Dan A. Simovici
Chabane Djeraba

Mathematical Tools for Data Mining

Set Theory, Partial Orders, Combinatorics

Springer

Advanced Information and Knowledge Processing

For other titles published in this series, go to
www.springer.com/series/4738

Dan A. Simovici · Chabane Djeraba

Mathematical Tools for Data Mining

Set Theory, Partial Orders, Combinatorics

Springer

Dan A. Simovici, MS, MS, PhD
University of Massachusetts, Boston
USA

Chabane Djeraba, BSc, MSc, PhD
University of Sciences and Technologies
of Lille (USTL)
France

AI&KP ISSN 1610-3947
ISBN: 978-1-84800-200-5 e-ISBN: 978-1-84800-201-2
DOI: 10.1007/978-1-84800-201-2

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2008932365

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

Preface

This volume was born from the experience of the authors as researchers and educators, which suggests that many students of data mining are handicapped in their research by the lack of a formal, systematic education in its mathematics.

The data mining literature contains many excellent titles that address the needs of users with a variety of interests ranging from decision making to pattern investigation in biological data. However, these books do not deal with the mathematical tools that are currently needed by data mining researchers and doctoral students. We felt it timely to produce a book that integrates the mathematics of data mining with its applications. We emphasize that this book is about mathematical tools for data mining and *not* about data mining itself; despite this, a substantial amount of applications of mathematical concepts in data mining are presented. The book is intended as a reference for the working data miner.

In our opinion, three areas of mathematics are vital for data mining: *set theory*, including partially ordered sets and combinatorics; *linear algebra*, with its many applications in principal component analysis and neural networks; and *probability theory*, which plays a foundational role in statistics, machine learning and data mining.

This volume is dedicated to the study of set-theoretical foundations of data mining. Two further volumes are contemplated that will cover linear algebra and probability theory.

The first part of this book, dedicated to set theory, begins with a study of functions and relations. Applications of these fundamental concepts to such issues as equivalences and partitions are discussed. Also, we prepare the ground for the following volumes by discussing indicator functions, fields and σ-fields, and other concepts.

In this part, we have also included a précis of universal and linear algebra that covers the needs of subsequent chapters. This part concludes with a chapter on graphs and hypergraphs.

The second part is centered around partially ordered sets. We present algebraic structures closely related to partial orders, namely lattices, and Boolean algebras. We study basic issues about lattices, such as their dual roles as special partially ordered sets and algebraic structures, the theory of complete lattices and Galois connections, and their applications to the study of association rules. Special attention is paid to Boolean algebras which are of increasing interest for data mining because they allow the discovery of minimal sets of features necessary for explaining observations and the discovery of hidden patterns.

An introduction to topology and measure theory, which is essential for the study of various concepts of dimension and the recent preoccupations of data mining researchers with the applications of fractal theory to data mining, is also a component of this part.

A variety of applications in data mining are discussed, such as the notion of entropy, presented in a new algebraic framework related to partitions rather than random distributions, levelwise algorithms that generalize the Apriori technique, and generalized measures and their use in the study of frequent item sets. This part concludes with a chapter on rough sets.

The third part is focused on metric spaces. Metrics play an important role in clustering, classification, and certain data preprocessing techniques. We study a variety of concepts related to metrics, from dissimilarities to metrics, tree metrics, and ultrametrics. This chapter is followed by an application chapter dedicated to clustering that includes basic types of clustering algorithms, limitations of clustering, and techniques for evaluating cluster quality.

The fourth part focuses on combinatorics, an area of mathematics dedicated to the study of finite collections of objects that satisfy certain criteria. The main topics discussed are the inclusion-exclusion principle, combinatorics of partitions, counting problems related to collections of sets, and the Vapnik-Chervonenkis dimension of collections of sets.

Each chapter ends with suggestions for further reading. The book contains more than 400 exercises; they form an integral part of the material. Some of the exercises are in reality supplemental material. For these, we include solutions. The mathematics required for making the best use of our book is a typical three-semester sequence in calculus.

We would like to thank Catherine Brett and Frank Ganz from Springer-Verlag for their professionalism and helpfulness.

Boston and Villeneuve d'Ascq — *Dan A. Simovici*
January 2008 — *Chabane Djeraba*

Contents

Part III Metric Spaces

Part I

Set Theory

1

Sets, Relations, and Functions

1.1 Introduction

In this chapter, dedicated to set-theoretical bases of data mining, we assume that the reader is familiar with the notion of a set, membership of an element in a set, and elementary set theory. After a brief review of set-theoretical operations we discuss collections of sets, ordered pairs, and set products.

The Axiom of Choice, a basic principle used in many branches of mathematics, is discussed in Section 1.4. This subject is approached again in the context of partially ordered sets in Chapter 4. Countable and uncountable sets are presented in Section 1.5. An introductory section on elementary combinatorics is expanded in Chapter 14. Finally, we introduce the basics of the relational database model.

1.2 Sets and Collections

If x is a member of a set S, this is denoted, as usual, by $x \in S$. To denote that x is not a member of the set S, we write $x \notin S$.

Throughout this book, we use standardized notations for certain important sets of numbers:

$\mathbb{C}$	the set of complex numbers
$\mathbb{R}$	the set of real numbers
$\mathbb{R}_{\geq 0}$	the set of nonnegative real numbers
$\mathbb{R}_{>0}$	the set of positive real numbers
$\hat{\mathbb{R}}_{\geq 0}$	the set $\mathbb{R}_{\geq 0} \cup \{+\infty\}$
$\hat{\mathbb{R}}$	the set $\mathbb{R} \cup \{-\infty, +\infty\}$
$\mathbb{Q}$	the set of rational numbers
$\mathbb{I}$	the set of irrational numbers
$\mathbb{Z}$	the set of integers
$\mathbb{N}$	the set of natural numbers
$\mathbb{N}_1$	the set of positive natural numbers

D.A. Simovici, C. Djeraba, *Mathematical Tools for Data Mining,*
DOI: 10.1007/978-1-84800-201-2_1,

The usual order of real numbers is extended to the set $\hat{\mathbb{R}}$ by $-\infty < x < +\infty$ for every $x \in \mathbb{R}$. In addition, we assume that

$$x + \infty = \infty + x = +\infty,$$
$$x - \infty = -\infty + x = -\infty,$$

for every $x \in \mathbb{R}$. Also,

$$x \cdot \infty = \infty \cdot x = \begin{cases} +\infty & \text{if } x > 0 \\ -\infty & \text{if } x < 0, \end{cases}$$

and

$$x \cdot (-\infty) = (-\infty) \cdot x = \begin{cases} -\infty & \text{if } x > 0 \\ \infty & \text{if } x < 0. \end{cases}$$

Note that the product of 0 with either $+\infty$ or $-\infty$ is not defined. Division is extended by $x/+\infty = x/-\infty = 0$ for every $x \in \mathbb{R}$.

