-\infty}^a g(t)dt + \int_a^\infty g(t)dt.$$

Note that $\int_{-\infty}^\infty f(t)dt$ is called the ***indefinite integral*** *of f and is often simply denoted $\int f(t)dt$.*

Let us now state a few important properties of integrals.

Proposition 1.71. *Let f be a bounded and Riemann integrable function on $[a,b]$. Then,*

$$|\int_a^b f(t)dt| \leq (b-a)\sup\{|f(t)| : a \leq t \leq b\}.$$

Proposition 1.72. *Let $\alpha \in \mathbb{R}$ and let f and g be two integrable real functions on $[a, b]$, then $\alpha f + g$ is Riemann integrable and*

$$\int_a^b (\alpha f + g)(t)dt = \alpha \int_a^b f(t)dt + \int_a^b g(t)dt.$$

Proposition 1.73. *Let f be a bounded and Riemann integrable real function on $[a, b]$ and $c \in [a, b]$. Then $f|_{[a,c]}$ and $f|_{[c,b]}$ are integrable as well and*

$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt.$$

Proposition 1.74. *Let f be a bounded and Riemann integrable real function on $[a, b]$. Then $|f|$ is Riemann integrable as well and*

$$|\int_a^b f(t)dt| \leq \int_a^b |f(t)|dt.$$

One important question is: what functions are integrable? One very simple and useful result is the following.

Proposition 1.75. *Let f be a continuous real function on $[a, b]$. Then f is Riemann integrable.*

For those of you who have had calculus previously, it is likely that the integral was simply introduced as the inverse operation of differentiation, but the introduction of the notion of integrability that I have given you hasn't made any reference to differentiability so far. The fact that integration and differentiation can be thought of as inverse operations is due to the following theorem which, given its importance, is called the **fundamental theorem of calculus**.

Theorem 1.76. *Let f be a continuous real function on $[a, b]$ and F be any real function on $[a, b]$. Then*

$$F(x) = F(a) + \int_a^x f(t)dt$$

if, and only if, $F \in \mathbf{C}^1 [a, b]$ and $F' = f$.

For your reference here are the formulas of the integrals for some important functions.

Proposition 1.77. *Let C be an arbitrary real number.*

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C,$$

$$\int (f(x))^n f'(x)dx = \frac{(f(x))^{n+1}}{n+1} + C.$$

Let us now see how to use the fundamental theorem of calculus to compute the integral of a few functions.

Example 1.78. *Let us compute* $\int_2^5 3x^2 + 5x^3 dx$*. By proposition 1.72 we get*

$$\int_2^5 3x^2 + 5x^3 dx = 3\int_2^5 x^2 + 5\int_2^5 x^3.$$

We now apply proposition 1.77 to compute $\int x^2 dx$ *and* $\int x^3$*. We get* $\int x^2 dx = \frac{x^3}{3}$ *and* $\int x^3 = \frac{x^4}{4}$*. It follows then from the fundamental theorem of calculus that*

$$\int_2^5 3x^2 + 5x^3 dx = 3\left[\frac{5^3}{3} - \frac{2^3}{3}\right] + 5\left[\frac{5^4}{4} - \frac{2^4}{4}\right] = 117 + 761.25 = 878.25.$$

Finally, here are some more rules that are helpful when computing integrals.

Proposition 1.79. *(Integration by Parts Formula) Let* f *and* g *be continuously differentiable real functions on* $[a, b]$*, then*

$$\int_a^b f(t)g'(t)dt = f(b)g(b) - f(a)g(a) - \int_a^b f'(t)g(t)dt.$$

Proposition 1.80. *(Integration by Substitution) Suppose we make the substitution* $x = g(u)$*, then*

$$\int f(x)dx = \int f(g(u))g'(u)du.$$

Here is an example of how integration by substitution may be used.

Example 1.81. *Consider the indefinite integral*

$$\int x\sqrt{x+1}dx.$$

Now let us make the substitution $x = u^2 - 11$*, then*

$$\begin{aligned}\int x\sqrt{x+1}dx &= \int 2(u^2-1)u^2 du = 2\int u^4 du - 2\int u^2 du \\ &= \frac{2}{5}u^5 - \frac{2}{3}u^3 + C = \frac{2}{5}(x+1)^{5/2} - \frac{2}{3}(x+1)^{3/2} + C.\end{aligned}$$

Exercise 1.82. *Compute the indefinite integral of each of the following functions*

(a) $4x^5 - x^4$,

(b) $12x^3 - 6x^{1/2}$,

(c) $(x^2 + 2x + 4)^{1/2}(x + 1)$,

(d) $(x^3 + 3x^2 + 1)^3(x^2 + 2x)$.

Exercise 1.83. *Compute the following definite integrals*

(a) $\int_{-2}^{4} 4x^5 - x^4 dx$,

(b) $\int_{3}^{8} 12x^3 - 6x^{1/2} dx$,

(c) $\int_{-1}^{5} (x^2 + 2x + 4)^{1/2}(x + 1)dx$,

(d) $\int_{1}^{3} (x^3 + 3x^2 + 1)^3(x^2 + 2x)dx$.

As in the case of differentiation we can extend the notion of the integral to functions with several variables. So let $f : \mathbb{R}^n \to \mathbb{R}$ be an integrable function. If we want to compute the area under this function we use the **multivariate definite integral**

$$\int_{a_n}^{b_n} \int_{a_{n-1}}^{b_{n-1}} \cdots \int_{a_1}^{b_1} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n.$$

The procedure here is to first integrate with respect to one of the variables treating all the others as constants (similar to what you do in partial differentiation). We then get a new multivariate integral which we integrate with respect to another variable and so on. An example might best explain the procedure.

Example 1.84. *Let us compute* $\int_{1/2}^{1} \int_{1}^{2} x^2 y dx dy$. *Let us integrate with respect to* y *which gives us the function* $\frac{1}{2}x^2y^2$. *We now evaluate the function at* $y = 1/2$ *and* $y = 1$ *respectively. It follows that*

$$\int_{1/2}^{1} \int_{1}^{2} x^2 y dx dy = \int_{1}^{2} \frac{1}{2}x^2 dx - \int_{1}^{2} \frac{1}{8}x^2 = \frac{1}{2}\left[\frac{8}{3} - \frac{1}{3}\right] - \frac{1}{8}\left[\frac{8}{3} - \frac{1}{3}\right] = \frac{7}{8}.$$

Exercise 1.85. *Compute the following multivariate definite integrals.*

(a) $\int_{1}^{3} \int_{0}^{2} x^3 y^2 dx dy$,

(b) $\int_{0}^{4} \int_{-1}^{3} 6x^2 y^2 dx dy$,

(c) $\int_{1}^{3} \int_{0}^{2} \int_{-1}^{4} x^3 y^2 z dx dy dz$.

1.4.5 Some Important Real Functions

Definition 1.35. *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is called a* ***monomial*** *if it can be written in the following form*

$$f(x_1, \ldots, x_n) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n},$$

where c *is any real number and* $a_i \in \mathbb{Z}_+$ *for all* $i \in \{1, \ldots n\}$. *The sum of the exponents* $(\sum_{i=1}^{n} a_i)$ *is called the* ***degree*** *of the monomial.*

Example 1.86. (a) $f(x_1, x_2) = -5x_1^2 x_2^4$ *is a monomial of degree* 6.

(b) $f(x_1, x_2, x_3) = 2x_1x_2^3x_3^4$ *is a monomial of degree* 8.

(c) Any constant real function can be written as $f(\mathbf{x}) = cx_1^0 \cdots x_n^0$. *It follows that a constant function is a monomial of degree 0.*

Definition 1.36. *A function* $f : \mathbb{R}^n \to \mathbb{R}$ *is called a* ***polynomial*** *if it can be written as the finite sum of monomials. The* ***degree*** *of the polynomial is the highest degree among its monomials.*

Example 1.87. $f(\mathbf{x}) = -5x_1^2x_2^4 + 2x_1x_2^3x_3^4$ *is a polynomial of degree* 8.

Definition 1.37. *A* ***rational function*** *is a function that can be written as the ratio of polynomials.*

Example 1.88. $f(x) = \frac{x^2+1}{x-1}$, $f(x) = \frac{x^5+4x}{5}$

We now consider two very important functions we haven't talked about yet: the logarithmic and the exponential function. There are a lot of different ways one can introduce these two functions. I will make use of integral calculus as it provides a rather short construction of these two functions.

Definition 1.38. *The* ***logarithmic function*** *is a function from* $\mathbb{R}_{++}$ *to* $\mathbb{R}$ *defined by*

$$\ln x := \int_1^x \frac{1}{t} dt.$$

Here are a few important properties of the logarithmic function that you should remember.

Proposition 1.89. *(a)* $\ln x$ *is strictly increasing and continuous. Moreover, it is a bijection.*

(b) $\ln 1 = 0$.

(c) $\ln x$ *is differentiable. Moreover,* $\frac{d}{dx} \ln x = \frac{1}{x}$ *for all* $x > 0$.

(d) If $u : \mathbb{R} \to \mathbb{R}_{++}$ *is a differentiable function, then* $\frac{d}{dx} \ln(u(x)) = \frac{u'(x)}{u(x)}$. *Moreover,* $\int \frac{u'(x)}{u(x)} dx = \ln u(x) + C$.

(e) $\ln xy = \ln x + \ln y$ *and* $\ln \frac{x}{y} = \ln x - \ln y$.

(f) $\ln x^n = n \ln x$.

As ln is a bijection, it is invertible with its inversion being a strictly increasing function from $\mathbb{R}$ onto $\mathbb{R}_{++}$. The inversion of ln is called the **exponential function** and is denoted e^x. As it is the inverse of ln, we get by definition

$$\ln e^x = x$$

and

$$e^{\ln x} = x$$

for all $x > 0$.

Again, let us go over a few properties of the exponential function.

Proposition 1.90. *(a) e^x is strictly increasing, continuous and a bijection. Its inverse is the logarithmic function.*

(b) $e^1 = e$. Hence, $\ln e = \int_1^e \frac{1}{t} dt = 1$. Note that as $\ln$ is strictly increasing, e is the only number which satisfies $\ln e = 1$.

(c) e^x is differentiable. Its derivative $\frac{d}{dx} e^x = e^x$ for all $x \in \mathbb{R}$. Moreover, $\int e^x dx = e^x + C$.

(d) If $u : \mathbb{R} \to \mathbb{R}$ is a differentiable function, then $\frac{d}{dx} e^{u(x)} = (e^{u(x)})u'(x)$. Moreover, $\int e^{u(x)} u'(x) dx = e^{u(x)} + C$.

(e) $e^{x+y} = e^x e^y$ and $e^{x-y} = \frac{e^x}{e^y}$ for all $x, y > 0$.

(f) $(e^x)^y = e^{xy}$.

