REAL ANALYSIS AND PROBABILITY

R. M. DUDLEY
Massachusetts Institute of Technology

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 1989 by Wadsworth, Inc.
Cambridge University Press edition published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface Times Roman 10.25/13 pt. *System* LaTeX 2_ε [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data available

ISBN 0 521 80972 X hardback
ISBN 0 521 00754 2 paperback

Contents

1

Foundations; Set Theory

In constructing a building, the builders may well use different techniques and materials to lay the foundation than they use in the rest of the building. Likewise, almost every field of mathematics can be built on a foundation of axiomatic set theory. This foundation is accepted by most logicians and mathematicians concerned with foundations, but only a minority of mathematicians have the time or inclination to learn axiomatic set theory in detail.

To make another analogy, higher-level computer languages and programs written in them are built on a foundation of computer hardware and systems programs. How much the people who write high-level programs need to know about the hardware and operating systems will depend on the problem at hand.

In modern real analysis, set-theoretic questions are somewhat more to the fore than they are in most work in algebra, complex analysis, geometry, and applied mathematics. A relatively recent line of development in real analysis, "nonstandard analysis," allows, for example, positive numbers that are infinitely small but not zero. Nonstandard analysis depends even more heavily on the specifics of set theory than earlier developments in real analysis did.

This chapter will give only enough of an introduction to set theory to define some notation and concepts used in the rest of the book. In other words, this chapter presents mainly "naive" (as opposed to axiomatic) set theory. Appendix A gives a more detailed development of set theory, including a listing of axioms, but even there, the book will not enter into nonstandard analysis or develop enough set theory for it.

Many of the concepts defined in this chapter are used throughout mathematics and will, I hope, be familiar to most readers.

1.1. Definitions for Set Theory and the Real Number System

Definitions can serve at least two purposes. First, as in an ordinary dictionary, a definition can try to give insight, to convey an idea, or to explain a less familiar idea in terms of a more familiar one, but with no attempt to specify or exhaust completely the meaning of the word being defined. This kind of definition will be called *informal*. A *formal* definition, as in most of mathematics and parts of other sciences, may be quite precise, so that one can decide scientifically whether a statement about the term being defined is true or not. In a formal definition, a familiar term, such as a common unit of length or a number, may be defined in terms of a less familiar one. Most definitions in set theory are formal. Moreover, set theory aims to provide a coherent logical structure not only for itself but for just about all of mathematics. There is then a question of where to begin in giving definitions.

Informal dictionary definitions often consist of synonyms. Suppose, for example, that a dictionary simply defined “high” as “tall” and “tall” as “high.” One of these definitions would be helpful to someone who knew one of the two words but not the other. But to an alien from outer space who was trying to learn English just by reading the dictionary, these definitions would be useless. This situation illustrates on the smallest scale the whole problem the alien would have, since all words in the dictionary are defined in terms of other words. To make a start, the alien would have to have some way of interpreting at least a few of the words in the dictionary other than by just looking them up.

In any case some words, such as the conjunctions “and,” “or,” and “but,” are very familiar but hard to define as separate words. Instead, we might have rules that define the meanings of phrases containing conjunctions given the meanings of the words or subphrases connected by them.

At first thought, the most important of all definitions you might expect in set theory would be the definition of “set,” but quite the contrary, just because the entire logical structure of mathematics reduces to or is defined in terms of this notion, it cannot necessarily be given a formal, precise definition. Instead, there are rules (axioms, rules of inference, etc.) which in effect provide the meaning of “set.” A preliminary, informal definition of *set* would be “any collection of mathematical objects,” but this notion will have to be clarified and adjusted as we go along.

The problem of defining *set* is similar in some ways to the problem of defining *number*. After several years of school, students “know” about the numbers $0, 1, 2, \ldots,$ in the sense that they know rules for operating with numbers. But many people might have a hard time saying exactly what a number is. Different people might give different definitions of the number 1, even though they completely agree on the rules of arithmetic.

In the late 19th century, mathematicians began to concern themselves with giving precise definitions of numbers. One approach is that beginning with 0, we can generate further integers by taking the “successor” or “next larger integer.”

If 0 is defined, and a successor operation is defined, and the successor of any integer n is called n', then we have the sequence $0, 0', 0'', 0''', \ldots$. In terms of 0 and successors, we could then write down definitions of the usual integers. To do this I'll use an equals sign with a colon before it, ":=," to mean "equals by definition." For example, $1 := 0'$, $2 := 0''$, $3 := 0'''$, $4 := 0''''$, and so on. These definitions are precise, as far as they go. One could produce a thick dictionary of numbers, equally precise (though not very useful) but still incomplete, since 0 and the successor operation are not formally defined. More of the structure of the number system can be provided by giving rules about 0 and successors. For example, one rule is that if $m' = n'$, then $m = n$.

Once there are enough rules to determine the structure of the nonnegative integers, then what is important is the structure rather than what the individual elements in the structure actually are.