If S is a finite set, we denote by $|S|$ the number of elements of S.

Sets may contain other sets as elements. For example, the set

$$\mathcal{C} = \{\emptyset, \{0\}, \{0,1\}, \{0,2\}, \{1,2,3\}\}$$

contains the empty set $\emptyset$ and $\{0\}, \{0,1\}$,$\{0,2\}$,$\{1,2,3\}$ as its elements. We refer to such sets as *collections of sets* or simply *collections*. In general, we use calligraphic letters $\mathcal{C}, \mathcal{D}, \ldots$ to denote collections of sets.

If $\mathcal{C}$ and $\mathcal{D}$ are two collections, we say that $\mathcal{C}$ is *included* in $\mathcal{D}$, or that $\mathcal{C}$ is a *subcollection* of $\mathcal{D}$, if every member of $\mathcal{C}$ is a member of $\mathcal{D}$. This is denoted by $\mathcal{C} \subseteq \mathcal{D}$.

Two collections $\mathcal{C}$ and $\mathcal{D}$ are equal if we have both $\mathcal{C} \subseteq \mathcal{D}$ and $\mathcal{D} \subseteq \mathcal{C}$. This is denoted by $\mathcal{C} = \mathcal{D}$.

Definition 1.1. *Let $\mathcal{C}$ be a collection of sets. The* union of $\mathcal{C}$*, denoted by $\bigcup \mathcal{C}$, is the set defined by*

$$\bigcup \mathcal{C} = \{x \mid x \in S \text{ for some } S \in \mathcal{C}\}.$$

If $\mathcal{C}$ is a nonempty *collection, its* intersection *is the set $\bigcap \mathcal{C}$ given by*

$$\bigcap \mathcal{C} = \{x \mid x \in S \text{ for every } S \in \mathcal{C}\}.$$

If $\mathcal{C} = \{S, T\}$, we have $x \in \bigcup \mathcal{C}$ if and only if $x \in S$ or $x \in T$ and $x \in \bigcup \mathcal{C}$ if and only if $x \in S$ and $y \in T$. The union and the intersection of this two-set collection are denoted by $S \cup T$ and $S \cap T$ and are referred to as the union and the intersection of S and T, respectively.

We give, without proof, several properties of union and intersection of sets:

1. $S \cup (T \cup U) = (S \cup T) \cup U$ (*associativity of union*),
2. $S \cup T = T \cup S$ (*commutativity of union*),
3. $S \cup S = S$ (*idempotency of union*),
4. $S \cup \emptyset = S$,
5. $S \cap (T \cap U) = (S \cap T) \cap U$ (*associativity of intersection*),
6. $S \cap T = T \cap S$ (*commutativity of intersection*),
7. $S \cap S = S$ (*idempotency of intersection*),
8. $S \cap \emptyset = \emptyset$,

for all sets S, T, U.

The associativity of union and intersection allows us to denote unambiguously the union of three sets S, T, U by $S \cup T \cup U$ and the intersection of three sets S, T, U by $S \cup T \cup U$.

Definition 1.2. *The sets S and T are* disjoint *if $S \cap T = \emptyset$.*

A collection of sets $\mathcal{C}$ is said to be a collection of pairwise disjoint sets *if for every S and T in $\mathcal{C}$, if $S \neq T$, S and T are disjoint.*

Definition 1.3. *Let S and T be two sets. The* difference *of S and T is the set $S - T$ defined by*

$$S - T = \{x \in S \mid x \notin T\}.$$

When the set S is understood from the context, we write $\overline{T}$ for $S - T$, and we refer to the set $\overline{T}$ as the *complement* of T with respect to S or simply the *complement of T*.

The relationship between set difference and set union and intersection is given in the following theorem.

Theorem 1.4. *For every set S and nonempty collection $\mathcal{C}$ of sets, we have*

$$S - \bigcup \mathcal{C} = \bigcap \{S - C \mid C \in \mathcal{C}\},$$
$$S - \bigcap \mathcal{C} = \bigcup \{S - C \mid C \in \mathcal{C}\}.$$

Proof. We leave the proof of these equalities to the reader. □

Corollary 1.5. *For any sets S, T, U, we have*

$$S - (T \cup U) = (S - T) \cap (S - U),$$
$$S - (T \cap U) = (S - T) \cup (S - U).$$

Proof. The corollary follows immediately from Theorem 1.4 by choosing $\mathcal{C} = \{T, U\}$. □

With the notation previously introduced for the complement of a set, the equalities of Corollary 1.5 become

$$\overline{T \cup U} = \overline{T} \cap \overline{U},$$
$$\overline{T \cap U} = \overline{T} \cup \overline{U}.$$

The link between union and intersection is given by the distributivity properties contained in the following theorem.

Theorem 1.6. *For any collection of sets* $\mathcal{C}$ *and set* T*, we have*

$$\left(\bigcup \mathcal{C}\right) \cap T = \bigcup \{C \cap T \mid C \in \mathcal{C}\}.$$

If $\mathcal{C}$ *is nonempty, we also have*

$$\left(\bigcap \mathcal{C}\right) \cup T = \bigcap \{C \cup T \mid C \in \mathcal{C}\}.$$

Proof. We shall prove only the first equality; the proof of the second one is left as an exercise for the reader.

Let $x \in (\bigcup \mathcal{C}) \cap T$. This means that $x \in \bigcup \mathcal{C}$ and $x \in T$. There is a set $C \in \mathcal{C}$ such that $x \in C$; hence, $x \in C \cap T$, which implies $x \in \bigcup \{C \cap T \mid C \in \mathcal{C}\}$.

Conversely, if $x \in \bigcup \{C \cap T \mid C \in \mathcal{C}\}$, there exists a member $C \cap T$ of this collection such that $x \in C \cap T$, so $x \in C$ and $x \in T$. It follows that $x \in \bigcup \mathcal{C}$, and this, in turn, gives $x \in (\bigcup \mathcal{C}) \cap T$. □

Corollary 1.7. *For any sets* T*,* U*,* V*, we have*

$$(U \cup V) \cap T = (U \cap T) \cup (V \cap T),$$
$$(U \cap V) \cup T = (U \cup T) \cap (V \cup T).$$

Proof. The corollary follows immediately by choosing $\mathcal{C} = \{U, V\}$ in Theorem 1.6. □

Note that if $\mathcal{C}$ and $\mathcal{D}$ are two collections such that $\mathcal{C} \subseteq \mathcal{D}$, then

$$\bigcup \mathcal{C} \subseteq \bigcup \mathcal{D}$$

and

$$\bigcap \mathcal{D} \subseteq \bigcap \mathcal{C}.$$

We initially excluded the empty collection from the definition of the intersection of a collection. However, within the framework of collections of subsets of a given set S, we will extend the previous definition by taking $\bigcap \emptyset = S$ for the empty collection of subsets of S. This is consistent with the fact that $\emptyset \subseteq \mathcal{C}$ implies $\bigcap \mathcal{C} \subseteq S$.