Exercise 1.91. *Solve the following equations for x:*

(a) $2e^{6x} = 18$,

(b) $e^{x^2} = 1$,

(c) $\ln x^2 = 5$,

(d) $\ln x^{5/2} - 0.5 \ln x = \ln 25$.

Exercise 1.92. *Compute the first and the second derivatives of the following functions:*

(a) xe^{3x},

(b) e^{x^2+3x-2},

(c) $\ln(x^4 + 2)^2$,

(d) $\frac{x}{e^x}$,

(e) $\frac{x}{\ln x}$,

(f) $\frac{\ln x}{x}$.

Exercise 1.93. *Compute the indefinite integral of each of the following functions:*

(a) $3x^{-1/2} - x^{-1}$,

(b) $6e^{7x}$,

(c) $e^{(3x^2+6x)}(x+1)$,

(d) $\frac{3x^{1/2}+x^{-1/2}}{x^{3/2}+x^{1/2}}$.

Exercise 1.94. *Use integration by parts (see proposition 1.73) to compute the following integrals:*

(a) $\int x \ln x dx$,

(b) $\int x^2 e^x dx$.

1.5 Countability

We now ask how we can compare the size of two sets. This question is trivial if both sets are finite. After all, we can simply count the number of elements of each of two finite sets and compare these two numbers. When sets are infinite the answer to the question becomes less clear. On the one hand one might think that if two sets are infinite then they are equally crowded. On the other hand, one might wonder how it would be possible for the set of natural numbers which is a proper subset of the rationals to be as crowded. The question we are about to answer is: does there exist one infinity or several infinities? We first introduce a way to measure the relative size of sets.

Definition 1.39. *A set S is said to be* ***countably infinite*** *(or denumerable) if there exists a bijection f that maps S onto $\mathbb{N}$. S is said to be* ***countable*** *if it is either finite or countably infinite. S is* ***uncountable*** *if it is not countable.*

The important thing to understand in this context is that two sets have the same number of elements if, and only if, there exists a bijection between the two sets. Again an example with finite sets may provide you with some intuition. Let $X := \{a, b, c\}$ and $Y := \{1, 2, 3, 4\}$. It should be obvious that there is no way to construct a bijection between these two sets. Similarly, if there is a bijection between two sets then any element in any of these sets is put in a direct relation only with one element of the other set. Hence, these two sets must have the same number of elements. It is this fact that allows us to study the relative size of infinite sets.

Let us now compare the size of some infinite sets. First of all, note that $\mathbb{Z}_+$ is countable. After all, $f : \mathbb{Z}_+ \to \mathbb{N}$ defined by $f(i) = i + 1$ is a bijection. Similarly, $\mathbb{Z}$ is countable as $g : \mathbb{Z} \to \mathbb{Z}_+$ defined by $g(i) := \begin{cases} 0 \text{ if } i = 0 \\ 2i - 1 \text{ if } i > 0 \\ -2i \text{ if } i < 0 \end{cases}$ is a bijection. Interestingly, the set of natural numbers contains subsets that are countably infinite as well. For example, the set of even natural numbers, let's call it E, is countably infinite as $f : E \to \mathbb{N}$ defined by $f(n) := n/2$ is a bijection. As a matter of fact any infinite subsets of $\mathbb{N}$ is countably infinite for the following is true:

Proposition 1.95. *Every subset of a countable set is countable.*

It remains to study the countability of $\mathbb{Q}$ and $\mathbb{R}$. The following result will prove useful shortly in this regard:

Proposition 1.96. *A countable union of countable sets is countable.*

Now remark that $\mathbb{Q} = \cup\{X_n : n \in \mathbb{N}\}$ with $X_n := \{\frac{m}{n} : m \in \mathbb{Z}\}$. Observe that X_n is countable for all $n \in \mathbb{N}$ as X_n has exactly the same number of elements as $\mathbb{Z}$. But then $\mathbb{Q}$ is nothing but a countable union of countable sets. Hence,

Proposition 1.97. $\mathbb{Q}$ *is countable.*

Although, I'm not going to explain you why, note however that the following is true:

Proposition 1.98. $\mathbb{R}$ *is uncountable.*

This implies that infinity comes in different sizes. In some ways it probably would be more adequate to speak of infinities as it is true that there are infinitely many different sizes of infinity.

Proposition 1.99. *For any nonempty set X, there is no surjection of the form $f : X \to 2^X$.*

Exercise 1.100. *Prove that the function $f : \mathbb{Z}_+ \times \mathbb{Z}_+ \to \mathbb{Z}_+$ defined by*

$$f(i, j) := \frac{(i + j - 2)(i + j - 1)}{2} + i$$

is one-to-one and onto. What does this imply for $\mathbb{Z}_+ \times \mathbb{Z}_+$?

Exercise 1.101. *Using the propositions 1.97 and 1.98 show that $\mathbb{R} \setminus \mathbb{Q}$ is uncountable.*

2 Probability Theory

2.1 Introduction to Probability

2.1.1 The Notion of Probability

In the social sciences we are often confronted with situations or events that appear to have a random component. This is usually expressed by saying that someone is uncertain about certain aspects of the world. Probability provides a formal model of uncertainty that allows us to measure the level of uncertainty. Other models of uncertainty can and have been developed. However, probability has been the most successful and is the most commonly used model of uncertainty. We start with the notion of an **experiment**. In probability theory, an experiment is any process whose outcome is not known in advance with certainty. The set of all possible outcomes of an experiment is called the **sample space** (or **outcome space**). An **event** is a subset of the sample space.

Example 2.1. *Consider the experiment of rolling a six-sided die. The sample space is then the set* $\{1, 2, 3, 4, 5, 6\}$*. The set of all events is then* $2^{\{1,2,3,4,5,6\}}$*. For example we may want to study the event that an even number occurred, i.e. the event* $\{2, 4, 6\}$*. Or we may want to study the event that a number greater than 3 occurs, i.e. the event* $\{4, 5, 6\}$*, and so forth.*

Our goal is now to define a function that assigns to any event A in the sample space S a number $Pr(A)$ that gives us the probability that A will occur. To form a sensitive model of uncertainty, this function is required to satisfy three axioms:

Axiom 1 For any event A, $Pr(A) \geq 0$.

Axiom 2 $Pr(S) = 1$.

Axiom 3 For any infinite sequence of disjoint events $A_1, A_2, \ldots$,

$$Pr(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i).$$

Any function that satisfies the three axioms above is called a **probability distribution** (or **probability measure**).

Proposition 2.2 (Properties of a probability distribution)**.** *1.* $Pr(\emptyset) = 0$.

2. (finite additivity) For any $n \in \mathbb{N}$*, and any pairwise disjoint events* $A_1, \ldots, A_n$*, we have*

$$Pr(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(A_i).$$

3. (subadditivity) For any sequence of events $A_1, A_2, \ldots$,

$$Pr(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} Pr(A_i).$$

4. For any event A*,* $Pr(A^c) = 1 - Pr(A)$.

5. (monotonicity) If $A \subseteq B$*, then* $Pr(A) \leq Pr(B)$.

6. For any event A*, we have* $0 \leq Pr(A) \leq 1$.

7. For any two events A *and* B*, we have* $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.

8. (Exclusion-Inclusion Formula) For any n *events* $A_1, \ldots A_n$,
$Pr(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(A_i) - \sum_{i<j} Pr(A_i \cap A_j) + \sum_{i<j<k} Pr(A_i \cap A_j \cap A_k) - \sum_{i<j<k<l} Pr(A_i \cap A_j \cap A_k \cap A_l) + \ldots + (-1)^{n+1} Pr(A_1 \cap A_2 \cap \cdots \cap A_n)$.

Exercise 2.3. *Consider a box containing five balls of different colors: red, green, yellow, blue, black. We are told that the probability that the selected ball will be blue is* $2/5$ *and the probability that it will be red is* $1/3$*. What is the probability that the selected ball will be green, yellow or black?*

Exercise 2.4. *Consider two events* A *and* B*. Let the respective probabilities of these two events be* $Pr(A) = 1/4$ *and* $Pr(B) = 1/3$ *respectively. Compute the probability of* $Pr(A^c \cap B)$ *in the three following cases: a)* $A \cap B = \emptyset$*, b)* $A \subset B$*, c)* $Pr(A \cap B) = 1/8$*.*

Exercise 2.5. *Let us again consider two events* A *and* B *which occur with probability* 0.4 *and* 0.7 *respectively. What is the maximum possible value of* $Pr(A \cap B)$*? What is the minimum value? Under what conditions is the maximum attained? Under what conditions the minimum?*

Exercise 2.6. *Prove that for any two events* A *and* B*,*

$$Pr(A) = Pr(A \cap B) + Pr(A \cap B^c).$$

Exercise 2.7. *A box contains 30 red balls, 30 white balls, and 30 blue balls. If 10 balls are selected randomly, without replacement, what is the probability that at least one color will be missing from the selection?*

If the sample space has only a finite number of outcomes $s_1, \ldots, s_n$, then we usually specify the probability distribution on S by assigning a number p_i to each outcome $s_i \in S$. By the axioms of probability, it must be the case that $p_i \geq 0$ for all $i = 1, \ldots, n$ and $\sum_{i=1}^{n} p_i = 1$. The probability of an event A is then simply the sum of the probabilities p_i of all outcomes $s_i \in A$. A finite sample space is called **simple** if each outcome s_i has the same probability $1/n$ of occurring. Hence, the probability of an event A in a simple probability space is simply the number of outcomes in A divided by the number of outcomes in S.

Example 2.8. *Let us consider the example of the six-sided die again. If we assume that the die is fair, then each number* 1 *through* 6 *occurs with probability* $\frac{1}{6}$*. The probability of the event* $\{2, 4, 6\}$ *is then equal to* $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$*. Note that this is an example of a simple probability space.*

Exercise 2.9. *Let us roll two balanced six-sided dice. What is the probability that the sum of the two numbers that appear is odd? What is the probability that it is even? What is the probability that the difference between the two numbers will be less than 4?*

Exercise 2.10. *Let us toss 4 fair coins. What is the probability that each coin will show the same face of the coin?*

Exercise 2.11. *A PhD program contains students in years 1,2,3,4,5,6. Imagine that the years 2 to 6 all have the same number of students but that the entering class has twice as many students. Now let us randomly choose a PhD student from the list of all PhD students. What is the probability that this student will be in his third year?*

2.1.2 Combinatorics

If the number of outcomes in S is very large, it becomes difficult or even impossible to list all the outcomes and count their number one by one. Combinatorial methods, to which we now turn, can be useful in such instances.

We start with the **multiplication rule**. Consider an experiment that takes place in several stages. For each stage n, there are m_n possible outcomes. Note that the number of outcomes can vary from stage to stage and does not depend on the number of outcomes in previous stages. We want to know how many different outcomes the grand experiment has. It turns out that the grand experiment has $\prod_{i=1}^{n} m_n$ possible outcomes.