In summary: if we want to be as precise as possible in building a rigorous logical structure for mathematics, then informal definitions cannot be part of the structure, although of course they can help to explain it. Instead, at least some basic notions must be left undefined. Axioms and other rules are given, and other notions are defined in terms of the basic ones.

Again, informally, a set is any collection of objects. In mathematics, the objects will be mathematical ones, such as numbers, points, vectors, or other sets. (In fact, from the set-theoretic viewpoint, all mathematical objects are sets of one kind or another.) If an object x is a member of a set y, this is written as "$x \in y$," sometimes also stated as "x belongs to y" or "x is in y." If S is a finite set, so that its members can be written as a finite list $x_1, \ldots, x_n$, then one writes $S = \{x_1, \ldots, x_n\}$. For example, $\{2, 3\}$ is the set whose only members are the numbers 2 and 3. The notion of membership, "$\in$," is also one of the few basic ones that are formally undefined.

A set can have just one member. Such a set, whose only member is x, is called $\{x\}$, read as "singleton x." In set theory a distinction is made between $\{x\}$ and x itself. For example if $x = \{1, 2\}$, then x has two members but $\{x\}$ only one.

A set A is *included* in a set B, or is a *subset* of B, written $A \subset B$, if and only if every member of A is also a member of B. An equivalent statement is that B *includes* A, written $B \supset A$. To say B *contains* x means $x \in B$. Many authors also say B contains A when $B \supset A$.

The phrase "if and only if" will sometimes be abbreviated "iff." For example, $A \subset B$ iff for all x, if $x \in A$, then $x \in B$.

One of the most important rules in set theory is called "extensionality." It says that if two sets A and B have the same members, so that for any object

x, $x \in A$ if and only if $x \in B$, or equivalently both $A \subset B$ and $B \subset A$, then the sets are equal, $A = B$. So, for example, $\{2, 3\} = \{3, 2\}$. The order in which the members happen to be listed makes no difference, as long as the members are the same. In a sense, extensionality is a definition of equality for sets. Another view, more common among set theorists, is that any two objects are equal if and only if they are identical. So "$\{2, 3\}$" and "$\{3, 2\}$" are two names of one and the same set.

Extensionality also contributes to an informal definition of *set*. A set is defined simply by what its members are—beyond that, structures and relationships between the members are irrelevant to the definition of the set.

Other than giving finite lists of members, the main way to define specific sets is to give a condition that the members satisfy. In notation, $\{x : \ldots\}$ means the set of all x such that.... For example, $\{x : (x-4)^2 = 4\} = \{2, 6\} = \{6, 2\}$.

In line with a general usage that a slash through a symbol means "not," as in $a \neq b$, meaning "a is not equal to b," the symbol "$\notin$" means "is not a member of." So $x \notin y$ means x is not a member of y, as in $3 \notin \{1, 2\}$.

Defining sets via conditions can lead to contradictions if one is not careful. For example, let $r = \{x : x \notin x\}$. Then $r \notin r$ implies $r \in r$ and conversely (Bertrand Russell's paradox). This paradox can be avoided by limiting the condition to some set. Thus $\{x \in A : \ldots x \ldots\}$ means "the set of all x in A such that $\ldots x \ldots$." As long as this form of definition is used when A is already known to be a set, new sets can be defined this way, and it turns out that no contradictions arise.

It might seem peculiar, anyhow, for a set to be a member of itself. It will be shown in Appendix A (Theorem A.1.9), from the axioms of set theory listed there, that no set is a member of itself. In this sense, the collection r of sets named in Russell's paradox is the collection of all sets, sometimes called the "universe" in set theory. Here the informal notion of set as any collection of objects is indeed imprecise. The axioms in Appendix A provide conditions under which certain collections are or are not sets. For example, the universe is not a set.

Very often in mathematics, one is working for a while inside a fixed set y. Then an expression such as $\{x : \ldots x \ldots\}$ is used to mean $\{x \in y : \ldots x \ldots\}$.

Now several operations in set theory will be defined. In cases where it may not be obvious that the objects named are sets, there are axioms which imply that they are (Appendix A).

There is a set, called $\emptyset$, the "empty set," which has no members. That is, for all x, $x \notin \emptyset$. This set is unique, by extensionality. If B is any set, then 2^B, also called the "power set" of B, is the set of all subsets of B. For example, if B has 3 members, then 2^B has $2^3 = 8$ members. Also, $2^{\emptyset} = \{\emptyset\} \neq \emptyset$.

$A \cap B$, called the intersection of A and B, is defined by $A \cap B := \{x \in A: x \in B\}$. In other words, $A \cap B$ is the set of all x which belong to both A and B. $A \cup B$, called the union of A and B, is a set such that for any x, $x \in A \cup B$ if and only if $x \in A$ or $x \in B$ (or both). Also, $A \setminus B$ (read "A minus B") is the set of all x in A which are not in B, sometimes called the *relative complement* (of B in A). The *symmetric difference* $A \,\Delta\, B$ is defined as $(A \setminus B) \cup (B \setminus A)$.