The *symmetric difference* of sets denoted by $\oplus$ is defined by

$$U \oplus V = (U - V) \cup (V - U)$$

for all sets U, V.

Theorem 1.8. *For all sets U, V, T, we have*
(i) $U \oplus U = \emptyset$;
(ii) $U \oplus V = V \oplus T$;
(iii) $(U \oplus V) \oplus T = U \oplus (V \oplus T)$.

Proof. The first two parts of the theorem are direct applications of the definition of $\oplus$. We leave to the reader the proof of the third part (the associativity of $\oplus$).

The next theorem allows us to introduce a type of set collection of fundamental importance.

Theorem 1.9. *Let $\{\{x, y\}, \{x\}\}$ and $\{\{u, v\}, \{u\}\}$ be two collections such that $\{\{x, y\}, \{x\}\} = \{\{u, v\}, \{u\}\}$. Then, we have $x = u$ and $y = v$.*

Proof. Suppose that $\{\{x, y\}, \{x\}\} = \{\{u, v\}, \{u\}\}$.

If $x = y$, the collection $\{\{x, y\}, \{x\}\}$ consists of a single set, $\{x\}$, so the collection $\{\{u, v\}, \{u\}\}$ will also consist of a single set. This means that $\{u, v\} = \{u\}$, which implies $u = v$. Therefore, $x = u$, which gives the desired conclusion because we also have $y = v$.

If $x \neq y$, then neither (x, y) nor (u, v) are singletons. However, they both contain exactly one singleton, namely $\{x\}$ and $\{u\}$, respectively, so $x = u$. They also contain the equal sets $\{x, y\}$ and $\{u, v\}$, which must be equal. Since $v \in \{x, y\}$ and $v \neq u = x$, we conclude that $v = y$. □

Definition 1.10. *An* ordered pair *is a collection of sets $\{\{x, y\}, \{x\}\}$.*

Theorem 1.9 implies that for an ordered pair $\{\{x, y\}, \{x\}\}$, x and y are uniquely determined. This justifies the following definition.

Definition 1.11. *Let $\{\{x, y\}, \{x\}\}$ be an ordered pair. Then x is* the first component of *p and y is* the second component of *p.*

From now on, an ordered pair $\{\{x, y\}, \{x\}\}$ will be denoted by (x, y). If both $x, y \in S$, we refer to (x, y) as an *ordered pair on the set S.*

Definition 1.12. *Let $\mathcal{C}$ and $\mathcal{D}$ be two collections of sets such that $\bigcup \mathcal{C} = \bigcup \mathcal{D}$. $\mathcal{D}$ is a* refinement *of $\mathcal{C}$ if, for every $D \in \mathcal{D}$, there exists $C \in \mathcal{C}$ such that $D \subseteq C$.*
This is denoted by $\mathcal{C} \sqsubseteq \mathcal{D}$.

Example 1.13. Consider the collection $\mathcal{C} = \{(a, \infty) \mid a \in \mathbb{R}\}$ and $\mathcal{D} = \{(a, b) \mid a, b \in \mathbb{R}, a < b\}$. It is clear that $\bigcup \mathcal{C} = \bigcup \mathcal{D} = \mathbb{R}$.

Since we have $(a, b) \subseteq (a, \infty)$ for every $a, b \in \mathbb{R}$ such that $a < b$, it follows that $\mathcal{D}$ is a refinement of $\mathcal{C}$.

Definition 1.14. *A collection of sets $\mathcal{C}$ is* hereditary *if $U \in \mathcal{C}$ and $W \subseteq U$ implies $W \in \mathcal{C}$.*

Example 1.15. Let S be a set. The collection of subsets of S, denoted by $\mathcal{P}(S)$, is a hereditary collection of sets since a subset of a subset T of S is itself a subset of S.

The set of subsets of S that contain k elements is denoted by $\mathcal{P}_k(S)$. Clearly, for every set S, we have $\mathcal{P}_0(S) = \{\emptyset\}$ because there is only one subset of S that contains 0 elements, namely the empty set. The set of all finite subsets of a set S is denoted by $\mathcal{P}_{fin}(S)$. It is clear that $\mathcal{P}_{fin}(S) = \bigcup k \in \mathbb{N}\mathcal{P}_k(S)$.

Example 1.16. If $S = \{a, b, c\}$, then $\mathcal{P}(S)$ consists of the following eight sets:

$$\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}.$$

For the empty set, we have $\mathcal{P}(\emptyset) = \{\emptyset\}$.

Definition 1.17. *A collection $\mathcal{C}$ has* finite character *if $C \in \mathcal{C}$ if and only if every finite subset of C belongs to $\mathcal{C}$.*

It is clear that, for a collection $\mathcal{C}$ of finite character, if $C \in \mathcal{C}$ and $D \subseteq C$, then we also have $D \in \mathcal{C}$. In other words, every collection of finite character is hereditary.

Theorem 1.18. *Let $\mathcal{C}$ be a collection of finite character that consists of subsets of a set S. If $U_0, \ldots, U_n, \ldots$ are members of $\mathcal{C}$ such that $U_0 \subseteq \cdots \subseteq U_n \subseteq \cdots$, then $U = \bigcup\{U_i \mid i \geq 0\} \in \mathcal{C}$.*

Proof. Let $W = \{w_i \mid 0 \leq i \leq n-1\}$ be a finite subset of U. For every $w_\ell \in W$, let w_ℓ be the least integer such that $w_\ell \in U_{q_\ell}$ for $0 \leq \ell \leq n-1$. If $q = \max\{q_0, \ldots, q_{k-1}\}$, then $W \subseteq U_q$, so $W \in \mathcal{C}$. Since every finite subset of U belongs to $\mathcal{C}$, we obtain $U \in \mathcal{C}$. □

Definition 1.19. *Let $\mathcal{C}$ be a collection of sets and let K be a set. The* trace of the collection $\mathcal{C}$ on the set K *is the collection $\{C \cap K \mid C \in \mathcal{C}\}$.*

An alternative notation for $\mathcal{C}_K$ is $\mathcal{C} \restriction_K$, a notation that we shall use when the collection $\mathcal{C}$ is adorned by other subscripts.

We conclude this presentation of collections of sets with two more operations on collections of sets.