Example 2.12. *Consider a survey of a randomly selected voter. We ask this voter three question sequentially. 1) On a scale of 1 to 7, where 1 is far to the left and 7 is far to the right, what do you think is the policy position of the democratic candidate? 2) On a scale of 1 to 7, where 1 is far to the left and 7 is far to the right, what do you think is the policy position of the republican candidate? 3) Are you going to vote on the next election (possible answers : yes, no, don't know)? Note that the person asking these questions does not know with certainty which responses the voter is going to give. Each question can therefore be thought of as an experiment. If we consider an outcome to be a list (x, y, z), where x is the answer to the first question, y to the second and z to the third question, then the overall number of possible outcomes is* $7 \cdot 7 \cdot 3 = 147$.

One implication of the multiplication rule is that it gives us a formula for the number of outcomes of a process of **sampling with replacement**. Sampling with replacement denotes the following type of experiment. Consider a box with n balls numbered 1,. . . , n. In the first stage, a ball is selected randomly from the box and its number is noted. Then, the ball is put back into the box and the experiment is repeated k times. An outcome of this experiment is an ordered list $(a_1, a_2, \ldots, a_k)$, where a_1 is the number of the ball that was selected in the first stage, a_2 the number of the ball selected in the second stage, and so forth. It follows trivially from the multiplication rule that the number of outcomes of the grand experiment is n^k.

Now imagine that in each round the selected ball is not put back into the box. Again an outcome of the experiment will be an ordered list $(a_1, a_2, \ldots, a_k)$. Note however that in this case $a_i \neq a_j$ for all $i \neq j$. Such a process is called **sampling without replacement**. Then the number of possible outcomes is $\frac{n!}{(n-k)!} = n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)$.

Example 2.13. *One very famous and interesting example of when these types of methods are going to be used is the* ***birthday problem****. Here is the question we ask: What is the probability that at least two people in a group with k people* $(2 \leq k \leq 365)$ *have their birthday on the same day? To make the problem simpler, let us assume that the birthdays of the k people are unrelated (no twins) and that the probability of being born on a given day is the same for all days of the year. Let us first compute the number of possible outcomes. This is a simple application of sampling with replacement, i.e. the number of possible outcomes is* 365^k*. Now note that the number of outcomes such that each person is born on a different day is a case of sampling without replacement, i.e. this number is* *equal to* $\frac{365!}{(365-k)!}$. *It follows that the probability that no two people have their birthday the same day is equal to*

$$\frac{365!}{(365-k)!} \cdot \frac{1}{365^k}.$$

The trick now is to recognize that the event that at least two people have their birthday the same day is nothing but the complement of the event that no two people have their birthday on the same day. Hence, the answer to the birthday problem is the following

$$p = 1 - \frac{365!}{(365-k)!} \cdot \frac{1}{365^k}.$$

Interestingly enough if $k = 23$, *this probability is* 0.507, *which is already quite high.*

Suppose now that we want to choose k balls out of the box but that unlike before we don't care about the order in which the balls were selected. If we use sampling with replacement then the number of outcomes is equal to $\binom{n+k-1}{k} := \frac{(n+k-1)!}{(n-1)!k!}$. If we use sampling without replacement then the number of outcomes is equal to $\binom{n}{k} := \frac{n!}{k!(n-k)!}$.

Exercise 2.14. *In how many different ways can you arrange the letters* a, b, c, d, e, f*?*

Exercise 2.15. *Let us roll four dice. What is the probability that we get four different numbers? Now let us roll six dice. What is the probability that each number appears exactly once?*

Exercise 2.16. *Consider a box containing 100 balls,* r *of which are red. Let us draw the balls randomly one by one out of the box without replacement. Compute*

(a) The probability that the first ball drawn is red.

(b) The probability that the 50th ball will be red.

(c) The probability that the last ball drawn will be red.

The number $\binom{n}{k}$ is called a **binomial coefficient** as it appears in the binomial theorem.

Theorem 2.17 (Binomial Theorem). *For all numbers* x *and* y *and any positive integer* n*,*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Suppose now that we want to divide n distinct elements into k different groups with $k \geq 2$ such that for $j = 1, \ldots, k$, the jth group contains exactly n_j elements with $n_1 + n_2 + \ldots n_k = n$. What is the number of different ways in which we can divide the n elements into k groups. It turns out that this number is equal to $\binom{n}{n_1, n_2, \ldots, n_k} := \frac{n!}{n_1! n_2! \cdots n_k!}$.

The number $\binom{n}{n_1, n_2, \ldots, n_k}$ is called a **multinomial coefficient** as it appears in the multinomial theorem.

Theorem 2.18 (**Multinomial Theorem**). *For all numbers $x_1, \ldots, x_k$ and any positive integer n,*

$$(x_1 + \cdots + x_k)^n = \sum \binom{n}{n_1, n_2, \ldots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k},$$

where the summation is taken over all possible combinations of nonnegative integers $n_1, \ldots, n_k$ such that $n_1 + \cdots + n_k = n$.

Exercise 2.19. *Let us form two committees out of a group of 400 people. The first committee has five members, while the second has seven. In how many different ways can these committees be selected?*

Exercise 2.20. *If the letters $s, s, s, t, t, t, i, i, a, c$ are arranged randomly, what is the probability that we will get the word 'statistics'?*

2.2 Conditional Probability and Independence

2.2.1 Conditional Probability

Imagine that we perform an experiment and we learn that the event B has occurred. We now ask how the probability of another event A changes as we learn that B has occurred. This new probability of A given that B has occurred is called the **conditional probability** of A given B and is denoted by $Pr(A|B)$. Obviously, if we know that B has occurred, we also learn that the outcome of the experiment is an element of B. Hence, the new probability of A will have to be based on the set of outcomes that are in B and in A. Obviously, the more outcomes in A are also outcomes in B, the higher the probability of A given B will be. This leads us to the following definition.

$$Pr(A|B) := \frac{Pr(A \cap B)}{Pr(B)}.$$

Note that we have to require that $Pr(B) > 0$ as otherwise $Pr(A|B)$ would not be defined.

Example 2.21. *Let us roll two fair dice. Note first that the total number of outcomes of this experiment is $6 \cdot 6 = 36$. Let us denote an outcome of this experiment by (a, b), where a refers to the number of the first die and b to the one of the second die. We are told that the sum of the two numbers that occurred (let's call this sum k) is odd. Let us call the event that an odd number has occurred B and note that $B = \{3, 5, 7, 9, 11\}$. We now want to find the probability that $k < 8$ given that k is odd. Let us call the event that $k < 8$ A and note that it is equal to the set $\{2, 3, 4, 5, 6, 7\}$. In order to compute the conditional probability of A given B, we first need to compute $Pr(A \cap B)$. Remark that $A \cap B = \{3, 5, 7\}$. Hence,*

$$Pr(A \cap B) = Pr(\{3\}) + Pr(\{5\}) + Pr(\{7\}).$$

So let us compute these probabilities in turn. First, note that two outcomes lead to $k = 3$, namely $(1, 2)$ and $(2, 1)$. Hence,

$$Pr(\{3\}) = \frac{2}{36}.$$

Similarly, the outcomes that lead to $k = 5$ *are* $(1,4),(4,1),(2,3),(3,2)$. *Hence,*

$$Pr(\{5\}) = \frac{4}{36}.$$

Finally, the outcomes that lead to $k = 7$ *are* $(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)$ *and thus*

$$Pr(\{7\}) = \frac{6}{36}.$$

It follows that

$$Pr(A \cap B) = \frac{2+4+6}{36} = \frac{1}{3}.$$

Next, we need to compute the $Pr(\{B\})$ *which is equal to* $Pr(\{3\}) + Pr(\{5\}) + Pr(\{7\}) + Pr(\{9\}) + Pr(\{11\})$. *We already know the first three numbers of this sum, so let us compute* $Pr(\{9\})$. *Note that the following outcomes lead to* $k = 9$*:* $(3,6),(6,3),(4,5),(5,4)$. *Hence,*

$$Pr(\{9\}) = \frac{4}{36}.$$

Finally, the outcomes that result in $k = 11$ *are* $(5,6)$ *and* $(6,5)$ *and thus*

$$Pr(\{11\}) = \frac{2}{36}.$$

It follows that

$$Pr(B) = \frac{2+4+6+4+2}{36} = \frac{1}{2}.$$

We conclude that

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

For certain experiments the computation of the conditional probabilities is straightforward. When this is the case the conditional probabilities can be used to compute the probability of the intersection of two events. In fact it follows directly from the definition of $Pr(A|B)$ that

$$Pr(A \cap B) = Pr(A|B)Pr(B) = Pr(B|A)Pr(A).$$

This principle can be extended to more than two events.

Proposition 2.22. *Consider the events* $A_1, A_2, \dots A_n$ *such that* $Pr(\cap_{i=1}^{n-1} A_i) > 0$. *Then*

$$Pr(\cap_{i=1}^{n} A_i) = Pr(A_1)Pr(A_2|A_1)Pr(A_3|A_1 \cap A_2) \cdots Pr(A_n| \cap_{i=1}^{n-1} A_i).$$

Example 2.23.

Similarly, the following is true.

Proposition 2.24. *Let* $A_1, A_2, \dots, A_n, B$ *be events such that* $Pr(\cap_{i=1}^{n-1} A_i|B) > 0$. *Then*

$$Pr(\cap_{i=1}^{n} A_i|B) = Pr(A_1|B)Pr(A_2|A_1 \cap B) \cdots Pr(A_n| \cap_{i=1}^{n-1} A_i \cap B).$$

Exercise 2.25. *Assume $A \subset B$ and $Pr(B) > 0$. Compute $Pr(A|B)$.*

Exercise 2.26. *Assume $A \cap B = \emptyset$ and $Pr(B) > 0$. Compute $Pr(A|B)$.*

Exercise 2.27. *Let S be the outcome space of the experiment and let $A \subseteq S$. What is $Pr(A|S)$?*

Exercise 2.28. *Consider a box that contains three cards. One card is red on both sides, one is green on both sides and one is green on one side and red on the other. Let us select one card at random out of the box and observe the color on one of its sides. If the color observed is green, what is the probability that the other side is green as well?*

2.2.2 Independence of Events

Imagine that the probability of A given B is different from the probability of A, i.e. $Pr(A) \neq Pr(A|B)$. Then the fact that B has occurred teaches us something about the probability that A has occurred as well. Note however that $Pr(A|B)$ need not be different from $Pr(A)$. There will be cases where $Pr(A) = Pr(A|B)$, and the occurrence of B won't teach us anything about the occurrence of A. In this latter case, we say that the events A and B are independent.

