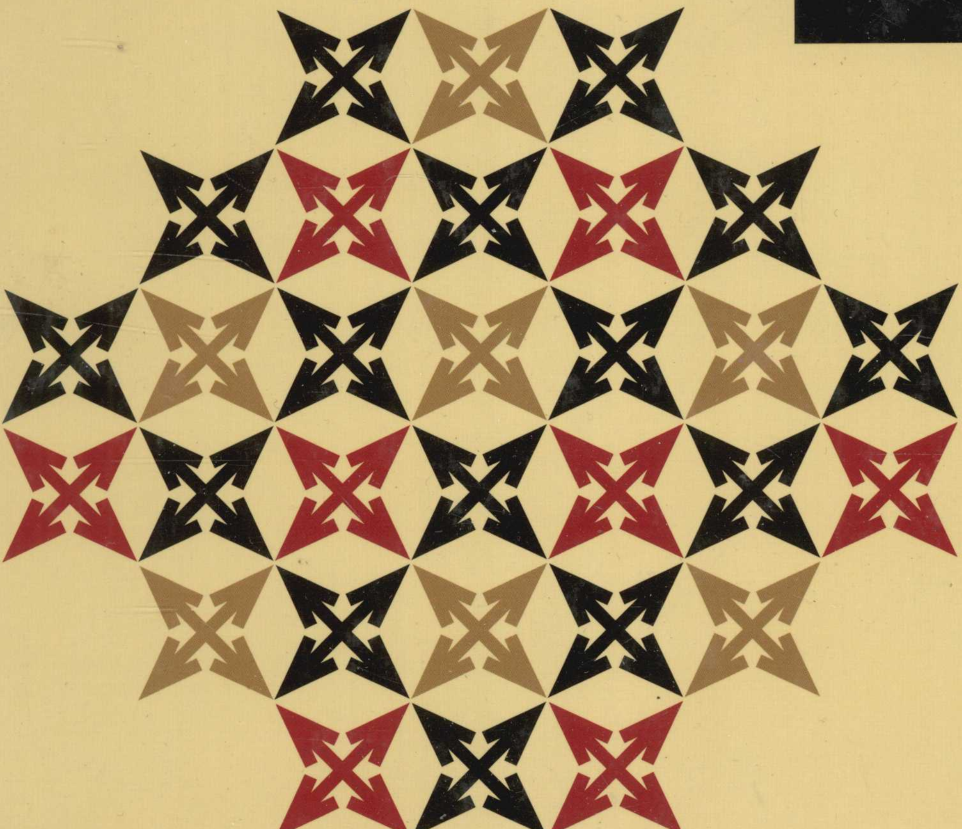
Advanced Series on Mathematical Psychology Vol. 2



Theories of Probability

An Examination of Logical and Qualitative Foundations

Louis Narens

Advanced Seri… …ychology Vol. 2

Theories of Probability

An Examination of Logical and Qualitative Foundations

Louis Narens
University of California, Irvine, USA

World Scientific
NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data
Narens, Louis.
Theories of probability : an examination of logical and qualitative foundations / Louis Narens.
p. cm. -- (Advanced series on mathematical psychology ; v. 2)
Includes bibliographical references and index.
ISBN-13: 978-981-270-801-4
ISBN-10: 981-270-801-4
1. Probabilities. I. Title.
QA273.4.N37 2007
519.2--dc22

2007018216

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

Printed in Singapore by B & JO Enterprise

Contents

Chapter 1

Elementary Concepts

1.1 Introduction

Traditional probability theory is founded on Kolmogorov's (1933) axiomatization of a probability function, which assumes probability is a σ-additive measure. This allows for the powerful and highly developed mathematics of measure theory to be immediately available as part of the theory. It is argued in this book, as well as in many places in the literature, that the measure theoretic foundation, while widely applicable, is overspecific for a general concept of probability. This book proposes two different approaches to a more general concept.