Definition 1.20. *Let $\mathcal{C}$ and $\mathcal{D}$ be two collections of sets. The collections $\mathcal{C} \vee \mathcal{D}$, $\mathcal{C} \wedge \mathcal{D}$, and $\mathcal{C} - \mathcal{D}$ are given by*

$$\begin{aligned}
\mathcal{C} \vee \mathcal{D} &= \{C \cup D \mid C \in \mathcal{C} \text{ and } D \in \mathcal{D}\},\\
\mathcal{C} \wedge \mathcal{D} &= \{C \cap D \mid C \in \mathcal{C} \text{ and } D \in \mathcal{D}\},\\
\mathcal{C} - \mathcal{D} &= \{C - D \mid C \in \mathcal{C} \text{ and } D \in \mathcal{D}\}.
\end{aligned}$$

Example 1.21. Let $\mathcal{C}$ and $\mathcal{D}$ be the collections of sets defined by

$$\mathcal{C} = \{\{x\}, \{y, z\}, \{x, y\}, \{x, y, z\}\},$$
$$\mathcal{D} = \{\{y\}, \{x, y\}, \{u, y, z\}\}.$$

We have

$$\mathcal{C} \vee \mathcal{D} = \{\{x, y\}, \{y, z\}, \{x, y, z\}, \{u, y, z\}, \{u, x, y, z\}\},$$
$$\mathcal{C} \wedge \mathcal{D} = \{\emptyset, \{x\}, \{y\}, \{x, y\}, \{y, z\}\},$$
$$\mathcal{C} - \mathcal{D} = \{\emptyset, \{x\}, \{z\}, \{x, z\}\},$$
$$\mathcal{D} - \mathcal{C} = \{\emptyset, \{u\}, \{x\}, \{y\}, \{u, z\}, \{u, y, z\}\}.$$

Unlike "$\cup$" and "$\cap$", the operations "$\vee$" and "$\wedge$" between collections of sets are not idempotent. Indeed, we have, for example,

$$\mathcal{D} \vee \mathcal{D} = \{\{y\}, \{x, y\}, \{u, y, z\}, \{u, x, y, z\}\} \neq \mathcal{D}.$$

The trace $\mathcal{C}_K$ of a collection $\mathcal{C}$ on K can be written as $\mathcal{C}_K = \mathcal{C} \wedge \{K\}$.

1.3 Relations and Functions

This section covers a number of topics that are derived from the notion of relation.

1.3.1 Cartesian Products of Sets

Definition 1.22. *Let X and Y be two sets. The* Cartesian product *of X and Y is the set $X \times Y$, which consists of all pairs (x, y) such that $x \in X$ and $y \in Y$.*

If either $X = \emptyset$ or $Y = \emptyset$, then $X \times Y = \emptyset$.

Example 1.23. Consider the sets $X = \{a, b, c\}$ and $Y = \{0, 1\}$. Their Cartesian product is the set:

$$X \times Y = \{(x, 0), (y, 0), (z, 0), (x, 1), (y, 1), (z, 1)\}.$$

Example 1.24. The Cartesian product $\mathbb{R} \times \mathbb{R}$ consists of all ordered pairs of real numbers (x, y). Geometrically, each such ordered pair corresponds to a point in a plane equipped with a system of coordinates. Namely, the pair $(u, v) \in \mathbb{R} \times \mathbb{R}$ is represented by the point P whose x-coordinate is u and y-coordinate is v (see Figure 1.1)

The Cartesian product is distributive over union, intersection, and difference of sets.

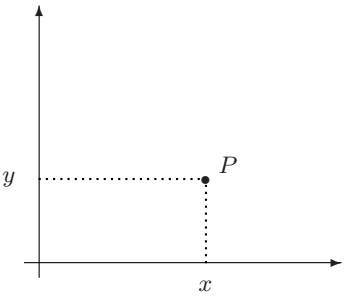

Fig. 1.1. Cartesian representation of the pair (x, y).

Theorem 1.25. *If $\star$ is one of $\cup, \cap$, or $-$, then for any sets R, S, and T, we have*

$$(R \star S) \times T = (R \times T) \star (S \times T),$$
$$T \times (R \star S) = (T \times R) \star (T \times S).$$

Proof. We prove only that $(R - S) \times T = (R \times T) - (S \times T)$. Let $(x, y) \in (R - S) \times T$. We have $x \in R - S$ and $y \in T$. Therefore, $(x, y) \in R \times T$ and $(x, y) \notin S \times T$, which show that $(x, y) \in (R \times T) - (S \times T)$.

Conversely, $(x, y) \in (R \times T) - (S \times T)$ implies $x \in$ and $y \in T$ and also $(x, y) \notin S \times T$. Thus, we have $x \notin S$, so $(x, y) \in (R - S) \times T$. □

It is not difficult to see that if $R \subseteq R'$ and $S \subseteq S'$, then $R \times S \subseteq R' \times S'$. We refer to this property as the *monotonicity of the Cartesian product with respect to set inclusion.*

1.3.2 Relations

Definition 1.26. *A* relation *is a set of ordered pairs.*

If S and T are sets and ρ is a relation such that $\rho \subseteq S \times T$, then we refer to ρ as a relation from S to T.

A relation from S to S is called a relation on S.

$\mathcal{P}(S \times T)$ is the set of all relations from S to T.

Among the relations from S to T, we distinguish the *empty relation* $\emptyset$ and the *full relation* $S \times T$.

The *identity relation* of a set S is the relation $\iota_S \subseteq S \times S$ defined by $\iota_S = \{(x, x) \mid x \in S\}$. The *full relation on S* is $\theta_S = S \times S$.

If $(x, y) \in \rho$, we sometimes denote this fact by $x \,\rho\, y$, and we write $x \not\rho\, y$ instead of $(x, y) \notin \rho$.

Example 1.27. Let $S \subseteq \mathbb{R}$. The relation "less than" on S is given by

$$\{(x, y) \mid x, y \in S \text{ and } y = x + z \text{ for some } z \in \mathbb{R}_{\geq 0}\}.$$

Example 1.28. Consider the relation $\nu \subseteq \mathbb{Z} \times \mathbb{Q}$ given by

$$\nu = \{(n, q) \mid n \in \mathbb{Z},\ q \in \mathbb{Q},\ \text{and } n \leq q < n + 1\}.$$

We have $(-3, -2.3) \in \nu$ and $(2, 2.3) \in \nu$. Clearly, $(n, q) \in \nu$ if and only if n is the integral part of the rational number q.

Example 1.29. The relation δ is defined by

$$\delta = \{(m, n) \in \mathbb{N} \times \mathbb{N} \mid n = km \text{ for some } k \in \mathbb{N}\}.$$

We have $(m, n) \in \delta$ if m divides n evenly.

Note that if $S \subseteq T$, then $\iota_S \subseteq \iota_T$ and $\theta_S \subseteq \theta_T$.

Definition 1.30. *The* domain *of a relation ρ from S to T is the set*

$$\mathrm{Dom}(\rho) = \{x \in S \mid (x, y) \in \rho \text{ for some } y \in T\}.$$

The range *of ρ from S to T is the set*

$$\mathrm{Ran}(\rho) = \{y \in T \mid (x, y) \in \rho \text{ for some } x \in S\}.$$

If ρ is a relation and S and T are sets, then ρ *is a relation from S to T* if and only if $\mathrm{Dom}(\rho) \subseteq S$ and $\mathrm{Ran}(\rho) \subseteq T$. Clearly, ρ is always a relation from $\mathrm{Dom}(\rho)$ to $\mathrm{Ran}(\rho)$.