Definition 2.1. *Two events A and B are **independent** if*

$$Pr(A \cap B) = Pr(A)Pr(B).$$

Note that if $Pr(A) > 0$ and $Pr(B) > 0$, then A and B independent can also be expressed as $Pr(A|B) = Pr(A)$ or $Pr(B|A) = Pr(B)$.

Example 2.29. *Let us roll the die again and consider the event $A := \{2, 4, 6\}$ and the event $B := \{1, 2, 3, 4\}$. Note that A is the event that an even number has occurred, while B is the event that a number lower than 5 has been obtained. Note that*

$$Pr(A) = Pr(\{2\}) + Pr(\{4\}) + Pr(\{6\}) = \frac{3}{6} = \frac{1}{2}.$$

Similarly,

$$Pr(B) = Pr(\{1\}) + Pr(\{2\}) + Pr(\{3\}) + Pr(\{4\}) = \frac{4}{6} = \frac{2}{3}.$$

Now note that $A \cap B = \{2, 4\}$, and hence

$$Pr(A \cap B) = Pr(\{2\}) + Pr(\{4\}) = \frac{2}{6} = \frac{1}{3} = \frac{1}{2} \cdot \frac{2}{3} = Pr(A)Pr(B).$$

It follows that the events A and B are independent.

Here is a useful observation about independent events.

Proposition 2.30. *If A and B are independent, then so are A and B^c, A^c and B, and A^c and B^c.*

Exercise 2.31. *Prove Proposition 2.26.*

We can extend the notion of independence to more than two events. The extension however is not entirely obvious.

Definition 2.2. *A nonempty collection of events $\mathcal{A}$ is said to be independent if for any finite subset $\mathcal{S}$ of $\mathcal{A}$, we have*

$$Pr(\cap \mathcal{S}) = \prod_{A \in \mathcal{S}} Pr(A).$$

Let us consider an example to make sure the notion of independence of several events is understood. Let $\mathcal{A}$ contain three events A_1, A_2, A_3, i.e. $\mathcal{A} := \{A_1, A_2, A_3\}$. Then these three events are independent if the following is true

$$Pr(A_1 \cap A_2) = Pr(A_1)Pr(A_2),$$

$$Pr(A_1 \cap A_3) = Pr(A_1)Pr(A_3),$$

$$Pr(A_2 \cap A_3) = Pr(A_2)Pr(A_3),$$

and

$$Pr(A_1 \cap A_2 \cap A_3) = Pr(A_1)Pr(A_2Pr(A_3).$$

One may wonder why it is not enough simply to require that the events are pairwise independent or only to require that $Pr(\cap \mathcal{A}) = \prod_{A \in \mathcal{A}} Pr(A)$. Well as it turns out none of these conditions necessarily follows from the others. As we will see in the next example it could be that the events are pairwise independent but fail to be independent all together. Similarly it could be that $Pr(\cap \mathcal{A}) = \prod_{A \in \mathcal{A}} Pr(A)$ holds yet some events fail to be pairwise independent.

Example 2.32. *Let us consider an experiment with four possible outcomes $\{s_1, s_2, s_3, s_4\}$. Let us assume that each of these outcomes occurs with equal probability $1/4$. Now consider the following three events $A := \{s_1, s_2\}$, $B := \{s_1, s_3\}$, and $C := \{s_1, s_4\}$. Note that*

$$Pr(A) = Pr(B) = Pr(C) = 1/2.$$

Moreover, $A \cap B = A \cap C = B \cap C = A \cap B \cap C = \{s_1\}$. It follows that

$$Pr(A \cap B) = Pr(A \cap C) = Pr(B \cap C) = Pr(A \cap B \cap C) = 1/4.$$

In words, this means that A, B, and C are pairwise independent while the three events are not.

Recall that if A and B both have positive probability and are independent then $Pr(A|B) = Pr(A)$. A similar connection between conditional probability and independence also holds for more than two events.

Proposition 2.33. *Let $A_1, \ldots, A_k$ be events such that $Pr(\cap_{i=1}^k A_i) > 0$. Then $A_1, \ldots A_k$ are independent iff for any two disjoint subsets $\{i_1, \ldots, i_m\}$ and $\{j_1, \ldots j_l\}$ of $\{1, \ldots, k\}$, the following holds*

$$Pr(A_{i_1} \cap \cdots \cap A_{i_m} | A_{j_1} \cap \cdots \cap A_{j_l}) = Pr(A_{i_1} \cap \cdots \cap A_{i_m}).$$

We now introduce the notion of conditional independence.

Definition 2.3. *A nonempty collection of events $\mathcal{A}$ is said to be conditionally independent given B if for any finite subset $\mathcal{S}$ of $\mathcal{A}$,*

$$Pr(\cap \mathcal{S} | B) = \prod_{A \in \mathcal{S}} Pr(A|B).$$

Proposition 2.34. *Let A_1, A_2, and B be events such that $Pr(A_1 \cap B) > 0$. Then A_1 and A_2 are conditionally independent given B iff $Pr(A_2|A_1 \cap B) = Pr(A_2|B)$.*

Exercise 2.35. *Let A be any event such that $Pr(A) = 0$. Prove that A is independent of any other event B.*

Exercise 2.36. *Let us roll two dice three times in a row. What is the probability that the sum of the two numbers that appear is 7 on all three rolls?*

Exercise 2.37. *Two students* jean *and* jacques *are registered for a class. Jean attends class 80% of the time and Jacques 60% of the time. Note that their absences are independent.*

(a) What is the probability that at least one of the two will be in class on a given day?

(b) If at least one of the two students is in class on a given day, what is the probability that Jean is in class that day?

Exercise 2.38. *A box contains 20 red balls, 30 green balls, and 50 blue balls. Suppose that 10 balls are successively selected at random with replacement. What is the probability that at least one color will be missing from the 10 selected balls?*

2.2.3 Bayes' Theorem

We now state and prove Bayes' Theorem which is extensively used in statistics as well as in game theory. But first we need some additional vocabulary.

Definition 2.4. *Consider any nonempty set S and let $\mathcal{F}$ be a subset of 2^S. In other words $\mathcal{F}$ is a set that contains subsets of S. We say that $\mathcal{F}$ is a partition of S, if the following three conditions hold*

(a) $\cup \mathcal{F} = S$,

(b) $\mathcal{F}$ *is pairwise disjoint,*

(c) *Any set* $A \in \mathcal{F}$ *is nonempty.*

Example 2.39. *Let* $S := \{1, 2, 3, 4, 5, 6\}$*. Then* $\mathcal{F} := \{\{1, 2\}, \{3\}, \{4, 5, 6\}\}$ *is a partition of* S*.*

Proposition 2.40. *(Law of total probability) Assume that the events* $A_1, \ldots A_k$ *form a partition of the outcome space* S *and that* $Pr(A_i) > 0$ *for all* $i \in \{1, \ldots, k\}$*. Then for any event* B *in* S*, we have*

$$Pr(B) = \sum_{i=1}^{k} Pr(A_i)Pr(B|A_i).$$

Proof. Let $A_1, \ldots, A_k$ form a partition of S and let B be any event in S. Then

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_k \cap B).$$

Moreover, $(A_i \cap B) \cap (A_j \cap B) = \emptyset$ for all $i \neq j$. Therefore

$$Pr(B) = \sum_{i=1}^{k} Pr(A_i \cap B). \tag{1}$$

Now as $Pr(A_i) > 0$ for all i, we have

$$Pr(A_i \cap B) = Pr(A_i)Pr(B|A_i). \tag{2}$$

Plugging (2) into (1) we get

$$Pr(B) = \sum_{i=1}^{k} Pr(A_i)Pr(B|A_i).$$

□

Proposition 2.41. *(Bayes' Theorem) Assume that the events* $A_1, \ldots A_k$ *form a partition of the outcome space* S *and that* $Pr(A_i) > 0$ *for all* $i \in \{1, \ldots, k\}$*. Assume further that* B *is an event such that* $Pr(B) > 0$*. Then, for all* $j \in \{1, \ldots, k\}$*, we have*

$$Pr(A_j|B) = \frac{Pr(A_j)Pr(B|A_j)}{\sum_{i=1}^{k} Pr(A_i)Pr(B|A_i)}.$$

Proof. By the definition of conditional probability, we have

$$Pr(A_j|B) = \frac{Pr(A_j \cap B)}{Pr(B)}.$$

Now as $Pr(A_j) > 0$ for all A_j, we have

$$Pr(A_j \cap B) = Pr(A_j)Pr(B|A_j).$$

Finally, by the law of total probability just stated, we have

$$Pr(B) = \sum_{i=1}^{k} Pr(A_i)Pr(B|A_i).$$

It follows that

$$Pr(A_j|B) = \frac{Pr(A_j)Pr(B|A_j)}{\sum_{i=1}^{k} Pr(A_i)Pr(B|A_i)}.$$

□

Example 2.42. *Let us consider the game* Let's Make a Deal *presented by Monte Hall. A contestant in this game has to choose one of three doors (let's call them A_1, A_2, and A_3). Behind one of the doors, there is a luxury car, while behind the other doors there are prices of comparatively low value. The ex ante probability that the car is behind one of the doors is equal for all doors, i.e. $Pr(A_1) = Pr(A_2) = Pr(A_3) = 1/3$. The contestant chooses a door, let's say door A_3. The show master now opens one of the remaining doors (i.e. A_1 or A_2) and reveals that the car is not behind that door. Once this is done, the contestant gets to choose between keeping door A_3 or switching to the only other door left. The question is the following: should the contestant switch?*

Obviously, the probability that the car is behind A_3 hasn't changed and is still $1/3$. To compute the probability of winning when switching let's call B_1 and B_2 the event that the show master has revealed that the car is not behind door 1 or door 2 respectively. We want to use Bayes rule to compute $Pr(A_1|B_2)$ and $Pr(A_2|B_1)$. To do so, we first need an expression for $Pr(B_i|A_j)$. Note that to make the show interesting for TV watchers, Monte will never expose the car. It follows that $Pr(B_1|A_1) = Pr(B_2|A_2) = 0$. Moreover, if the car is behind door A_3, we assume that the show master opens one of the remaining doors randomly and hence $Pr(B_1|A_2) = Pr(B_2|A_1) = 1$, and $Pr(B_1)|A_3 = Pr(B_2|A_3) = 1/2$. Now everything is set for us to compute $Pr(A_1|B_2)$ via Bayes' rule.

$$\begin{aligned} Pr(A_1|B_2) &= \frac{Pr(B_2|A_1)Pr(A_1)}{Pr(B_2|A_1)Pr(A_1) + Pr(B_2|A_2)Pr(A_2) + Pr(B_2|A_3)Pr(A_3)} \\ &= \frac{1 \cdot 1/3}{1 \cdot 1/3 + 0 \cdot 1/3 + 1/2 \cdot 1/3} \\ &= \frac{2}{3}. \end{aligned}$$

The computation of $Pr(A_2|B_1)$ is similar and leads to the same result. It follows that if you ever happen to appear on the show Let's Make a Deal and Monte Hall offers you to switch, you should switch.