$\mathbb{N}$ will denote the set of all nonnegative integers 0, 1, 2, ... (Formally, nonnegative integers are usually defined by defining 0 as the empty set $\emptyset$, 1 as $\{\emptyset\}$, and generally the successor operation mentioned above by $n' = n \cup \{n\}$, as is treated in more detail in Appendix A.)

Informally, an *ordered pair* consists of a pair of mathematical objects in a given order, such as $\langle x, y\rangle$, where x is called the "first member" and y the "second member" of the ordered pair $\langle x, y\rangle$. Ordered pairs satisfy the following axiom: for all x, y, u, and v, $\langle x, y\rangle = \langle u, v\rangle$ if and only if both $x = u$ and $y = v$. In an ordered pair $\langle x, y\rangle$ it may happen that $x = y$. Ordered pairs can be defined formally in terms of (unordered, ordinary) sets so that the axiom is satisfied; the usual way is to set $\langle x, y\rangle := \{\{x\}, \{x, y\}\}$ (as in Appendix A). Note that $\{\{x\}, \{x, y\}\} = \{\{y, x\}, \{x\}\}$ by extensionality.

One of the main ideas in all of mathematics is that of function. Informally, given sets D and E, a function f on D is defined by assigning to each x in D one (and only one!) member $f(x)$ of E. Formally, a *function* is defined as a set f of ordered pairs $\langle x, y\rangle$ such that for any x, y, and z, if $\langle x, y\rangle \in f$ and $\langle x, z\rangle \in f$, then $y = z$. For example, $\{\langle 2, 4\rangle, \langle -2, 4\rangle\}$ is a function, but $\{\langle 4, 2\rangle, \langle 4, -2\rangle\}$ is not a function. A set of ordered pairs which is (formally) a function is, informally, called the *graph* of the function (as in the case $D = E = \mathbb{R}$, the set of real numbers).

The *domain*, *dom f*, of a function f is the set of all x such that for some y, $\langle x, y\rangle \in f$. Then y is uniquely determined, by definition of function, and it is called $f(x)$. The *range*, *ran f*, of f is the set of all y such that $f(x) = y$ for some x.

A function f with domain A and range included in a set B is said to be *defined on A* or *from A into B*. If the range of f equals B, then f is said to be *onto B*.

The symbol "$\mapsto$" is sometimes used to describe or define a function. A function f is written as "$x \mapsto f(x)$." For example, "$x \mapsto x^3$" or "$f: x \mapsto x^3$" means a function f such that $f(x) = x^3$ for all x (in the domain of f). To specify the domain, a related notation in common use is, for example, "$f: A \mapsto B$," which together with a more specific definition of f indicates that it is defined from A into B (but does not mean that $f(A) = B$; to distinguish the two related usages of $\mapsto$, A and B are written in capitals and members of them in small letters, such as x).

If X is any set and A any subset of X, the *indicator function* of A (on X) is the function defined by

$$1_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

(Many mathematicians call this the *characteristic* function of A. In probability theory, "characteristic function" happens to mean a Fourier transform, to be treated in Chapter 9.)

A *sequence* is a function whose domain is either $\mathbb{N}$ or the set $\{1, 2, \ldots\}$ of all positive integers. A sequence f with $f(n) = x_n$ for all n is often written as $\{x_n\}_{n\geq 1}$ or the like.

Formally, every set is a set of sets (every member of a set is also a set). If a set is to be viewed, also informally, as consisting of sets, it is often called a family, class, or collection of sets. Let $\mathcal{V}$ be a family of sets. Then the *union* of $\mathcal{V}$ is defined by

$$\bigcup \mathcal{V} := \{x : x \in A \text{ for some } A \in \mathcal{V}\}.$$

Likewise, the *intersection* of a non-empty collection $\mathcal{V}$ is defined by

$$\bigcap \mathcal{V} := \{x : x \in A \text{ for all } A \in \mathcal{V}\}.$$

So for any two sets A and B, $\bigcup\{A, B\} = A \cup B$ and $\bigcap\{A, B\} = A \cap B$. Notations such as $\bigcup \mathcal{V}$ and $\bigcap \mathcal{V}$ are most used within set theory itself. In the rest of mathematics, unions and intersections of more than two sets are more often written with indices. If $\{A_n\}_{n\geq 1}$ is a sequence of sets, their union is written as

$$\bigcup_n A_n := \bigcup_{n=1}^{\infty} A_n := \{x : x \in A_n \text{ for some } n\}.$$

Likewise, their intersection is written as

$$\bigcap_{n\geq 1} A_n := \bigcap_{n=1}^{\infty} A_n := \{x : x \in A_n \text{ for all } n\}.$$

The union of finitely many sets $A_1, \ldots, A_n$ is written as

$$\bigcup_{1\leq i\leq n} A_i := \bigcup_{i=1}^{n} A_i := \{x : x \in A_i \text{ for some } i = 1, \ldots, n\},$$

and for intersections instead of unions, replace "some" by "all."