The first is qualitative. Kolmogorov's axiomatization assumes numbers (probabilities) have been assigned to events, and his axioms involve both properties of numbers and events. But where did the numbers come from? Some have tried to answer this by having probabilistic assignments be determined by some rule involving random processes. For example, in von Mises (1936) probabilities are limits of relative frequencies arising from random sequences. Obviously approaches based on randomness are limited to situations where assumptions about randomness are appropriate for the generation of the kind of uncertainty under consideration. It is unlikely, for example, that such assumptions apply to the kind of uncertain events encountered in everyday situations. The qualitative approach introduces numbers (probabilities) without making assumptions about randomness. It assumes that *some pairs* of events are comparable in terms of their likelihood of occurrence; that is, some pairs of events are comparable through the relation $\precsim$, where $A \precsim B$ stands for "A is less or equally likely to occur as B." Qualitative axioms are given in terms of events and the relation $\precsim$ that guarantee the existence of a function φ on events such that for the sure event, X, the null event, $\varnothing$, and all A and B in the domain of φ,

(*i*) φ is into the close interval $[0, 1]$ of the reals, $\varphi(X) = 1$, and $\varphi(\varnothing) = 0$,

(*ii*) if $A \cap B = \varnothing$, then $\varphi(A \cup B) = \varphi(A) + \varphi(B)$, and

(*iii*) if $A \precsim B$, then $\varphi(A) \leq \varphi(B)$.

This qualitative approach yields a more general theory than Kolmogorov's, and it applies to important classes of probabilistic situations for which Kolmogorov's axiomatization is overspecific. Additional qualitative axioms can be added so that φ satisfies Kolmogorov's axioms.

I view this book's axiomatic, qualitative approach as being essentially about the same kind of uncertainty covered by the Kolmogorov axiomatization. This kind of uncertainty is one dimensional in nature and is measurable through probability functions or a modest generalization of them. The second approach is about a different kind of uncertainty.

In the decision theory literature, many have suggested that the utility of a gamble involving uncertain events is not its expectation with respect to utility of outcomes, but a more complicated function involving utility of outcomes, subjective probabilities, *and other factors of uncertainty,* for example, knowledge or hypotheses about the processes giving rise to the uncertainty inherent in the events. I find it reasonable to suppose that uncertainty with such "other factors" give rise to a subjective belief function that does not necessarily have properties (*i*) and (*ii*) above of a Kolmogorov probability function. In the models presented in the book, the "other factors" impact belief in two different, but related, ways: (1) by distorting in a systematic manner a Kolmogorov probability function to produce a non-additive belief function (i.e., a belief function $\mathbb{B}$ such that $\mathbb{B}(A \cup B) \neq \mathbb{B}(A) + \mathbb{B}(B)$ for some disjoint events A and B); and (2) by changing the nature of the event space so that it is no longer properly modeled as a boolean algebra of events. Quantum mechanics employs (2) in its modeling of uncertainty. This book's implementation of (2) uses a different kind of event space than those found in quantum mechanics. However, as in quantum mechanics, the belief functions for these event spaces retain abstract properties similar to those of a Kolmogorov probability function. In particular, generalized versions of (*i*) and (*ii*) above are retained.

The book's two approaches can be read separately using the following plan:

Qualitative Foundation: Chapters 1 to 5 and 11.

New Event Space: Chapters 1 and 8 to 10.[1]

[1]One proof in Chapter 9 use concepts of Chapter 4.

Chapters 6 and 7 can be added to either plan. Chapter 7 (which depends on Chapter 6) provides a qualitative foundation for a descriptive theory of human probability judgments known as Support Theory. It employs a boolean event space and axiomatizes a belief function that has a more generalized form than a Kolmogorov probability function. A different foundation for Support Theory is given in Chapter 10. It is based on a non-boolean event space.

The book is not intended to be comprehensive. Much of its material comes from articles by the author. The good part of such a limitation is that it makes for a compact book with unified themes and methods of proof. The bad part is that many excellent results of the literature are left out.

The book is self-contained. The mathematics in it is at the level of upper division mathematics courses taught in the United States. However, many of its concepts are abstract and require mathematical sophistication and abstract thinking beyond that level, but not beyond what is usually achieved by researchers in applied mathematical disciplines like theoretical physics, theoretical computer science, philosophical logic, theoretical economics, etc.