Exercise 2.43. *A box contains three coins with head on each side, four coins with a tail on each side, and two fair coins. Let us select one coin out of the box at random and let us toss it once. What is the probability that a head is obtained?*

Exercise 2.44. *In a given town,* 30% *of the voters are Democrats and* 70% *are Republican.* 40% *of the Democrats support the president's budget, while it is supported by* 80% *of the Republicans. If a randomly selected voter supports the president's budget, what is the probability that the voter is a Democrat?*

Exercise 2.45. *Assume* 2% *of the population of the USA are members of some extremist militia group ($Pr(M) = 0.02$), a fact that some members may not want to admit to an interviewer. We develop a survey that is* 95% *accurate on positive classification,* $Pr(C|M) = 0.95$, *and* 97% *accurate on negative classification,* $Pr(C^c|M^c) = 0.97$. *Using Bayes' rule, what is the probability that someone positively classified by the survey as being a militia member really is a militia member?*

2.3 Random Variables and Their Distributions

2.3.1 Random Variables and Distributions

Definition 2.5. *By a* ***random variable*** *we mean a real function on the outcome space S, i.e. $X : S \to \mathbb{R}$ is a random variable on S.*

Example 2.46. *Consider the experiment of tossing a coin 5 times. An outcome of this experiment is then a list of five elements which indicate whether the outcome was head (H) or Tail (T) at the n-th repetition. Here is an example (H,T,T,H,T). By the multiplication rule the total number of outcomes of this experiment is* 2^5. *Imagine that we are interested in the number of times we obtained head in our five tosses. We could then define a random variable X which would assign the number of heads obtained for all outcomes $s \in S$. For example, if we let $s := (H, T, T, H, T)$, then $X(s)$ would be* 2.

The reason we introduce random variables is that it is often easier to work with real numbers than it is to work with the outcomes of the original experiment.

Note that for any experiment we usually have a probability distribution on the outcome space S. From this distribution, we can deduce a probability distribution for the random variable. Note that the original distribution we start with is a function that assigns a probability to subsets of S. The probability distribution of our random variable X however is a function that assigns a probability to subsets of the real line. Here is its definition.

Definition 2.6. *Let A be a subset of the real line and let $Pr(X \in A)$ be the probability that the value of X will belong to A. Then $Pr(X \in A)$ is called the* ***distribution*** *of the random variable X and is defined by*

$$Pr(X \in A) := Pr(\{s \in S : X(s) \in A\}).$$

Definition 2.7. *We say that a random variable has a* ***discrete distribution*** *or is a* ***discrete random variable*** *if its range is finite or if its range contains only an infinite sequence of different values $x_1, x_2, \ldots$. If a random variable is discrete we define its probability distribution by the means of a* ***probability mass function*** *(pmf)*

$$f(x) = Pr(X = x).$$

Example 2.47. *(a) One simple example is the* ***uniform distribution*** *over a finite set of integers. Let's say that the range of our random variable is the set* $\{1, 2, \ldots, k\}$ *and that each of these integers occurs with equal probability. Then, the probability mass function of our random variable is*

$$f(x) = \begin{cases} \frac{1}{k} \text{ if } x = 1, \ldots, k, \\ 0 \text{ otherwise} \end{cases}.$$

1. *(b) Suppose we have an experiment which either fails (with probability* $1 - p$*) or succeeds (with probability* p*) and let's suppose we repeat the experiment* n *times. Then the probability of* x *successes out of the* n *trials is given by*

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} \text{ if } x = 0, 1, \ldots, k, \\ 0 \text{ otherwise} \end{cases}.$$

This distribution is called the ***binomial distribution****.*

Exercise 2.48. *Suppose that a random variable* X *has a uniform distribution on the integers* $10, \ldots, 20$*. What is the probability that* X *is even?*

Exercise 2.49. *Suppose that a random variable* X *has a discrete distribution with the following probability distribution:*

$$f(x) = \begin{cases} cx \text{ for } x = 1, \ldots, 5, \\ 0 \text{ otherwise }. \end{cases}$$

What is the value of the constant c*?*

Exercise 2.50. *Let us roll two dice and let* X *be the absolute value of the difference between the two numbers that appear. Determine the probability distribution of* X*.*

Exercise 2.51. *Suppose that a fair coin is tossed 5 times independently. Determine the probability distribution of the number of heads that will be obtained.*

Exercise 2.52. *Suppose that a box contains seven red balls and three blue balls. If five balls are selected at random, without replacement, determine the probability distribution of the number of red balls that will be obtained.*

Exercise 2.53. *Show that there does not exist any number* c *such that the following function would be a probability distribution:*

$$f(x) = \begin{cases} \frac{c}{x} \text{ for } x = 1, 2, \ldots \\ 0 \text{ otherwise.} \end{cases}$$

Definition 2.8. *We say that a random variable has* ***continuous distribution*** *or is a* ***continuous random variable*** *if there exists a function* $f : \mathbb{R} \to \mathbb{R}$ *that satisfies the following properties*

(a) $f(x) \geq 0$ *for all* $x \in \mathbb{R}$,

(b) f *is bounded and Riemann integrable,*

(c) $\int_{-\infty}^{\infty} f(x)dx = 1$,

(d) $Pr(a < X \leq B) = \int_a^b f(x)dx$ *for any interval* $(a, b] \subseteq \mathbb{R}$.

Note that f is called the ***probability density function*** *(or simply pdf) of the random variable* X.

A few remarks on the interpretation of the density are warranted. The density does not give us the probability that our random variable X takes on a particular value x as this probability is always 0 in the case of continuous random variables, i.e. $Pr(X = x) = 0$. However, it gives us the limit of $Pr(X \in [x - \epsilon, x + \epsilon])$ as ϵ goes to 0.

Example 2.54. *We will now introduce the* ***uniform distribution on an interval*** *(also called the* ***continuous uniform distribution****). Let* $a, b \in \mathbb{R}$ *with* $a < b$ *and let us assume that the probability that our random variable* X *takes on a value in any subinterval of* $[a, b]$ *is proportional to the length of this subinterval. Then the probability density function of* X *will be*

$$f(x) = \begin{cases} \frac{1}{b-a} \text{ if } x \in [a, b] \\ 0 \text{ otherwise} \end{cases}.$$

Exercise 2.55. *Let*

$$f(x) = \begin{cases} \frac{4}{3}(1 - x^3) \text{ for } 0 < x < 1, \\ 0 \text{ otherwise.} \end{cases}$$

be the probability density function of a random variable X.
Compute $Pr(X < \frac{1}{2})$, $Pr(\frac{1}{4} < X < \frac{3}{4})$, *and* $Pr(X > \frac{1}{3})$.

Exercise 2.56. *Let* X *have a uniform distribution on the interval* $[-2, 8]$. *Compute the probability density function of* X *and the value of* $Pr(0 < X < 7)$.

Exercise 2.57. *Suppose the probability density function of a random variable* X *is defined by*

$$f(x) := \begin{cases} cx^2 \text{ for } 1 \leq x \leq 2, \\ 0 \text{ otherwise.} \end{cases}$$

(a) *Compute* c.

(b) *Compute* $Pr(X > \frac{3}{2})$.

Exercise 2.58. *Show that there does not exist a number c such that the following function* $f(x)$ *would be a probability density function:*

$$f(x) = \begin{cases} \frac{c}{x} \text{ for } 0 < x < 1, \\ 0 \text{ otherwise.} \end{cases}$$

2.3.2 The Distribution Function

Definition 2.9. *The function $F : \mathbb{R} \to [0, 1]$ defined by*

$$F(x) := Pr(X \leq x)$$

is called the ***distribution function*** *(or* ***cumulative distribution function****) of the random variable X. Note that this function is defined for discrete random variables as well as for continuous random variables. When X is a discrete random variable, we have*

$$F(x) = \sum_{i \in \mathbb{N}: x_i \leq x} Pr(X = x_i).$$

When X is continuous however,

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Note that this last equation implies that $F'(x) = f(x)$ at any point $x \in \mathbb{R}$ at which F is continuous.

Example 2.59. *(a) Let X be a discrete random variable with pmf f defined by*

$$f(x) := \begin{cases} \frac{1}{15}x \text{ if } x = 1, 2, 3, 4, 5 \\ 0 \text{ otw} \end{cases}.$$

Then, $F(x) = \begin{cases} 0 \text{ if } -\infty < x < 1 \\ \frac{1}{15} \text{ if } 1 \leq x < 2 \\ \frac{3}{15} \text{ if } 2 \leq x < 3 \\ \frac{6}{15} \text{ if } 3 \leq x < 4 \\ \frac{10}{15} \text{ if } 4 \leq x < 5 \\ 1 \text{ if } 5 \leq x < \infty \end{cases}.$

(b) Let

$$f(x) = \begin{cases} \frac{4}{3}(1 - x^3) \text{ for } 0 < x < 1, \\ 0 \text{ otherwise.} \end{cases}$$

be the probability density function of a random variable X. Then,

$$F(x) = \begin{cases} 0 \text{ if } -\infty < x \leq 0 \\ \frac{4}{3}(x - \frac{1}{4}x^4) \text{ if } 0 < x < 1 \\ 1 \text{ if } 1 \leq x < \infty \end{cases}.$$

One important thing to remember about the distribution function is that it is defined for any real number x and not only for those numbers which occur with positive probabilities.

Here are three important properties of the distribution function that you should remember.

Proposition 2.60. *(a) F is increasing,*

(b) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

(c) F is continuous from the right.

If the distribution function of a random variable is known, it can be used to compute the distribution of a random variable.

Proposition 2.61. *(a) For all $x \in \mathbb{R}$, $Pr(X > x) = 1 - F(x)$.*

(b) For all $x_1, x_2 \in \mathbb{R}$ such that $x_1 < x_2$, $Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1)$.

Exercise 2.62. *Let X be a random variable that can only take the values $-1, 1, 3,$ and 4 with $Pr(-1) = 0.4$, $Pr(1) = 0.2$, $Pr(3) = 0.1$, and $Pr(4) = 0.3$. Draw the graph of the distribution function of X.*

Exercise 2.63. *Let X have a uniform distribution on the interval $[0, 5]$ and define the random variable Y by $Y = 0$ if $X \leq 1$, $Y = 5$ if $X \geq 3$, and $Y = X$ otherwise. Compute the distribution function of Y.*

Exercise 2.64. *Suppose that the distribution function of a random variable X is defined by:*

$$F(x) = \begin{cases} 0 \text{ if } x \leq 0, \\ \frac{1}{9}x^2 \text{ if } 0 < x \leq 3, \\ 1 \text{ if } x > 3. \end{cases}$$

Compute the probability density function of X.