More generally, let I be any set, and suppose A is a function defined on I whose values are sets $A_i := A(i)$. Then the union of all these sets A_i is written

$$\bigcup_i A_i := \bigcup_{i \in I} A_i := \{x : x \in A_i \text{ for some } i\}.$$

A set I in such a situation is called an *index set*. This just means that it is the domain of the function $i \mapsto A_i$. The index set I can be omitted from the notation, as in the first expression above, if it is clear from the context what I is. Likewise, the intersection is written as

$$\bigcap_i A_i := \bigcap_{i \in I} A_i := \{x : x \in A_i \text{ for all } i \in I\}.$$

Here, usually, I is a non-empty set. There is an exception when the sets under discussion are all subsets of one given set, say X. Suppose $t \notin I$ and let $A_t := X$. Then replacing I by $I \cup \{t\}$ does not change $\bigcap_{i \in I} A_i$ if I is non-empty. In case I is empty, one can set $\bigcap_{i \in \emptyset} A_i = X$.

Two more symbols from mathematical logic are sometimes useful as abbreviations: $\forall$ means "for all" and $\exists$ means "there exists." For example, $(\forall x \in A)(\exists y \in B)\ldots$ means that for all x in A, there is a y in B such that....

Two sets A and B are called *disjoint* iff $A \cap B = \emptyset$. Sets A_i for $i \in I$ are called disjoint iff $A_i \cap A_j = \emptyset$ for all $i \neq j$ in I.

Next, some definitions will be given for different classes of numbers, leading up to a definition of real numbers. It is assumed that the reader is familiar with integers and rational numbers. A somewhat more detailed and formal development is given in Appendix A.4.

Recall that $\mathbb{N}$ is the set of all nonnegative integers $0, 1, 2, \ldots$, $\mathbb{Z}$ denotes the set of all integers $0, \pm 1, \pm 2, \ldots$, and $\mathbb{Q}$ is the set of all rational numbers m/n, where $m \in \mathbb{Z}$, $n \in \mathbb{Z}$, and $n \neq 0$.

Real numbers can be defined in different ways. A familiar way is through decimal expansions: x is a real number if and only if $x = \pm y$, where $y = n + \sum_{j=1}^{\infty} d_j/10^j$, $n \in \mathbb{N}$, and each digit d_j is an integer from 0 to 9. But decimal expansions are not very convenient for proofs in analysis, and they are not unique for rational numbers of the form $m/10^k$ for $m \in \mathbb{Z}$, $m \neq 0$, and $k \in \mathbb{N}$. One can also define real numbers x in terms of more general sequences of rational numbers converging to x, as in the completion of metric spaces to be treated in §2.5.

The formal definition of real numbers to be used here will be by way of Dedekind cuts, as follows: A *cut* is a set $C \subset \mathbb{Q}$ such that $C \notin \emptyset$; $C \neq \mathbb{Q}$; whenever $q \in C$, if $r \in \mathbb{Q}$ and $r < q$ then $r \in C$, and there exists $s \in \mathbb{Q}$ with $s > q$ and $s \in C$.

Let $\mathbb{R}$ be the set of all real numbers; thus, formally, $\mathbb{R}$ is the set of all cuts. Informally, a one-to-one correspondence between real numbers x and cuts C, written $C = C_x$ or $x = x_C$, is given by $C_x = \{q \in \mathbb{Q}: q < x\}$.

The ordering $x \le y$ for real numbers is defined simply in terms of cuts by $C_x \subset C_y$. A set E of real numbers is said to be *bounded above* with an *upper bound* y iff $x \le y$ for all $x \in E$. Then y is called the *supremum* or *least upper bound* of E, written $y = \sup E$, iff it is an upper bound and $y \le z$ for every upper bound z of E. A basic fact about $\mathbb{R}$ is that for every non-empty set $E \subset \mathbb{R}$ such that E is bounded above, the supremum $y = \sup E$ exists. This is easily proved by cuts: C_y is the union of the cuts C_x for all $x \in E$, as is shown in Theorem A.4.1 of Appendix A.

Similarly, a set F of real numbers is *bounded below* with a *lower bound* v if $v \le x$ for all $x \in F$, and v is the *infimum* of F, $v = \inf F$, iff $t \le v$ for every lower bound t of F. Every non-empty set F which is bounded below has an infimum, namely, the supremum of the lower bounds of F (which are a non-empty set, bounded above).

The maximum and minimum of two real numbers are defined by $\min(x, y) = x$ and $\max(x, y) = y$ if $x \le y$; otherwise, $\min(x, y) = y$ and $\max(x, y) = x$.

For any real numbers $a \le b$, let $[a, b] := \{x \in \mathbb{R}: a \le x \le b\}$.

For any two sets X and Y, their *Cartesian product*, written $X \times Y$, is defined as the set of all ordered pairs $\langle x, y\rangle$ for x in X and y in Y. The basic example of a Cartesian product is $\mathbb{R} \times \mathbb{R}$, which is also written as $\mathbb{R}^2$ (pronounced r-two, not r-squared), and called the *plane*.