If ρ and σ are relations and $\rho \subseteq \sigma$, then $\mathrm{Dom}(\rho) \subseteq \mathrm{Dom}(\sigma)$ and $\mathrm{Ran}(\rho) \subseteq \mathrm{Ran}(\sigma)$.

If ρ and σ are relations, then so are $\rho \cup \sigma$, $\rho \cap \sigma$, and $\rho - \sigma$, and in fact if ρ and σ are both relations from S to T, then these relations are also relations from S to T.

Definition 1.31. *Let ρ be a relation. The* inverse *of ρ is the relation ρ^{-1} given by*

$$\rho^{-1} = \{(y, x) \mid (x, y) \in \rho\}.$$

The proofs of the following simple properties are left to the reader:

(i) $\mathrm{Dom}(\rho^{-1}) = \mathrm{Ran}(\rho)$,
(ii) $\mathrm{Ran}(\rho^{-1}) = \mathrm{Dom}(\rho)$,
(iii) if ρ is a relation from A to B, then ρ^{-1} is a relation from B to A, and
(iv) $(\rho^{-1})^{-1} = \rho$

for every relation ρ. Furthermore, if ρ and σ are two relations such that $\rho \subseteq \sigma$, then $\rho^{-1} \subseteq \sigma^{-1}$ (monotonicity of the inverse).

Definition 1.32. *Let ρ and σ be relations. The* product *of ρ and σ is the relation $\rho\sigma$, where*

$$\rho\sigma = \{(x, z) \mid \text{ for some } y,\ (x, y) \in \rho,\ \text{and } (y, z) \in \sigma\}.$$

It is easy to see that $\mathrm{Dom}(\rho\sigma) \subseteq \mathrm{Dom}(\rho)$ and $\mathrm{Ran}(\rho\sigma) \subseteq \mathrm{Ran}(\sigma)$. Further, if ρ is a relation from A to B and σ is a relation from B to C, then $\rho\sigma$ is a relation from A to C.

Several properties of the relation product are given in the following theorem.

Theorem 1.33. *Let ρ_1, ρ_2, and ρ_3 be relations. We have*

(i) $\rho_1(\rho_2\rho_3) = (\rho_1\rho_2)\rho_3$ *(associativity of relation product).*

(ii) $\rho_1(\rho_2 \cup \rho_3) = (\rho_1\rho_2) \cup (\rho_1\rho_3)$ *and* $(\rho_1 \cup \rho_2)\rho_3 = (\rho_1\rho_3) \cup (\rho_2\rho_3)$ *(distributivity of relation product over union).*

(iii) $(\rho_1\rho_2)^{-1} = \rho_2^{-1}\rho_1^{-1}$.

(iv) *If $\rho_2 \subseteq \rho_3$, then $\rho_1\rho_2 \subseteq \rho_1\rho_3$ and $\rho_2\rho_1 \subseteq \rho_3\rho_1$ (monotonicity of relation product).*

(v) *If S and T are any sets, then $\iota_S\rho_1 \subseteq \rho_1$ and $\rho_1\iota_T \subseteq \rho_1$. Further, $\iota_S\rho_1 = \rho_1$ if and only if $\mathrm{Dom}(\rho_1) \subseteq S$, and $\rho_1\iota_T = \rho_1$ if and only if $\mathrm{Ran}(\rho_1) \subseteq T$. (Thus, ρ_1 is a relation from S to T if and only if $\iota_S\rho_1 = \rho_1 = \rho_1\iota_T$.)*

Proof. We prove (i), (ii), and (iv) and leave the other parts as exercises.

To prove Part (i), let $(a, d) \in \rho_1(\rho_2\rho_3)$. There is a b such that $(a, b) \in \rho_1$ and $(b, d) \in \rho_2\rho_3$. This means that there exists c such that $(b, c) \in \rho_2$ and $(c, d) \in \rho_3$. Therefore, we have $(a, c) \in \rho_1\rho_2$, which implies $(a, d) \in (\rho_1\rho_2)\rho_3$. This shows that $\rho_1(\rho_2\rho_3) \subseteq (\rho_1\rho_2)\rho_3$.

Conversely, let $(a, d) \in (\rho_1\rho_2)\rho_3$. There is a c such that $(a, c) \in \rho_1\rho_2$ and $(c, d) \in \rho_3$. This implies the existence of a b for which $(a, b) \in \rho_1$ and $(b, c) \in \rho_3$. For this b, we have $(b, d) \in \rho_2\rho_3$, which gives $(a, d) \in \rho_1(\rho_2\rho_3)$. We have proven the reverse inclusion, $(\rho_1\rho_2)\rho_3 \subseteq \rho_1(\rho_2\rho_3)$, which gives the associativity of relation product.

For Part (ii), let $(a, c) \in \rho_1(\rho_2 \cup \rho_3)$. Then, there is a b such that $(a, b) \in \rho_1$ and $(b, c) \in \rho_2$ or $(b, c) \in \rho_3$. In the first case, we have $(a, c) \in \rho_1\rho_2$; in the second, $(a, c) \in \rho_1\rho_3$. Therefore, we have $(a, c) \in (\rho_1\rho_2) \cup (\rho_1\rho_3)$ in either case, so $\rho_1(\rho_2 \cup \rho_3) \subseteq (\rho_1\rho_2) \cup (\rho_1\rho_3)$.

Let $(a, c) \in (\rho_1\rho_2) \cup (\rho_1\rho_3)$. We have either $(a, c) \in \rho_1\rho_2$ or $(a, c) \in \rho_1\rho_3$. In the first case, there is a b such that $(a, b) \in \rho_1$ and $(b, c) \in \rho_2 \subseteq \rho_2 \cup \rho_3$. Therefore, $(a, c) \in \rho_1(\rho_2 \cup \rho_3)$. The second case is handled similarly. This establishes

$$(\rho_1\rho_2) \cup (\rho_1\rho_3) \subseteq \rho_1(\rho_2 \cup \rho_3).$$

The other distributivity property has a similar argument.

Finally, for Part (iv), let ρ_2 and ρ_3 be such that $\rho_2 \subseteq \rho_3$. Since $\rho_2 \cup \rho_3 = \rho_3$, we obtain from (ii) that

$$\rho_1\rho_3 = (\rho_1\rho_2) \cup (\rho_1\rho_3),$$

which shows that $\rho_1\rho_2 \subseteq \rho_1\rho_3$. The second inclusion is proven similarly. □

Definition 1.34. *The n-power of a relation $\rho \subseteq S \times S$ is defined inductively by $\rho^0 = \iota_S$ and $\rho^{n+1} = \rho^n\rho$ for $n \in \mathbb{N}$.*

Note that $\rho^1 = \rho^0\rho = \iota_S\rho = \rho$ for any relation ρ.