2.3.3 Bivariate Distributions

Imagine that an experiment involves two discrete random variables X and Y and that we are interested in the values of both random variables simultaneously. In such an instance, it may be useful to be able to talk about the joint probability function of the two random variables. Our goal is to produce a probability model that deals with more than one random variable at a time. To do this we first generalize the notion of a random variable.

Definition 2.10. *By an* ***n-dimensional random vector*** *we mean a $\mathbb{R}^n$- valued function on the outcome space S, i.e. $X : S \to \mathbb{R}^n$ is an n-dimensional random vector on S.*

Example 2.65. *Let us again consider the experiment in which we toss two fair dice. Remember that the outcome space is $S := \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$. In previous exercises, you have been asked to work with two different random variables defined on S: (1) the sum of the two dice and (2) the absolute difference of the two dice. Formally, the first of these random variables is defined by $X(i, j) := i + j$, while the second is defined by*

$Y(i,j) := |i-j|$. *If we combine these two random variables we get the bivariate random vector (X,Y) defined by*

$$(X,Y)(i,j) := (X(i,j), Y(i,j)).$$

Definition 2.11. *We say that a random vector has* ***discrete distribution*** *or is a* ***discrete random vector*** *if its range is finite or countably infinite. If a random vector is discrete we define its probability distribution by the means of* ***joint probability mass function*** *(or joint pmf)*

$$f(x,y) = Pr(X = x, Y = y).$$

Example 2.66. *Consider the bivariate random vector defined in example 2.65. Let us compute $f(2,x)$, which is interpreted as the probability that the sum of the tow dice is 2 and the absolute difference is 0. Obviously, the only outcome $(i,j) \in S$ that satisfies $i+j=2$ and $|i-j|=0$ is $(1,1)$. Hence, $f(2,0)=\frac{1}{36}$. Next let us compute $f(3,0)$. As there does not exist $(i,j \in S)$ such that $i+j=3$ and $|i-j|=0$, we have $f(3,0)=0$.*

As in the case of a probability mass function, the joint probability mass function can be used to compute the probability of any event associated with the bivariate random vector (X,Y). Let $A \subseteq \mathbb{R}^2$. Then,

$$Pr((X,Y) \in A) = \sum_{(x,y)\in A} f(x,y).$$

Example 2.67. *For example,*

$$\begin{aligned} Pr((X,Y) \in [0,3]\times[0,2]) &= f(2,0)+f(2,1)+f(2,2)+f(3,0)+f(3,1)+f(3,2) \\ &= \frac{1}{36}+0+0+0+\frac{1}{18}+0 = \frac{1}{12}. \end{aligned}$$

Exercise 2.68. *Suppose X and Y have a discrete joint distribution for which the joint pmf is defined by*

$$f(x,y) = \begin{cases} c|x+y| \text{ for } x,y = -2,-1,0,1,2 \\ 0 \;\; otw \end{cases}.$$

Determine (a) the value of the constant c, (b) $Pr(X=0, Y=-2)$, (c) $Pr(X=1)$, (d) $Pr(|X-Y| \le 1)$.

We can similarly define a joint probability for continuous random variables.

Definition 2.12. *Two random variables X and Y have a* ***continuous joint distribution*** *if there exists a function $f : \mathbb{R}^2 \to \mathbb{R}$ that satisfies the following properties:*

(a) $f(x,y) \ge 0$ for all $(x,y) \in \mathbb{R}^2$,

(b) f is bounded and Riemann integrable,

(c) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$,

(d) $Pr[(X,Y) \in A] = \int_A \int f(x,y)dxdy$ *for all* $A \subseteq \mathbb{R}^2$.

The function f *is called the* ***joint probability density function*** *of* X *and* Y*.*

Example 2.69. *Consider the joint pdf defined by*

$$f(x,y) := \begin{cases} 6xy^2 \text{ if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 \text{ otw} \end{cases}.$$

Let $A := \{(x,y) : x + y \geq 1\}$*. We want to compute* $Pr((X,Y) \in A)$*. Note that the density is equal to* 0 *for any* (x,y) *not contained in the unit square. Hence, we only need to integrate over*

$$\begin{aligned} A \cap (0,1) \times (0,1) &= \{(x+y) : x + y \geq 1, 0 < x < 1, 0 < y < 1\} \\ &= \{(x,y) : x \geq 1 - y, 0 < x < 1, 0 < y < 1\} \\ &= \{(x,y) : 1 - y \leq x < 1, 0 < y < 1\} \end{aligned}$$

Everything is now set to compute the sought probability:

$$Pr(X + Y \geq 1) = \int_A \int f(x,y)dxdy = \int_0^1 \int_{1-y}^1 6xy^2 dxdy = \frac{9}{10}.$$

Exercise 2.70. *Suppose that* X *and* Y *have a continuous joint distribution for which the pdf is defined by*

$$f(x,y) = \begin{cases} cy^2 \text{ for } 0 \leq x \leq 2 \text{ and } 0 \leq y \leq 1, \\ 0 \text{ otw} \end{cases}.$$

Determine (a) the value of the constant c, (b) $Pr(X + Y > 2)$*, (c)* $Pr(Y < 1/2)$*, (d)* $Pr(X \leq 1)$*, (e)* $Pr(X = 3Y)$*.*

Finally, let us introduce the joint distribution function of two random variables X and Y.

Definition 2.13. *The function* $F : \mathbb{R}^2 \to [0,1]$ *defined by*

$$F(x,y) := Pr(X \leq x \text{ and } Y \leq y)$$

is called the ***joint distribution function*** *of the random variables* X *and* Y*.*

In the case of a discrete bivariate random vector the joint distribution function is often difficult to manipulate. Hence, we rather try to work with the joint probability mass function. In the continuous case however we have the following useful relationships:

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)dtds,$$

and

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y).$$

Exercise 2.71. *Suppose that X and Y are random variables such that (X,Y) must belong to $[0,3]\times[0,4]$. Suppose that the joint distribution function of (X,Y) is defined by*

$$F(x,y)=\frac{1}{156}xy(x^2+y)$$

for all $(x,y)\in[0,3]\times[0,4]$.

Determine (a) $Pr(1\le X\le 2, 1\le Y\le 2)$, (b) $Pr(2\le X\le 4, 2\le Y\le 4)$, (c) the joint pdf of (X,Y), and (e) $Pr(Y\le X)$.

2.3.4 Marginal Distributions

Suppose we know the joint probability distribution of two random variables. We will now study how to derive the distribution of one these two variables from the joint distribution.

So let X and Y have a discrete joint probability mass function f. We deduce the probability mass function of X, let's call it f_1, as follows

$$f_1(x) := Pr(X=x)=\sum_y Pr(X=x \text{ and } Y=y)=\sum_y f(x,y).$$

Obviously, the probability mass function of Y, which we call f_2 is defined similarly as

$$f_2(y) := Pr(Y=y)=\sum_x Pr(X=x \text{ and } Y=y)=\sum_x f(x,y).$$

Note that f_1 (f_2) is called the **marginal probability mass function** of X (Y).

Example 2.72. *Let us consider the random vector defined in example 2.65. We want to compute the marginal pmf of the absolute difference random variable. We get*

$$f_2(0)=f(2,0)+f(4,0)+f(6,0)+f(8,0)+f(10,0)+f(12,0)=\frac{1}{6}.$$

Similarly, $f_2(1)=\frac{5}{18}$, $f_2(2)=\frac{2}{9}$, $f_2(3)=\frac{1}{6}$, $f_2(4)=\frac{1}{9}$, $f_2(5)=\frac{1}{18}$.

Of course we can also deduce the **marginal probability density functions** from a joint probability density function. We then have

$$f_1(x)=\int_{-\infty}^{\infty} f(x,y)dy,$$

and

$$f_2(x)=\int_{-\infty}^{\infty} f(x,y)dx.$$

Example 2.73. *Consider the following joint pdf:*

$$f(x,y) := \begin{cases} \frac{21}{4}x^2y \textit{ for } x^2\le y\le 1, \\ 0 \textit{ otw.} \end{cases}$$

It should be obvious that X can only take values in the interval $[-1,1]$. Hence, $f_1(x) = 0$ whenever $x < -1$ or $x > 1$. Moreover, even if $x \in [-1,1]$ $f(x,y) = 0$ if $x^2 > y$. It follows that for $-1 \le x \le 1$,

$$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{x^2}^{1} \left(\frac{21}{4}\right) x^2 y dy = \left(\frac{21}{8}\right) x^2(1-x^4).$$

Similarly, Y can only take values in $[0,1]$. Hence, $f_2(y) = 0$ if $y < 0$ or $y > 1$. Moreover, $f(x,y) = 0$ unless $-\sqrt{y} \le x \le \sqrt{y}$. It follows that for $0 \le y \le 1$,

$$f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\sqrt{y}}^{\sqrt{y}} \left(\frac{21}{4}\right) x^2 y dx = \left(\frac{7}{2}\right) y^{5/2}.$$

Exercise 2.74. *Suppose that X and Y have a joint pdf defined by:*

$$f(x,y) = \begin{cases} k \text{ if } a \le x \le b, c \le y \le d \\ 0 \text{ otw.,} \end{cases}$$

where $a < b$, $c < d$, $k > 0$. Find the marginal pdfs of X and Y.

2.3.5 Conditional Distributions

The notion of conditional probability can be extend to random variables.

Definition 2.14. *As before let f denote the joint pmf (pdf) of two random variables X and Y and let f_1 and f_2 be their respective marginal pmfs (pdfs). Then, the **conditional probability mass (density) function** of X given Y is defined as*

$$\begin{aligned} Pr(X = x|Y = y) &:= \frac{Pr(X = x \text{ and } Y = y)}{Pr(Y = y)} \\ &= \frac{f(x,y)}{f_2(y)}. \end{aligned}$$

Example 2.75. *As in example 2.73, let our joint pdf be defined by:*

$$f(x,y) := \begin{cases} \frac{21}{4} x^2 y \text{ for } x^2 \le y \le 1, \\ 0 \text{ otw.} \end{cases}$$

In example 2.73 we showed that

$$f_1(x) = \begin{cases} \left(\frac{21}{8}\right) x^2(1-x^4) \text{ if } -1 \le x \le x \\ 0 \text{ otw.} \end{cases}$$

It follows that the conditional pdf of Y given X, denoted $g_2(y|x)$ is given by:

$$g_2(y|x) = \begin{cases} \frac{2y}{1-x^4} \text{ if } x^2 \le y \le 1 \\ 0 \text{ otw.} \end{cases}$$

Exercise 2.76. *Suppose that the joint pdf of X and Y is as follows:*

$$f(x,y) = \begin{cases} \frac{3}{16}(4-2x-y) \text{ for } x > 0, y > 0, \text{ and } 2x + y < 4, \\ 0 \text{ otw.} \end{cases}$$

Determine (a) the conditional pdf of Y given X, and (b) $Pr(Y \geq 2|X = .5)$.