Problems

1. Let $A := \{3, 4, 5\}$ and $B := \{5, 6, 7\}$. Evaluate: (a) $A \cup B$. (b) $A \cap B$. (c) $A \backslash B$. (d) $A \,\Delta\, B$.
2. Show that $\emptyset \ne \{\emptyset\}$ and $\{\emptyset\} \ne \{\{\emptyset\}\}$.
3. Which of the following three sets are equal? (a) $\{\{2, 3\}, \{4\}\}$; (b) $\{\{4\}, \{2, 3\}\}$; (c) $\{\{4\}, \{3, 2\}\}$.
4. Which of the following are functions? Why?
   (a) $\{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 3, 1\rangle\}$.
   (b) $\{\langle 1, 2\rangle, \langle 2, 3\rangle, \langle 2, 1\rangle\}$.
   (c) $\{\langle 2, 1\rangle, \langle 3, 1\rangle, \langle 1, 2\rangle\}$.
   (d) $\{\langle x, y\rangle \in \mathbb{R}^2: x = y^2\}$.
   (e) $\{\langle x, y\rangle \in \mathbb{R}^2: y = x^2\}$.
5. For any relation V (that is, any set of ordered pairs), define the *domain* of

V as $\{x: \langle x, y\rangle \in V$ for some $y\}$, and the *range* of V as $\{y: \langle x, y\rangle \in V$ for some $x\}$. Find the domain and range for each relation in the last problem (whether or not it is a function).

6. Let $A_{1j} := \mathbb{R} \times [j-1, j]$ and $A_{2j} := [j-1, j] \times \mathbb{R}$ for $j = 1, 2$. Let $B := \bigcup_{m=1}^{2} \bigcap_{n=1}^{2} A_{mn}$ and $C := \bigcap_{n=1}^{2} \bigcup_{m=1}^{2} A_{mn}$. Which of the following is true: $B \subset C$ and/or $C \subset B$? Why?
7. Let $f(x) := \sin x$ for all $x \in \mathbb{R}$. Of the following subsets of $\mathbb{R}$, which is f into, and which is it onto? (a) $[-2, 2]$. (b) $[0, 1]$. (c) $[-1, 1]$. (d) $[-\pi, \pi]$.
8. How is Problem 7 affected if x is measured in degrees rather than radians?
9. Of the following sets, which are included in others? $A := \{3, 4, 5\}$; $B := \{\{3, 4\}, 5\}$; $C := \{5, 4\}$; and $D := \{\{4, 5\}\}$. Assume that no nonobvious relations, such as $4 = \{3, 5\}$, are true. More specifically, you can assume that for any two sets x and y, at most one of the three relations holds: $x \in y$, $x = y$, or $y \in x$, and that each nonnegative integer k is a set with k members. Please *explain* why each inclusion does or does not hold. Sample: If $\{\{6, 7\}, \{5\}\} \subset \{3, 4\}$, then by extensionality $\{6, 7\} = 3$ or 4, but $\{6, 7\}$ has two members, not three or four.
10. Let $I := [0, 1]$. Evaluate $\bigcup_{x\in I}[x, 2]$ and $\bigcap_{x\in I}[x, 2]$.
11. "Closed half-lines" are subsets of $\mathbb{R}$ of the form $\{x \in \mathbb{R}: x \le b\}$ or $\{x \in \mathbb{R}: x \ge b\}$ for real numbers b. A *polynomial of degree n* on $\mathbb{R}$ is a function $x \mapsto a_n x^n + \cdots + a_1 x + a_0$ with $a_n \neq 0$. Show that the range of any polynomial of degree $n \ge 1$ is $\mathbb{R}$ for n odd and a closed half-line for n even. *Hints:* Show that for large values of $|x|$, the polynomial has the same sign as its leading term $a_n x^n$ and its absolute value goes to ∞. Use the intermediate value theorem for a continuous function such as a polynomial (Problem 2.2.14(d) below).
12. A polynomial on $\mathbb{R}^2$ is a function of the form $\langle x, y\rangle \mapsto \sum_{0\le i\le k, 0\le j\le k} a_{ij} x^i y^j$. Show that the ranges of nonconstant polynomials on $\mathbb{R}^2$ are either all of $\mathbb{R}$, closed half-lines, or open half-lines $(b, \infty) := \{x \in \mathbb{R}: x > b\}$ or $(-\infty, b) := \{x \in \mathbb{R}: x < b\}$, where each open or closed half-line is the range of some polynomial. *Hint:* For one open half-line, try the polynomial $x^2 + (xy - 1)^2$.