Example 1.35. Let $\rho \subseteq \mathbb{R} \times \mathbb{R}$ be the relation defined by

$$\rho = \{(x, x+1) \mid x \in \mathbb{R}\}.$$

The zero-th power of ρ is the relation $\iota_{\mathbb{R}}$. The second power of ρ is

$$\rho^2 = \rho \cdot \rho = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid (x, z) \in \rho \text{ and } (z, y) \in \rho \text{ for some } z \in \mathbb{R}\}.$$

In other words, $\rho^2 = \{(x, x+2) \mid x \in \mathbb{R}\}$. In general, $\rho^n = \{(x, x+n) \mid x \in \mathbb{R}\}$.

Definition 1.36. *A relation ρ is a* function *if for all x, y, z, $(x, y) \in \rho$ and $(x, z) \in \rho$ imply $y = z$; ρ is a* one-to-one relation *if, for all x, x', and y, $(x, y) \in \rho$ and $(x', y) \in \rho$ imply $x = x'$.*

Observe that $\emptyset$ is a function (referred to in this context as the *empty function*) because $\emptyset$ satisfies vacuously the defining condition for being a function.

Example 1.37. Let S be a set. The relation ρ on $S \times \mathcal{P}(S)$ given by

$$\rho = \{(x, \{x\}) \mid x \in S\}$$

is a function.

Example 1.38. For every set S, the relation ι_S is both a function and a one-to-one relation. The relation ν from Example 1.28 is a one-to-one relation, but it is not a function.

Theorem 1.39. *For any relation ρ, ρ is a function if and only if ρ^{-1} is a one-to-one relation.*

Proof. Suppose that ρ is a function, and let $(y_1, x), (y_2, x) \in \rho^{-1}$. Definition 1.31 implies that $(x, y_1), (x, y_2) \in \rho$; hence, $y_1 = y_2$ because ρ is a function. This proves that ρ^{-1} is one-to-one.

Conversely, assume that ρ^{-1} is one-to-one and let $(x, y_1), (x, y_2) \in \rho$. Applying Definition 1.31, we obtain $(y_1, x), (y_2, x) \in \rho^{-1}$ and, since ρ^{-1} is one-to-one, we have $y_1 = y_2$. This shows that ρ is a function. □

Example 1.40. We observed that the relation ν introduced in Example 1.28 is one-to-one. Therefore, its inverse $\nu^{-1} \subseteq \mathbb{Q} \times \mathbb{Z}$ is a function. In fact, ν^{-1} associates to each rational number q its integer part $\lfloor q \rfloor$.

Definition 1.41. *A relation ρ from S to T is* total *if* $\mathrm{Dom}(\rho) = S$ *and is* onto *if* $\mathrm{Ran}(\rho) = T$.

Any relation ρ is a total and onto relation from $\mathrm{Dom}(\rho)$ to $\mathrm{Ran}(\rho)$. If both S and T are nonempty, then $S \times T$ is a total and onto relation from S to T.

It is easy to prove that a relation ρ from S to T is a total relation from S to T if and only if ρ^{-1} is an onto relation from T to S.

If ρ is a relation, then one can determine whether or not ρ is a function or is one-to-one just by looking at the ordered pairs of ρ. Whether ρ is a total or onto relation from A to B depends on what A and B are.

Theorem 1.42. *Let ρ and σ be relations.*
(i) *If ρ and σ are functions, then $\rho\sigma$ is also a function.*
(ii) *If ρ and σ are one-to-one relations, then $\rho\sigma$ is also a one-to-one relation.*
(iii) *If ρ is a total relation from R to S and σ is a total relation from S to T, then $\rho\sigma$ is a total relation from R to T.*
(iv) *If ρ is an onto relation from R to S and σ is an onto relation from S to T, then $\rho\sigma$ is an onto relation from R to T.*

Proof. To show Part (i), suppose that ρ and σ are both functions and that (x, z_1) and (x, z_2) both belong to $\rho\sigma$. Then, there exists a y_1 such that $(x, y_1) \in \rho$ and $(y_1, z_1) \in \sigma$, and there exists a y_2 such that $(x, y_2) \in \rho$ and $(y_2, z_2) \in \sigma$. Since ρ is a function, $y_1 = y_2$, and hence, since σ is a function, $z_1 = z_2$, as desired.

Part (ii) follows easily from Part (i). Suppose that relations ρ and σ are one-to-one (and hence that ρ^{-1} and σ^{-1} are both functions). To show that $\rho\sigma$ is one-to-one, it suffices to show that $(\rho\sigma)^{-1} = \sigma^{-1}\rho^{-1}$ is a function. This follows immediately from Part (i).

We leave the proofs for the last two parts of the theorem to the reader. □

The properties of relations defined next allow us to define important classes of relations.

Definition 1.43. *Let S be a set and let $\rho \subseteq S \times S$ be a relation. The relation ρ is:*
(i) reflexive *if $(s, s) \in \rho$ for every $s \in S$;*
(ii) irreflexive *if $(s, s) \notin \rho$ for every $s \in S$;*
(iii) symmetric *if $(s, s') \in \rho$ implies $(s', s) \in \rho$ for $s, s' \in S$;*
(iv) antisymmetric *if $(s, s'), (s', s) \in \rho$ implies $s = s'$ for $s, s' \in S$;*
(v) asymmetric *if $(s, s') \in \rho$ implies $(s', s) \notin \rho$; and*
(vi) transitive *if $(s, s'), (s', s'') \in \rho$ implies $(s, s'') \in \rho$.*

Example 1.44. The relation ι_S is reflexive, symmetric, antisymmetric, and transitive for any set S.

Example 1.45. The relation δ introduced in Example 1.29 is reflexive since $n \cdot 1 = n$ for any $n \in \mathbb{N}$.

Suppose that $(m, n), (n, m) \in \delta$. There are $p, q \in \mathbb{N}$ such that $mp = n$ and $nq = m$. If $n = 0$, then this also implies $m = 0$; hence, $m = n$. Let us assume

that $n \neq 0$. The previous equalities imply $nqp = n$, and since $n \neq 0$, we have $qp = 1$. In view of the fact that both p and q belong to $\mathbb{N}$, we have $p = q = 1$; hence, $m = n$, which proves the antisymmetry of ρ.

Let $(m, n), (n, r) \in \delta$. We can write $n = mp$ and $r = nq$ for some $p, q \in \mathbb{N}$, which gives $r = mpq$. This means that $(m, r) \in \delta$, which shows that δ is also transitive.

Definition 1.46. *Let S and T be two sets and let $\rho \subseteq S \times T$ be a relation.*

The image of an element $s \in S$ under the relation ρ *is the set* $\rho(s) = \{t \in T \mid (s, t) \in \rho\}$.