2.3.6 Independent Random Variables

Definition 2.15. *Two random variables X and Y are said to be **independent** if for any two subsets A and B of $\mathbb{R}$, we have*

$$Pr(X \in A \text{ and } Y \in B) = Pr(X \in A)Pr(Y \in B).$$

The notion of independence of random variables can also be expressed in terms of their marginal and joint distributions (or distribution functions). Indeed the following is true:

Proposition 2.77. *Two random variables X and Y are independent iff for all $(x,y) \in \mathbb{R}^2$,*

$$f(x,y) = f_1(x)f_2(y).$$

Similarly, if we let F_1 and F_2 denote the distribution functions of X and Y respectively, then X and Y are independent iff for all $(x,y) \in \mathbb{R}^2$,

$$F(x,y) = F_1(x)F_2(y).$$

Example 2.78. *Let the discrete random vector (X,Y) have the following joint pmf: $f(10,1) = f(20,1) = f(20,2) = \frac{1}{10}$, $f(10,2) = f(10,3) = \frac{1}{5}$, and $f(20,3) = \frac{3}{10}$.*

Some easy computations give us the marginal pmfs:
$f_1(10) = f_2(10) = \frac{1}{2}$ and $f_2(1) = \frac{1}{5}$, $f_2(2) = \frac{3}{10}$, and $f_2(3) = \frac{1}{2}$.

Now observe that

$$f(10,3) = \frac{1}{5} \neq \frac{1}{2}\frac{1}{2} = f_1(10)f_2(3).$$

It follows that X and Y are not independent.

Exercise 2.79. *Suppose that X and Y have a joint pmf defined by:*

$$f(x,y) = \begin{cases} \frac{1}{30}(x+y) \text{ for } x = 0,1,2, \text{ and } y = 0,1,2,3, \\ 0 \text{ otw.} \end{cases}$$

(a) Determine the marginal pmfs of X and Y.

(b) Are X and Y independent?

Exercise 2.80. *Suppose that X and Y have joint pdf defined by*

$$f(x,y) = \begin{cases} \frac{3}{2}y^2 \text{ for } 0 \leq x \leq 2, \text{ and } 0 \leq y \leq 1, 0 \text{ otw.} \end{cases}$$

(a) Determine the marginal pdfs of X and Y.

(b) Are X and Y independent?

(c) Are the event $\{X < 1\}$ and the event $\{Y \geq 1/2\}$ independent?

2.3.7 Multivariate Distributions

We now extend the notion of the bivariate joint distribution to more than two random variables.

Definition 2.16. *The **joint probability mass function** of the random variables $X_1, \ldots, X_n$ is the function f that assigns to any vector $(x_1, \ldots, x_n) \in \mathbb{R}^n$ the value $f(x_1, \ldots, x_n)$ defined by*

$$f(x_1, \ldots, x_n) := Pr(X_1 = x_1, \ldots, X_n = x_n).$$

Definition 2.17. *The random variables $X_1, \ldots, X_n$ have a **joint probability density function** if there exists a function $f : \mathbb{R}^n \to \mathbb{R}$, with $f(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$, such that for every $A \subset \mathbb{R}^n$,*

$$Pr\left[(X_1, \ldots, X_n) \in A\right] = \int_A \cdots \int f(x_1, \ldots, x_n) dx_1 \cdots dx_n.$$

Definition 2.18. *Let $X_1, \ldots X_n$ be random variables. The function $F : \mathbb{R}^n \to [0, 1]$ defined by*

$$F(x_1, \ldots, x_n) := Pr(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

*is called the **joint distribution function** of the random variables $X_1, \ldots X_n$.*

As with bivariate distributions you can extract the marginal distributions from the joint distribution. Let $X_1, \ldots, X_n$ have a continuous joint distribution f then the marginal distribution f_1 of X_1 can be computed in the following manner

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_2 \cdots dx_n.$$

Note that we integrate over all variables except X_1.

We can also extract the joint marginal distribution of k of the n random variables by integrating the joint distribution over all the $n - k$ other variables.

We now extend the notion of independence to several random variables.

Definition 2.19. *We say that the random variables $X_1, \ldots, X_n$ are **independent** if, for any n sets of real numbers $A_1, \ldots, A_n$,*

$$Pr(X_1 \in A_1, \ldots, X_n \in A_n) = Pr(X_1 \in A_1) \cdots Pr(X_n \in A_n).$$

Proposition 2.81. *Let $X_1, \ldots, X_n$ be random variables with joint distribution F and let F_i be the distribution function of random variable X_i. Then, $X_1, \ldots, X_n$ are independent iff*

$$F(x_1, \ldots, x_n) = F_1(x_1) \cdots F_n(x_n).$$

Finally, we generalize the notion of conditional distribution to the multivariate case.

Let $X_1, \ldots, X_n$ have a continuous joint distribution with probability density function f. Let f_0 denote the joint marginal distribution of the $n-1$ variables $X_2, \ldots, X_n$. Then for all values $x_2, \ldots, x_n$ such that $f_0(x_2, \ldots, x_n) > 0$ the conditional probability density function of X_1 given $X_2 = x_2, \ldots, X_n = x_n$ is defined by

$$g_1(x_1|x_2, \ldots, x_n) = \frac{f(x_1, \ldots, x_n)}{f_0(x_2, \ldots, x_n)}.$$

Exercise 2.82. *Suppose that the random variables X_1, X_2, and X_3 have a joint pdf defined by:*

$$f(x_1, x_2, x_3) = \begin{cases} c(x_1 + x_2 + x_3) \text{ for } 0 \le x_i \le 1 \quad (i = 1, 2, 3), \\ 0 \text{ otw.} \end{cases}$$

Determine (a) the value of the constant c, (b) the marginal joint pdf of X_1 and X_3, and (c) $Pr(X_3 < \frac{1}{2}|X_1 = \frac{1}{4}, X_2 = \frac{3}{4})$.

Exercise 2.83. *Suppose that the random variables $X_1, \ldots, X_n$ are independent and all have the same marginal pdf f. Determine the probability that at least k of these n random variables will lie in a specified interval $[a, c]$.*

2.4 Moments

Although the distribution of a random variable X contains all the probabilistic information about this variable, it is at times cumbersome to work with probability distributions and quickly get the information we're actually interested in. We thus now consider ways to summarize some of the important information of a random variable.

2.4.1 The Expectation of a Random Variable

Definition 2.20. *Let X be a random variable. If X has a discrete distribution f, then its **expectation** is defined by*

$$\mathbb{E}(X) := \sum_{x \in X(S)} x f(x).$$

*If X has a continuous distribution for which the probability density function is f, then its **expectation** is defined by*

$$\mathbb{E}(X) := \int_{-\infty}^{\infty} x f(x) dx.$$

Note that the expectation $\mathbb{E}(X)$ of a random variable X is also called the **expected value** or the **mean** of the random variable.

Here is a way to get an intuition about what the expectation tells us about our random variable.

We can also compute the expectation of a function of a random variable. Let X, and Y be random variables such that Y can be defined as a function of X, i.e. $Y = r(X)$. We have now two ways to compute the expectation of Y. First, we could determine the distribution of Y, and then simply apply the definition of the expected value. The second method does not require us to compute the distribution of Y, indeed we have the following:

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \sum_{r(x) \in Y(S)} r(x)f(x)$$

if X has a discrete distribution, and

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int_{-\infty}^{\infty} r(x)f(x)dx$$

if X has a continuous distribution.

Finally, here are some important properties of the expected value that you should know.

Proposition 2.84. *Let X be a random variable for which $\mathbb{E}(X)$ exists. Then*

(a) If a and b are any real numbers, then

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$

(b) If there exists a constant a such that $Pr(X \geq a) = 1$, then $\mathbb{E}(x) \geq a$. If there exists a constant b such that $Pr(X \leq b) = 1$, then $\mathbb{E}(x) \leq b$.

(c) If $X_1, \dots, X_n$ are n random variables such that $\mathbb{E}(X_i)$ exists for all $i \in \{1, \dots, n\}$, then

$$\mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots \mathbb{E}(X_n).$$

(d) Let $X_1, \dots, X_n$ be n independent random variables such that $\mathbb{E}(X_i)$ exists for all $i \in \{1, \dots, n\}$, then

$$\mathbb{E}(\prod_{i=1}^{n} X_i) = \prod_{i=1}^{n} \mathbb{E}(X_i).$$

2.4.2 Variance

Definition 2.21. *Let X be a random variable with expectation $\mu := \mathbb{E}(X)$. Then, the **variance** of X, denoted by $Var(X)$, is defined by:*

$$Var(X) = \mathbb{E}\left[(X - \mu)^2\right].$$

interpretation

Proposition 2.85. *(a) For all random variables X, we have $Var(X) \geq 0$.*

(b) $Var(X) = 0$ if and only if there exists a real number a such that $Pr(X = a) = 1$.

(c) For any two real numbers a, and b,

$$Var(aX+b) = a^2 Var(X).$$

(d) For any random variable X,

$$Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

(e) If $X_1, \ldots, X_n$ are independent random variables, then

$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n).$$

2.4.3 Moments

Definition 2.22. *For any random variable X and any $k \in \mathbb{N}$, the expectation $\mathbb{E}(X^k)$ is called the* ***k-th moment*** *of X. It follows in particular that the expectation of X is the first moment of X.*

Definition 2.23. *Let X be a random variable with expectation $\mu := \mathbb{E}(X)$. Then for any $k \in \mathbb{N}$, the expectation $\mathbb{E}\left[X - \mu)^k\right]$ is called the* ***k-th central moment*** *of X. Note that the variance is the second central moment of X.*

2.5 The Normal Distribution

We now have a quick look at the normal distribution which is certainly the most important continuous distribution.

Definition 2.24. *A continuous random variable X has a* ***normal distribution*** *with expected value μ and variance σ^2 if its probability density function is defined by*

$$f(x|\mu, \sigma^2) := \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}.$$

The normal distribution with expected value 0 and variance 1 is called the ***standard normal distribution****.*

3 Linear Algebra

3.1 Matrix Algebra

Definition 3.1. *By an $m \times n$* ***matrix*** *we mean a function from $\{1, \ldots, m\} \times \{1, \ldots, n\}$ into $\mathbb{R}$. The matrix is however usually presented as an array*

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

where a_{ij} is the real number that is assigned to the ordered pair $(i,j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}$. A compact way to describe the matrix is simply to write $A = (a_{ij})$.