1.2 Preliminary Conventions and Definitions

Convention 1.1 Throughout the book, the following notation, conventions and definitions are observed:

$\mathbb{R}$ denotes the set of reals, $\mathbb{R}^+$ the set of positive reals, $\mathbb{I}$ the integers, $\mathbb{I}^+$ the positive integers, and $*$ the operation of function composition. Usual set-theoretic notation is employed throughout, for example, $\cup$, $\cap$, $-$, and $\in$ are respectively, set-theoretic intersection, union, difference, and membership. $\subseteq$ is the subset relation, and $\subset$ is the proper subset relation, $\varnothing$ is the empty set, and $\wp(A)$ is the power set of A, $\{B|B \subseteq A\}$. $\notin$ stands for "is not a member of" and $\not\subseteq$ for "is not a subset of." For nonempty sets $\mathcal{E}$, $\bigcup\mathcal{E}$ and $\bigcap\mathcal{E}$ have the following definitions:

$$\bigcup\mathcal{E} = \{x|x \in E \text{ for some } E \text{ in } \mathcal{E}\} \quad \text{and} \quad \bigcap\mathcal{E} = \{x|x \in E \text{ for all } E \text{ in } \mathcal{E}\}\,.$$

"iff" stands for "if and only if." □

Definition 1.1 Let X be a set. Then X is said to be *denumerable* if and only if there exists a one-to-one function from $\mathbb{I}^+$ onto X. X is said to be *countable* if and only if X is denumerable or X is finite. □

Definition 1.2 Let X be a nonempty set and $\precsim$ be a binary relation on X. Then $\precsim$ is said to be:

Reflexive if and only if for all x in X, $x \precsim x$.

Transitive if and only if for all x, y, and z in X, if $x \precsim y$ and $y \precsim z$ then $x \precsim z$.

Symmetric if and only if for all x and y in X, if $x \precsim y$ then $y \precsim x$.

Connected if and only if for all x and y in X, either $x \precsim y$ or $y \precsim x$.

Antisymmetric if and only if for all x and y in X, if $x \precsim y$ and $y \precsim x$, then $x = y$.

The binary relations $\prec$, $\succsim$, $\succ$, and $\sim$ are defined in terms of $\precsim$ as follows: For all x and y in X,

$x \prec y$ if and only if $x \precsim y$ and not $y \precsim x$.

$x \succsim y$ if and only if $y \precsim x$.

$x \succ y$ if and only if $y \prec x$.

$x \sim y$ if and only if $x \precsim y$ and $y \precsim x$. □

Definition 1.3 Let $\precsim$ be a binary relation on the nonempty set X. Then $\precsim$ is said to be a:

Partial ordering on X if and only if X is a nonempty set and $\precsim$ is a reflexive, transitive, and antisymmetric relation on X.

Weak ordering if and only if $\precsim$ is transitive and connected.

Total ordering if and only if $\precsim$ is a weak ordering and is antisymmetric.

It is immediate that weak and total orderings are reflexive. By convention, partial orderings and total orderings $\precsim$ are often written as $\preceq$ to emphasize the fact that the relation $\sim$ defined in terms of $\precsim$ is the identity relation $=$. □

Definition 1.4 $\equiv$ is said to be an *equivalence relation* on X if and only if X is a nonempty set and $\equiv$ is a reflexive, transitive, and symmetric relation on X. □

It easily follows that if $\precsim$ is a weak odering on X, then $\sim$ is an equivalence relation on X.

The following definition is useful for distinguishing the usual total ordering of the real numbers from the usual total ordering of the rational numbers.

Definition 1.5 Suppose $\preceq$ is a total ordering on X. Then (A, B) is said to be a *Dedekind cut* of $\langle X, \preceq \rangle$ if and only if

(*i*) A and B are nonempty subsets of X,

(*ii*) $A \cup B = X$, and

(*iii*) for each x in A and each y in B, $x \prec y$.

Suppose (A, B) is a Dedekind cut of $\langle X, \preceq \rangle$, where $\preceq$ is a total ordering on X. Then c is said to be a *cut element* of (A, B) if and only if either

(1) c is in A and $x \preceq c \prec y$ for each x in A and each y in B, or

(2) c is in B and $x \prec c \preceq y$ for each x in A and each y in B.

$\langle X, \preceq \rangle$ is said to be *Dedekind complete* if and only if each Dedekind cut of $\langle X, \preceq \rangle$ has a cut element. □

The following theorem is well-known.

Theorem 1.1 *$\langle \mathbb{R}, \leq \rangle$ is Dedekind complete, and for each Dedekind cut (A, B) of $\langle \mathbb{R}, \leq \rangle$, if r and s are cut elements of (A, B), then $r = s$.* □

Definition 1.6 $A_1, \ldots, A_n$ is said to be a *partition of* X if and only if n is an integer ≥ 2, $A_1, \ldots, A_n$ are nonempty and pairwise disjoint and

$$A_1 \cup \cdots \cup A_n = X . \quad \square$$

Let $\mathcal{P} = A_1, \ldots, A_n$ be a partition of X. Note that by Definition 1.6, X is nonempty, $\varnothing$ is not an element of $\mathcal{P}$, and $\mathcal{P}$ has at least two elements.