1.2. Relations and Orderings

A *relation* is any set of ordered pairs. For any relation E, the inverse relation is defined by $E^{-1} := \{\langle y, x\rangle: \langle x, y\rangle \in E\}$. Thus, a function is a special kind of relation. Its inverse f^{-1} is not necessarily a function. In fact, a function f is called 1–1 or *one-to-one* if and only if f^{-1} is also a function. Given a relation E, one often writes xEy instead of $\langle x, y\rangle \in E$ (this notation is used not for functions but for other relations, as will soon be explained). Given a set X, a relation $E \subset X \times X$ is called *reflexive* on X iff xEx for all $x \in X$. E is called *symmetric* iff $E = E^{-1}$. E is called *transitive* iff whenever xEy and yEz, we have xEz. Examples of transitive relations are orderings, such as $x \leq y$.

A relation $E \subset X \times X$ is called an *equivalence relation* iff it is reflexive on X, symmetric, and transitive. One example of an equivalence relation is equality. In general, an equivalence relation is like equality; two objects x and y satisfying an equivalence relation are equal in some way. For example, two integers m and n are said to be equal mod p iff $m - n$ is divisible by p. Being equal mod p is an equivalence relation. Or if f is a function, one can define an equivalence relation E_f by $xE_f y$ iff $f(x) = f(y)$.

Given an equivalence relation E, an *equivalence class* is a set of the form $\{y \in X: yEx\}$ for any $x \in X$. It follows from the definition of equivalence relation that two equivalence classes are either disjoint or identical. Let $f(x) := \{y \in X: yEx\}$. Then f is a function and xEy if and only if $f(x) = f(y)$, so $E = E_f$, and every equivalence relation can be written in the form E_f.

A relation E is called *antisymmetric* iff whenever xEy and yEx, then $x = y$. Given a set X, a *partial ordering* is a transitive, antisymmetric relation $E \subset X \times X$. Then $\langle X, E\rangle$ is called a *partially ordered set*. For example, for any set Y, let $X = 2^Y$ (the set of all subsets of Y). Then $\langle 2^Y, \subset\rangle$, for the usual inclusion $\subset$, gives a partially ordered set. (*Note:* Many authors require that a partial ordering also be reflexive. The current definition is being used to allow not only relations '$\leq$' but also '$<$' to be partial orderings.) A partial ordering will be called *strict* if xEx does not hold for any x. So "strict" is the opposite of "reflexive." For any partial ordering E, define the relation $\leq$ by $x \leq y$ iff (xEy or $x = y$). Then $\leq$ is a reflexive partial ordering. Also, define the relation $<$ by $x < y$ iff (xEy and $x \neq y$). Then $<$ is a strict partial ordering. For example, the usual relations $<$ and $\leq$ between real numbers are connected in the way just defined. A one-to-one correspondence between strict partial orderings E and reflexive partial orderings F on a set X is given by $F = E \cup D$ and $E = F \backslash D$, where D is the "diagonal," $D := \{\langle x, x\rangle : x \in X\}$. From here on, the partial orderings considered will be either reflexive, usually written $\leq$ (or $\geq$), or strict, written $<$ (or $>$). Here, as usual, "$<$" is read "less than," and so forth.

Two partially ordered sets $\langle X, E\rangle$ and $\langle Y, G\rangle$ are said to be *order-isomorphic* iff there exists a 1–1 function f from X onto Y such that for any

u and x in X, uEx iff $f(u)Gf(x)$. Then f is called an *order-isomorphism*. For example, the intervals [0, 1] and [0, 2], with the usual ordering for real numbers, are order-isomorphic via the function $f(x) = 2x$. The interval [0, 1] is not order-isomorphic to $\mathbb{R}$, which has no smallest element.

From here on, an ordered pair $\langle x, y\rangle$ will often be written as (x, y) (this is, of course, still different from the unordered pair $\{x, y\}$).

A *linear ordering* E of X is a partial ordering E of X such that for all x and $y \in X$, either xEy, yEx, or $x = y$. Then $\langle X, E\rangle$ is called a *linearly ordered set*. The classic example of a linearly ordered set is the real line $\mathbb{R}$, with its usual ordering. Actually, $(\mathbb{R}, <)$, $(\mathbb{R}, \leq)$, $(\mathbb{R}, >)$, and $(\mathbb{R}, \geq)$ are all linearly ordered sets.

If (X, E) is any partially ordered set and A is any subset of X, then $\{\langle x, y\rangle \in E : x \in A \text{ and } y \in A\}$ is also a partial ordering on A. Suppose we call it E_A. For most orderings, as on the real numbers, the orderings of subsets will be written with the same symbol as on the whole set. If (X, E) is linearly ordered and $A \subset X$, then (A, E_A) is also linearly ordered, as follows directly from the definitions.

Let W be a set with a reflexive linear ordering $\leq$. Then W is said to be *well-ordered* by $\leq$ iff for every non-empty subset A of W there is a smallest $x \in A$, so that for all $y \in A$, $x \leq y$. The corresponding strict linear ordering $<$ will also be called a well-ordering. If X is a finite set, then any linear ordering of it is easily seen to be a well-ordering. The interval [0, 1] is not well-ordered, although it has a smallest element 0, since it has subsets, such as $\{x: 0 < x \leq 1\}$, with no smallest element.