The preimage of an element $t \in T$ under ρ *is the set* $\{s \in S \mid (s, t) \in \rho\}$, *which equals* $\rho^{-1}(t)$, *using the previous notation.*

The collection of images of S under ρ *is*

$$\mathit{IM}_\rho = \{\rho(s) \mid s \in S\},$$

while the collection of preimages of T *is*

$$\mathit{PIM}_\rho = \mathit{IM}_{\rho^{-1}} = \{\rho^{-1}(t) \mid t \in T\}.$$

If $\mathcal{C}$ and $\mathcal{C}'$ are two collections of subsets of S and T, respectively, and $\mathcal{C}' = \mathit{IM}_\rho$ and $\mathcal{C} = \mathit{PIM}_\rho$ for some relation $\rho \subseteq S \times T$, we refer to $\mathcal{C}'$ as the dual class relative to ρ *of $\mathcal{C}$.*

Example 1.47. Any collection $\mathcal{D}$ of subsets of S can be regarded as the collection of images under a suitable relation. Indeed, let $\mathcal{C}$ be such a collection. Define the relation $\rho \subseteq S \times \mathcal{C}$ as $\rho = \{(s, C) \mid s \in S, C \in \mathcal{C} \text{ and } c \in C\}$. Then, IM_ρ consists of all subsets of $\mathcal{P}(\mathcal{C})$ of the form $\rho(s) = \{C \in \mathcal{C} \mid s \in C\}$ for $s \in S$. It is easy to see that $\mathsf{PIM}_\rho(\mathcal{C}) = \mathcal{C}$.

The collection IM_ρ defined in this example is referred to as the *bi-dual collection of* $\mathcal{C}$.

1.3.3 Functions

We saw that a function is a relation ρ such that, for every x in $\mathrm{Dom}(\rho)$, there is only one y such that $(x, y) \in \rho$. In other words, a function assigns a unique value to each member of its domain.

From now on, we will use the letters f, g, h, and k to denote functions, and we will denote the identity relation ι_S, which we have already remarked is a function, by 1_S.

If f is a function, then, for each x in $\mathrm{Dom}(f)$, we let $f(x)$ denote the unique y with $(x, y) \in f$, and we refer to $f(x)$ as the *image of x under f*.

Definition 1.48. *Let S and T be sets. A* partial function from S to T *is a relation from S to T that is a function.*

A total function from S to T *(also called a* function from S to T *or a* mapping from S to T*) is a partial function from S to T that is a total relation from S to T.*

The set of all partial functions from S to T is denoted by $S \rightsquigarrow T$ and the set of all total functions from S to T by $S \longrightarrow T$. We have $S \longrightarrow T \subseteq S \rightsquigarrow T$ for all sets S and T.

The fact that f is a partial function from S to T is indicated by writing $f : S \rightsquigarrow T$ rather than $f \in S \rightsquigarrow T$. Similarly, instead of writing $f \in S \longrightarrow T$, we use the notation $f : S \longrightarrow T$.

For any sets S and T, we have $\emptyset \in S \rightsquigarrow T$. If either S or T is empty, then $\emptyset$ is the only partial function from S to T. If $S = \emptyset$, then the empty function is a total function from S to any T. Thus, for any sets S and T, we have

$$\begin{aligned} S \rightsquigarrow \emptyset &= \{\emptyset\}, \\ \emptyset \rightsquigarrow T &= \{\emptyset\}, \\ \emptyset \longrightarrow T &= \{\emptyset\}. \end{aligned}$$

Furthermore, if S is nonempty, then there can be no (total) function from S to the empty set, so we have

$$S \longrightarrow \emptyset = \emptyset \text{ (if } S \neq \emptyset).$$

Definition 1.49. *A one-to-one function is called an* injection.

A function $f : S \rightsquigarrow T$ is called a surjection (from S to T) *if f is an onto relation from S to T, and it is called a* bijection (from S to T) *or a* one-to-one correspondence between S and T *if it is total, an injection, and a surjection.*

Using our notation for functions, we can restate the definition of injection as follows: f is an injection if for all $s, s' \in \mathrm{Dom}(f)$, $f(s) = f(s')$ implies $s = s'$. Likewise, $f : S \rightsquigarrow T$ is a surjection if for every $t \in T$ there is an $s \in S$ with $f(s) = t$.

Example 1.50. Let S and T be two sets and assume that $S \subseteq T$. The *containment mapping* $c : S \longrightarrow T$ defined by $c(s) = s$ for $s \in S$ is an injection. We denote such a containment by $c : S \hookrightarrow T$.

Example 1.51. Let $m \in \mathbb{N}$ be a natural number, $m \geq 2$. Consider the function $r_m : \mathbb{N} \longrightarrow \{0, \ldots, m-1\}$, where $r_m(n)$ is the remainder when n is divided by m. Obviously, r_m is well-defined since the remainder p when a natural number is divided by m satisfies $0 \leq p \leq m-1$. The function r_m is onto because of the fact that, for any $p \in \{0, \ldots, m-1\}$, we have $r_m(km + p) = p$ for any $k \in \mathbb{N}$.

For instance, if $m = 4$, we have $r_4(0) = r_4(4) = r_4(8) = \cdots = 0$, $r_4(1) = r_4(5) = r_4(9) = \cdots = 1$, $r_4(2) = r_4(6) = r_4(10) = \cdots = 2$ and $r_4(3) = r_4(7) = r_4(11) = \cdots = 3$.

Example 1.52. Let $\mathcal{P}_{fin}(\mathbb{N})$ be the set of finite subsets of $\mathbb{N}$. Define the function $\phi : \mathcal{P}_{fin}(\mathbb{N}) \longrightarrow \mathbb{N}$ as

$$\phi(K) = \begin{cases} 0 & \text{if } K = \emptyset, \\ \sum_{i=1}^{p} 2^{n_i} & \text{if } K = \{n_1, \ldots, n_p\}. \end{cases}$$

It is easy to see that ϕ is a bijection.

Since a function is a relation, the ideas introduced in the previous section for relations in general can be equally well applied to functions. In particular, we can consider the inverse of a function and the product of two functions.

If f is a function, then, by Theorem 1.39, f^{-1} is a one-to-one relation; however, f^{-1} is not necessarily a function. In fact, by the same theorem, if f is a function, then f^{-1} is a function if and only if f is an injection.

Suppose now that $f : S \rightsquigarrow T$ is an injection. Then, $f^{-1} : T \rightsquigarrow S$ is also an injection. Further, $f^{-1} : T \rightsquigarrow S$ is total if and only if $f : S \rightsquigarrow T$ is a surjection, and $f^{-1} : T \rightsquigarrow S$ is a surjection if and only if $f : S \rightsquigarrow T$ is total. It follows that $f : S \rightsquigarrow T$ is a bijection if and only if $f^{-1} : T \rightsquigarrow S$ is a bijection.

If f and g are functions, then we will always use the alternative notation gf instead of the notation fg used for the relation product. We will refer to gf as the *composition* of f and g rather than the product.