Note that there are even more ways to represent a matrix. Let

$$\begin{bmatrix} a_{i1} & \dots & a_{in} \end{bmatrix}$$

be the i-th row of the matrix A, which we denote A_i^r, $i = 1, \dots, m$, and let

$$\begin{bmatrix} a_{ij} \\ \vdots \\ a_{mj} \end{bmatrix}$$

be the j-th column vector of A, which we denote A_j^c, $j = 1, \dots, n$, then the matrix A can be represented as

$$A = \begin{bmatrix} A_1^r \\ \vdots \\ A_m^r \end{bmatrix} = \begin{bmatrix} A_1^c \dots A_n^c \end{bmatrix}.$$

Definition 3.2. *Let A and B be two $m \times n$ matrices then their sum $A + B$ is the $m \times n$ matrix whose $(i, j)-th$ entry is $a_{ij} + b_{ij}$:*

$$A + B = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m_1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix}.$$

Remark 3.1. *An important point to understand is that the sum of two matrices is only defined if they have the same number of rows and the same number of columns.*

Definition 3.3. *Let x and y be two vectors in $\mathbb{R}^n$. Then the inner product of x and y, denoted $x \cdot y$, is defined by:*

$$x \cdot y = \sum_{i=1}^{n} x_i y_i.$$

Definition 3.4. *If A is an $m \times n$ matrix and B is an $n \times k$ matrix then their product AB is the $m \times k$ matrix whose (i, j)-th entry is the inner product of the i-th row A_i^r of A and the j-th column B_j^c of B:*

$$AB = \begin{bmatrix} A_1^r \cdot B_1^c & A_1^r \cdot B_2^c & \dots & A_1^r \cdot B_k^c \\ A_2^r \cdot B_1^c & A_2^r \cdot B_2^c & \dots & A_2^r \cdot B_k^c \\ \vdots & \vdots & \ddots & \vdots \\ A_m^r \cdot B_1^c & A_m^r \cdot B_2^c & \dots & A_m^r \cdot B_k^c \end{bmatrix}$$

Remark 3.2. *(a) Note that the product AB is only well-defined if the number of columns of A is the same as the number of rows of B. Otherwise, the inner product $A_i^r \cdot B_j^c$ would not be defined.*

(b) It follows from (a) that it is not in general the case that $AB = BA$, as BA may not even exist.

Exercise 3.3. *For those of you who feel the need to perform some computations of additions and multiplications of matrices to make sure that you understand the concepts, I recommend that you do exercise 8.1. on page 159 in Simon and Blume.*

Exercise 3.4. *Let A and B be 2×2 matrices. Identify general conditions on the entries (a_{ij}) and (b_{ij}) under which $AB = BA$. Using these conditions find a numerical example of such matrices.*

Proposition 3.5. *(a) Commutativity of matrix addition: $A + B = B + A$.*

(b) Associativity of matrix addition: $(A + B) + C = A + (B + C)$.

Proposition 3.6. *(a) Associativity of matrix multiplication: $(AB)C = A(BC)$.*

(b) Multiplication distributes over addition: $A(B + C) = AB + AC$.

Exercise 3.7. *Which of the following statements are true, and why?*

(a) $(A + B)^2 = (B + A)^2$,

(b) $(A + B)^2 = A^2 + 2AB + B^2$,

(c) $(A + B)^2 = A(A + B) + B(A + B)$,

(d) $(A + B)^2 = (A + B)(B + A)$,

(e) $(A + B)^2 = A^2 + AB + BA + B^2$.

Definition 3.5. *The transpose of a matrix A, denoted A^T, is the matrix whose $(i, j)-th$ entry is a_{ji}. Hence, if A is an $m \times n$ matrix then A^T is an $n \times m$ matrix.*

Example 3.8. *Let* $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$. *Then* $A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$.

Proposition 3.9. *(Properties of transposition)*

(a) $(A + B)^T = A^T + B^T$

(b) $(AB)^T = B^T A^T \quad (ABC)^T = C^T B^T A^T$

(c) $(A^T)^T = A$

(d) $(rA)^T = rA^T$

Exercise 3.10. *Show that if AB is defined, then $B^T A^T$ is well defined but $A^T B^T$ need not.*

3.1.1 Some Important Classes of Matrices

Definition 3.6. *The $m \times n$ matrix A is said to be a* ***square matrix*** *if $m = n$, i.e. the number of rows of A is equal to its number of columns. The number of rows (columns) is called the* **order** *of the square matrix. The terms a_{ii} are called the* ***diagonal entries*** *of A, while the terms a_{ij} with $i \neq j$ are called the* ***off-diagonal entries****.*

Definition 3.7. *The square matrix A of order n is said to be* ***symmetric*** *if $A = A^T$, i.e. if for all $(i, j) \in \{1, \ldots, n\} \times \{1, \ldots, n\}$ we have $a_{ij} = a_{ji}$.*

Definition 3.8. *A square matrix A whose off-diagonal entries are all zero is called a* ***diagonal matrix****:*

$$D = \begin{bmatrix} d_{11} & 0 & \ldots & 0 \\ 0 & d_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{nn} \end{bmatrix}.$$

Note that any diagonal matrix is also symmetric.

Definition 3.9. *The* ***identity matrix*** *of order n is a square $n \times n$ matrix whose diagonal elements are all equal to 1 and whose off-diagonal elements are all zero:*

$$I = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix}.$$

As the identity matrix is a special kind of diagonal matrix it is also symmetric.

Proposition 3.11. *Let A be an $m \times n$ matrix and B a $n \times k$ matrix, then $AI = A$ and $IB = B$. Moreover, $I^2 := I \times I = I$.*

Definition 3.10. *A square matrix of order n which has the property that all the entries above the diagonal are zero is called a* ***lower-triangular matrix*** *of order n:*

$$D = \begin{bmatrix} d_{11} & 0 & \ldots & 0 \\ d_{21} & d_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \ldots & d_{nn} \end{bmatrix}.$$

Definition 3.11. *A square matrix of order n which has the property that all the entries above the diagonal are zero is called a* ***lower-triangular matrix*** *of order n:*

$$D = \begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1n} \\ 0 & d_{22} & \ldots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{nn} \end{bmatrix}.$$

Definition 3.12. *A square matrix A of order n is* ***idempotent*** *if $A^2 = A$. Idempotent matrices arise frequently in econometrics.*

Definition 3.13. *A square matrix A is called a* ***permutation matrix*** *if each row and each column contains a single entry 1, while all other entries are zero.*

Exercise 3.12. *Let $\mathbf{0}$ denote a matrix whose entries are all zero. If $A = \mathbf{0}$ and $B = \mathbf{0}$, is $A = B$?*

Exercise 3.13. *(a) Is the 3×3 matrix $A := \begin{bmatrix} 0 & 0 & a \\ 0 & b & 0 \\ c & 0 & 0 \end{bmatrix}$ a diagonal matrix?*

(b) Show that AA^T and A^TA are diagonal matrices.

Exercise 3.14. *Let A be a square matrix.*

(a) Show that $A + A^T$ is symmetric even if A is not.

(b) Show that AB is not necessarily symmetric even if A and B are.

(c) Show that A^TBA is symmetric if B is symmetric.

3.1.2 Rank of a Matrix

Exercise 3.15. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 5 & 6 \\ 1 & 6 & 8 \\ 7 & 1 & 8 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 5 & 6 & 3 \\ 2 & 1 & 4 & 1 \\ 3 & 5 & 5 & 4 \end{bmatrix}$$

(a) Compute the rank of A and B.

(b) Show that $\rho(A) = \rho(A^T) = \rho(AA^T) = \rho(A^TA)$

(c) Is it possible to construct a 3×4 matrix of rank 4?

Exercise 3.16. *Find the inverse of the following matrices:*

$$A = \begin{bmatrix} 6 & 13 & 4 \\ 8 & 15 & 2 \\ 7 & 14 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 15 & 24 & 12 \\ 10 & 16 & 10 \\ 2 & 3 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 8 & 3 \\ 11 & 4 & 0 \end{bmatrix}.$$

Exercise 3.17. *Consider the matrices*

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

(a) Does A have an inverse? What is the rank of A?

(b) Does B have an inverse? What is the rank of B?

Exercise 3.18. *Let A and B be $n \times n$ matrices. Show that $AB = I$ iff $BA = I$.*

Exercise 3.19. *Let* $A = \begin{bmatrix} 1 & \alpha & 0 \\ 4 & 5 & 3 \\ 1 & 0 & 2 \end{bmatrix}$

(a) For which α is $|A| = 0$. Show that the columns of A are linearly dependent in that case.

(b) If we interchange the second and third row, show that $|A|$ changes sign.

(c) If we multiply the first column by 2, show that $|A|$ is also multiplied by 2.

(d) If we substract 4 times the first row from the second row, show that $|A|$ does not change.

Exercise 3.20. *Compute the determinant of the following three matrices:*

$$A = \begin{bmatrix} -1 & 3 & 2 \\ 6 & -2 & 3 \\ 7 & 10 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 13 & 4 & 1 \\ 0 & 4 & 1 \\ -7 & 2 & 3 \end{bmatrix} \quad C = \begin{bmatrix} -4 & 4 & 12 \\ 3 & -3 & -9 \\ 8 & 2 & 6 \end{bmatrix}.$$

Exercise 3.21. *Let A be a diagonal matrix. Show that A is positive definite iff $a_{ii} > 0$ for all $i \in \{1, \ldots, n\}$*

Exercise 3.22. *Let A be a positive definite matrix. Show that the diagonal elements of A are positive.*

Exercise 3.23. *Prove the following statement or provide a counterexample to show it is false: if A is a positive definite matrix, then A^{-1} is a negative definite matrix.*

Exercise 3.24. *Give an example of matrices A and B which are each negative semidefinite but not negative definite, and which are such that $A + B$ is negative definite.*

Exercise 3.25. *Find the Hessians $D^2 f$ of each of the following functions. Evaluate the hessians at the specified points and examine if the hessian is positive definite, negative definite, positive semidefinite, negative semidefinite or indefinite.*

(a) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1^2 + \sqrt{x_2}$, at $x = (1, 1)$.

(b) $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x) = (x_1 x_2)^2$ at an arbitrary point $x \in \mathbb{R}^2_{++}$.

1. Matrix Algebra
   (a) nonsingular.
   (b) Elementary Matrices.
   (c) inverse, invertible, uniqueness of the inverse, right (left) inverse, inversion and solution to $A\mathbf{x} = \mathbf{b}$, nonsingular and invertible, $(A^{-1})^{-1}$, $(A^T)^{-1}$, $(AB)^{-1}$, A^m.

2. Determinants: minor, cofactor, determinant, computing the determinant, determinant and nonsingularity, A^{-1}, Cramer's rule, A square, then $\det A^T$, $\det(A \cdot B)$, $\det(A + B)$.

4 First Take on Optimization Theory

References