The method of proof by mathematical induction can be extended to well-ordered sets, as follows. Suppose $(X, <)$ is a well-ordered set and that we want to prove that some property holds for all elements of X. If it does not, then there is a smallest element for which the property fails. It suffices, then, to prove that for each $x \in X$, if the property holds for all $y < x$, then it holds for x. This "induction principle" will be treated in more detail in §1.3.

Problems

1. For any partial ordering E, show that E^{-1} is also a partial ordering.
2. For two partially ordered sets $\langle A, \leq\rangle$ and $\langle B, \leq\rangle$, the *lexicographical ordering* on the Cartesian product $A \times B$ is defined by $\langle a, b\rangle \leq \langle c, d\rangle$ iff $a < c$ or ($a = c$ and $b \leq d$). (For example, if A and B are both an alphabet with the usual ordering, then we have the dictionary or "alphabetical" ordering of two-letter words or strings.) If the orderings on A and B are linear, show that the lexicographical ordering is linear on $A \times B$. If A

and B are well-ordered by the given relations, show that $A \times B$ is also well-ordered.

3. Instead, let $\langle a, b\rangle \le \langle c, d\rangle$ iff both $a \le c$ and $b \le d$. Show that this is a partial ordering. If A and B each contain more than one element, show that $A \times B$ is never linearly ordered by such an ordering.
4. On $\mathbb{R}^2$ let $\langle x, y\rangle E\langle u, v\rangle$ iff $x+y=u+v$, $\langle x, y\rangle F\langle u, v\rangle$ iff $x+y \le u+v$, and $\langle x, y\rangle G\langle u, v\rangle$ iff $x+u \le y+v$. Which of E, F, and G is an equivalence relation, a partial ordering, or a linear ordering? Why?
5. For sequences $\{x_n\}$ of real numbers let $\{x_n\}E\{y_n\}$ iff $\lim_{n\to\infty} x_n - y_n = 0$ and $\{x_n\}F\{y_n\}$ iff $\lim_{n\to\infty} x_n - y_n = 1$. Which of E and F is an equivalence relation and/or a partial ordering? Why?
6. For any two relations E and F on the same set X, define a relation $G := E \circ F$ by xGz iff for some y, xEy and yFz. For each of the following properties, if E and F both have the property, prove, or disprove by an example, that G also has the property: (a) reflexive, (b) symmetric, (c) transitive.
7. Refer to Problem 6 and answer the same question in regard to the following properties: (d) antisymmetric, (e) equivalence relation, (f) function.

*1.3. Transfinite Induction and Recursion

Mathematical induction is a well-known and useful method of proving facts about nonnegative integers. If $F(n)$ represents a statement that one wants to prove for all $n \in \mathbb{N}$, and a direct proof is not apparent, one first proves $F(0)$. Then, in proving $F(n+1)$, one can assume that $F(n)$ is true, which is often helpful. Or, if you prefer, you can assume that $F(0), F(1), \ldots, F(n)$ are all true. More generally, let $(X, <)$ be any partially ordered set. A subset $Y \subset X$ will be called *inductive* if, for every $x \in X$ such that $y \in Y$ for all $y \in X$ such that $y < x$, we have $x \in Y$. If X has a least element x, then there are no $y < x$, so x must belong to any inductive subset Y of X. In ordinary induction, Y is the set of all n for which $F(n)$ holds. Proving that Y is inductive gives a proof that $Y = \mathbb{N}$, so that $F(n)$ holds for all n. In $\mathbb{R}$, the set $(-\infty, 0)$ is inductive, but it is not all of $\mathbb{R}$. The set $\mathbb{N}$ is well-ordered, but $\mathbb{R}$ is not: the set $\{x \in \mathbb{R}: x > 1\}$ has no least element. One of the main advantages of well-orderings is that they allow the following extension of induction:

1.3.1. Induction Principle Let X be any set well-ordered by a relation $<$. Let Y be any inductive subset of X. Then $Y = X$.

Proof. If $X \setminus Y = \emptyset$, the conclusion holds. Otherwise, if not, let y be the least element of $X \setminus Y$. Then $x \in Y$ for all $x < y$ (perhaps vacuously, if y is the least element of X), so $y \in Y$, a contradiction. □

For any linearly ordered set $(X, \leq)$, an *initial segment* is a subset $Y \subset X$ such that whenever $x < y$ and $y \in Y$, then also $x \in Y$. Then if $(X, \leq)$ is the real line with usual ordering and Y an initial segment, then either $Y = X$ or, for some y, either $Y = \{x: x < y\}$ or $Y = \{x: x \leq y\}$.