A frequently employed principle of set theory is the Axiom of Choice. This axiom is often needed in mathematics to show the existence of various set-theoretic objects. In this book, a well-known equivalent of the Axiom of Choice, called "Zorn's Lemma," is sometimes used in proofs.

Definition 1.7 (Axiom of Choice) *For each nonempty set $\mathcal{Y}$ of nonempty sets there exists a function f on $\mathcal{Y}$ such that for each A in $\mathcal{Y}$, $f(A) \in A$.* □

Definition 1.8 Suppose $\mathcal{Y}$ is a nonempty set of sets. Then $A \in \mathcal{Y}$ is said to be *a maximal element of $\mathcal{Y}$ with respect to* $\subseteq$ if and only if for each B in $\mathcal{Y}$, if $A \subseteq B$ then $A = B$. □

Definition 1.9 $\mathcal{Y}$ is said to be a *chain* if and only if $\mathcal{Y}$ is a nonempty set of sets and for all A and B in $\mathcal{Y}$, either $A \subseteq B$ or $B \subseteq A$. □

Definition 1.10 (Zorn's Lemma) *Suppose $\mathcal{Y}$ is a nonempty set of sets such that for each subset $\mathcal{Z}$ of $\mathcal{Y}$, if $\mathcal{Z}$ is a chain then $\bigcup \mathcal{Z}$ is in $\mathcal{Y}$. Then $\mathcal{Y}$ has a maximal element with respect to $\subseteq$.* □

Definition 1.11 $\mathcal{F}$ is said to be a *ratio scale* family of functions if and only if $\mathcal{F}$ is a nonempty set of functions from some nonempty set into $\mathbb{R}^+$ such that (*i*) rf is in $\mathcal{F}$ for each r in $\mathbb{R}^+$ and each f in $\mathcal{F}$, and (*ii*) for all g and h in $\mathcal{F}$, there exists s in $\mathbb{R}^+$ such that $g = sh$. □

Convention 1.2 In Definition 1.11, "ratio scale" is defined for a family of functions that are into $\mathbb{R}^+$. Occasionally, this concept of "ratio scale" needs to be expanded to include cases where the elements of $\mathcal{F}$ are into $\mathbb{R}^+ \cup \{0\}$ while satisfying the rest of Definition 1.11. The expanded concept is also called a "ratio scale." When the context does not make clear which concept of "ratio scale" is involved, the concept in Definition 1.11 should be used. □

Definition 1.12 Then $\mathcal{F}$ is said to be an *interval scale* family of functions if and only if $\mathcal{F}$ is a nonempty set of functions from some nonempty set into $\mathbb{R}$ such that (*i*) $rf + s$ is in $\mathcal{F}$ for each r in $\mathbb{R}^+$, each s in $\mathbb{R}$, and each f in $\mathcal{F}$, and (*ii*) for all g and h in $\mathcal{F}$, there exist q in $\mathbb{R}^+$ and t in $\mathbb{R}$ such that $g = qh + t$. □

Convention 1.3 The notation (a, b) will often stand for the ordered pair of elements a and b, and in general $(a_1, \ldots, a_n)$ will stand for the ordered n-tuple of elements $a_1, \ldots, a_n$. The notation $\langle a_1, \ldots, a_n \rangle$ will also be used to stand for the ordered n-tuple of elements $a_1, \ldots, a_n$. $\langle \cdots \rangle$ is usually used to describe *relational structures with finitely many primitives.* These structures have the form

$$\mathfrak{A} = \langle A, R_1, \ldots, R_m, a_1, \ldots, a_n \rangle ,$$

where A is a nonempty set, $R_1, \ldots, R_m$ are relations on A, and $a_1, \ldots, a_n$ are elements of A. $A, R_1, \ldots, R_m, a_1, \ldots, a_n$ are called the *primitives* of $\mathfrak{A}$. □

Definition 1.13 Let R be an n-ary relation and A be a set. Then the *restriction of R to A,* in symbols, $R \restriction A$, is

$$\{(a_1, \ldots, a_n) \,|\, a_1 \in A, \;\ldots,\; a_n \in A, \;\text{ and }\; R(a_1, \ldots, a_n)\} . \quad \square$$