By Theorem 1.42, the composition of two functions is a function. In fact, it follows from the definition of composition that

$$\mathrm{Dom}(gf) = \{s \in \mathrm{Dom}(f) \mid f(s) \in \mathrm{Dom}(g)\}$$

and, for all $s \in \mathrm{Dom}(gf)$,

$$gf(s) = g(f(s)).$$

This explains why we use gf rather than fg. If we used the other notation, the previous equation would become $fg(s) = g(f(s))$, which is rather confusing.

Definition 1.53. *Let $f : S \longrightarrow T$. A* left inverse (relative to S and T) *for f is a function $g : T \longrightarrow S$ such that $gf = 1_S$. A* right inverse (relative to S and T) *for f is a function $g : T \longrightarrow S$ such that $fg = 1_T$.*

Theorem 1.54. *Let $f : S \longrightarrow T$.*

(i) *f is a surjection if and only if f has a right inverse (relative to S and T).*

(ii) *If S is nonempty, then f is an injection if and only if f has a left inverse (relative to S and T).*

Proof. To prove the first part, suppose first that $f : S \longrightarrow T$ is a surjection. Define a function $g : T \longrightarrow S$ as follows: For each $y \in T$, let $g(y)$ be some arbitrarily chosen element $x \in S$ such that $f(x) = y$. (Such an x exists because f is surjective.) Then, by definition, $f(g(y)) = y$ for all $y \in T$, so g is a right inverse for f. Conversely, suppose that f has a right inverse g. Let $y \in T$ and let $x = g(y)$. Then, we have $f(x) = f(g(y)) = 1_T(y) = y$. Thus, f is surjective.

To prove the second part, first suppose that $f : S \longrightarrow T$ is an injection and S is nonempty. Let x_0 be some fixed element of S. Define a function

$g : T \longrightarrow S$ as follows: If $y \in \text{Ran}(f)$, then, since f is an injection, there is a unique element $x \in S$ such that $f(x) = y$. Define $g(y)$ to be this x. If $y \in T - \text{Ran}(f)$, define $g(y) = x_0$. Then, it is immediate from the definition of g that, for all $x \in S$, $g(f(x)) = x$, so g is a left inverse for f. Conversely, suppose that f has a left inverse g. For all $x_1, x_2 \in S$, if $f(x_1) = f(x_2)$, we have $x_1 = 1_S(x_1) = g(f(x_1)) = g(f(x_2)) = 1_S(x_2) = x_2$. Hence, f is an injection. □

We have used in this proof (without an explicit mention) an axiom of set theory that we discuss in Section 1.4. For a proof that makes explicit use of this axiom, see Supplement 38.

Theorem 1.55. *Let $f : S \longrightarrow T$. Then, the following statements are equivalent:*

(i) *f is a bijection.*
(ii) *There is a function $g : T \longrightarrow S$ that is both a left and a right inverse for f.*
(iii) *f has both a left inverse and a right inverse.*

Further, if f is a bijection, then f^{-1} is the only left inverse that f has, and it is the only right inverse that f has.

Proof. (i) implies (ii): If $f : S \longrightarrow B$ is a bijection, then $f^{-1} : T \longrightarrow S$ is both a left and a right inverse for f.

(ii) implies (iii): This implication is obvious.

(iii) implies (i): If f has both a left inverse and a right inverse and $S \neq \emptyset$, then it follows immediately from Theorem 1.54 that f is both injective and surjective, so f is a bijection. If $S = \emptyset$, then the existence of a left inverse function from T to S implies that T is also empty; this means that f is the empty function, which is a bijection from the empty set to itself.

Finally, suppose that $f : S \longrightarrow T$ is a bijection and that $g : T \longrightarrow S$ is a left inverse for f. Then, we have

$$f^{-1} = 1_S f^{-1} = (gf)f^{-1} = g(ff^{-1}) = g1_T = g.$$

Thus, f^{-1} is the unique left inverse for f. A similar proof shows that f^{-1} is the unique right inverse for f. □

To prove that $f : S \longrightarrow T$ is a bijection one could prove directly that f is both one-to-one and onto. Theorem 1.55 provides an alternative way. If we can define a function $g : T \longrightarrow S$ and show that g is both a left and a right inverse for f, then f is a bijection and $g = f^{-1}$.

The next definition provides another way of viewing a subset of a set S.

Definition 1.56. *Let S be a set. An* indicator function *over S is a function $I : S \longrightarrow \{0, 1\}$.*

If P is a subset of S, then the indicator function *of P (as a subset of S) is the function $I_P : S \longrightarrow \{0, 1\}$ given by*

$$I_P(x) = \begin{cases} 1 & \text{if } x \in P \\ 0 & \text{otherwise,} \end{cases}$$

for every $x \in S$.

It is easy to see that

$$\begin{aligned} I_{P \cap Q}(x) &= I_P(x) \cdot I_Q(x), \\ I_{P \cup Q}(x) &= I_P(x) + I_Q(x) - I_P(x) \cdot I_Q(x), \\ I_{\bar{P}}(x) &= 1 - I_P(x), \end{aligned}$$

for every $P, Q \subseteq S$ and $x \in S$.

The relationship between the subsets of a set and indicator functions defined on that set is discussed next.

Theorem 1.57. *There is a bijection* $\Psi : \mathcal{P}(S) \longrightarrow (S \longrightarrow \{0, 1\})$ *between the set of subsets of* S *and the set of indicator functions defined on* S.

Proof. For $P \in \mathcal{P}(S)$, define $\Psi(P) = I_P$. The mapping Ψ is one-to-one. Indeed, assume that $I_P = I_Q$, where $P, Q \in \mathcal{P}(S)$. We have $x \in P$ if and only if $I_P(x) = 1$, which is equivalent to $I_Q(x) = 1$. This happens if and only if $x \in Q$; hence, $P = Q$ so Ψ is one-to-one.

Let $f : S \longrightarrow \{0, 1\}$ be an arbitrary function. Define the set $T_f = \{x \in S \mid f(x) = 1\}$. It is easy to see that f is the indicator function of the set T_f. Hence, $\Psi(T_f) = f$, which shows that the mapping Ψ is also onto and hence it is a bijection. □

Definition 1.58. *A* simple function on a set S *is a function* $f : S \longrightarrow \mathbb{R}$ *that has a finite range.*

Simple functions are linear combinations of indicator functions, as we show next.

Theorem 1.59. *Let* $f : S \longrightarrow \mathbb{R}$ *be a simple function such that* $\text{Ran}(f) = \{y_1, \ldots, y_n\} \subseteq \mathbb{R}$. *Then,*

$$f = \sum_{i=1}^{n} y_i I_{f^{-1}(y_i)}.$$

Proof. Let $x \in \mathbb{R}$. If $f(x) = y_j$, then

$$I_{f^{-1}(y_\ell)}(x) = \begin{cases} 1 & \text{if } \ell = j, \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$\left(\sum_{i=1}^{n} y_i I_{f^{-1}(y_i)} \right)(x) = y_j,$$

which shows that $f(x) = \left(\sum_{i=1}^{n} y_i I_{f^{-1}(y_i)}\right)(x)$. □