In ordinary mathematical induction, the set $(X, <)$ is order-isomorphic to $\mathbb{N}$, the set of nonnegative integers, or to some initial segment of it (finite integer) with usual ordering. Transfinite induction refers to induction for an $(X, <)$ with a more complicated well-ordering. One example is "double induction." To prove a statement $F(m, n)$ for all nonnegative integers m and n, one can first prove $F(0, 0)$. Then in proving $F(m, n)$ one can assume that $F(j, k)$ is true for all $j < m$ and all $k \in \mathbb{N}$, and for $j = m$ and $k < n$. (In this case the well-ordering is the "lexicographical" ordering mentioned in Problem 2 of §1.2.) Other well-orderings of $\mathbb{N} \times \mathbb{N}$ may also be useful. Much of set theory is concerned with well-orderings more general than those of sequences, such as well-orderings of $\mathbb{R}$, although these are in a sense nonconstructive (well-ordering of general sets, and of $\mathbb{R}$ in particular, depends on the axiom of choice, to be treated in §1.5).

Another very important method in mathematics, definition by recursion, will be developed next. In its classical form, a function f is defined by specifying $f(0)$, then defining $f(n)$ in terms of $f(n-1)$ and possibly other values of $f(k)$ for $k < n$. Such recursive definitions will also be extended to well-ordered sets. For any function f and $A \subset \operatorname{dom} f$, the restriction of f to A is defined by $f \upharpoonright A := \{\langle x, f(x)\rangle : x \in A\}$.

1.3.2. Recursion Principle Let $(X, <)$ be a well-ordered set and Y any set. For any $x \in X$, let $I(x) := \{u \in X: u < x\}$. Let g be a function whose domain is the set of all j such that for some $x \in X$, j is a function from $I(x)$ into Y, and such that $\operatorname{ran} g \subset Y$. Then there is a unique function f from X into Y such that for every $x \in X$, $f(x) = g(f \upharpoonright I(x))$.

Note. If b is the least element of X and we want to define $f(b) = c$, then we set $g(\emptyset) = c$ and note that $I(b) = \emptyset$.

Proof. If $X = \emptyset$, then $f = \emptyset$ and the conclusion holds. So suppose X is non-empty and let b be its smallest element. Let $J(x) := \{u \in X: u \leq x\}$ for each $x \in X$. Let T be the set of all $x \in X$ such that on $J(x)$, there is a function f such that $f(u) = g(f \restriction I(u))$ for all $u \in J(x)$. Let us show that if such an f exists, it is unique. Let h be another such function. Then $h(b) = g(\emptyset) = f(b)$. By induction (1.3.1), for each $u \in J(x)$, $h(u) = g(f \restriction I(u)) = f(u)$. So f is unique. If $x < u$ for some u in T and f as above is defined on $J(u)$, then $f \restriction J(x)$ has the desired properties and is the f for $J(x)$ by uniqueness. Thus T is an initial segment of X. The union of all the functions f for all $x \in T$ is a well-defined function, which will also be called f. If $T \neq X$, let u be the least element of $X \backslash T$. But then $T = I(u)$ and $f \cup \{\langle u, g(f)\rangle\}$ is a function on $J(u)$ with the desired properties, so $u \in T$, a contradiction. So f exists. As it is unique on each $J(x)$, it is unique. $\square$

For any function f on a Cartesian product $A \times B$, one usually writes $f(a, b)$ rather than $f(\langle a, b\rangle)$. The classical recursion on the nonnegative integers can then be described as follows.

1.3.3. Corollary (Simple Recursion) Let Y be any set, $c \in Y$, and h any function from $\mathbb{N} \times Y$ into Y. Then there is a function f from $\mathbb{N}$ into Y with $f(0) = c$ and for each $n \in \mathbb{N}$, $f(n + 1) = h(n, f(n))$.

Proof. To apply 1.3.2, let $g(\emptyset) = c$. Let j be any function from some non-empty $I(n)$ into Y. (Note that $I(n)$ is empty if and only if $n = 0$.) Then $n - 1$ is the largest member of $I(n)$. Let $g(j) = h(n - 1, j(n - 1))$. Then the function g is defined on all such functions j, and 1.3.2 applies to give a function f. Now $f(0) = c$, and for any $n \in \mathbb{N}$, $f(n+1) = g(f \restriction I(n+1)) = h(n, f(n))$. $\square$

Example. Let t be a function with real values defined on $\mathbb{N}$. Let

$$f(n) = \sum_{j=0}^{n} t(j).$$

To obtain f by simple recursion (1.3.3), let $c = t(0)$ and $h(n, y) = t(n+1) + y$ for any $n \in \mathbb{N}$ and $y \in \mathbb{R}$. A computer program to compute f, given a program for t, could well be written along the lines of this recursion, which in a sense reduces the summation to simple addition.

Example. General recursion (1.3.2) can be used to define the function f such that for $n = 1, 2, \ldots, f(n)$ is the nth prime: $f(1) = 2$, $f(2) = 3$, $f(3) = 5$, $f(4) = 7$, $f(5) = 11$, and so on. On the empty function, g is defined as 2, and so $f(1) = 2$. Given j on $J(n) = \{1, 2, \ldots, n\} = I(n + 1)$, let $g(j)$