Convention 1.4 The convention of mathematics is often employed of having the same symbol denote different relations when a structure and substructure are simultaneously considered, for example, $+$ denoting addition of positive integers in $\langle \mathbb{I}^+, + \rangle$ as well as addition of real numbers in $\langle \mathbb{R}, + \rangle$. □

Chapter 2

Kolmogorov Probability Theory

Since the 1930's, the probability calculus of Kolmogorov (1933, 1950) has become the standard theory of probability for mathematics and science. Many philosophers of science and statisticians consider it to be the foundation for a general, rational theory of belief involving uncertainty. The author and others have been critical of this view and consider it to be a theory of probability that is at best only rationally justifiable in certain narrow kinds of probabilistic situations, for example, continuous situations in physics. That is, we believe the Kolmogorov theory is overspecific for a general, rational theory of belief.

The Kolmogorov theory assumes the following definition and six axioms:

Definition 2.1 $\mathcal{A}$ is said to be a *boolean algebra of subsets* of X if and only if the following five conditions hold:

(1) X is a nonempty set and $\mathcal{A}$ is a set of subsets of X;

(2) X is in $\mathcal{A}$, and the empty set, $\varnothing$, is in $\mathcal{A}$;

(3) for all A and B, if A is in $\mathcal{A}$ and B is in $\mathcal{A}$, then $A \cap B$ is in $\mathcal{A}$;

(4) for all A and B, if A is in $\mathcal{A}$ and B is in $\mathcal{A}$, then $A \cup B$ is in $\mathcal{A}$; and

(5) for all A in $\mathcal{A}$, $X - A$ is in $\mathcal{A}$.

$\mathcal{A}$ is said to be a *boolean σ-algebra of subsets* if and only if $\mathcal{A}$ is a boolean algebra of subsets such that if $A_i \in \mathcal{A}$ for each $i \in \mathbb{I}^+$, then $\bigcup_{i=1}^{\infty} A_i$ is in $\mathcal{A}$.

Suppose $\mathcal{A}$ is a boolean algebra of sets. Then $\mathcal{B}$ is said to be a *subalgebra* of $\mathcal{A}$ if and only if $\mathcal{B} \subseteq \mathcal{A}$ and $\mathcal{B}$ is a boolean algebra of sets. □

The following six axioms summarize Kolmogorov's axiomatic treatment of probability.

Axiom 2.1 *Uncertainty is captured by a unique function* $\mathbb{P}$. □

Axiom 2.2 *The domain of* $\mathbb{P}$ *is a boolean* σ*-algebra of subsets (Definition 2.1).* □

Axiom 2.3 *The codomain of* $\mathbb{P}$ *is a subset of the closed interval of real numbers* $[0,1]$. □

Axiom 2.4 $\mathbb{P}(\varnothing) = 0$. □

Axiom 2.5 $\mathbb{P}(X) = 1$. □

Axiom 2.6 (σ-additivity) *If* A_i, $i \in \mathbb{I}^+$, *is a sequence of pairwise disjoint sets, then*

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)\,. \quad \square$$

Definition 2.2 A function $\mathbb{P}$ satisfying Axioms 2.2 to 2.6 is called a *σ-additive probability function (on $\mathcal{A}$).* □

Except for Axiom 2.4, $\mathbb{P}(\varnothing) = 0$, it will be argued at various places in this book that each of the other five Kolmogorov axioms are overspecific. Axiom 2.6, σ-additivity, is generally singled out in the literature as being overspecific, and it is often suggested that to achieve a more general theory of probability, Axioms 2.2 and 2.6 should be replaced by Axioms 2.7 and 2.8 below.

Axiom 2.7 *The domain of* $\mathbb{P}$ *is a boolean algebra of subsets (Definition 2.1).* □

Axiom 2.8 (finite additivity) *For all* A *and* B *in* $\mathcal{A}$, *if* $A \cap B = \varnothing$, *then*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)\,. \quad \square$$

However, it is argued in this book that Axioms 2.7 and 2.8 are still overspecific.

Definition 2.3 A function $\mathbb{P}$ satisfying Axioms 2.3, 2.4, 2.5, 2.7, and 2.8 is called a *finitely additive probability function (on $\mathcal{A}$).* □

This book emphasizes the more general situation of finitely additive probability functions instead of σ-additive probability functions. By convention, the term, "the Kolmogorov theory," applies to both types of functions.

Convention 2.1 By convention, when the term "probability function" is used without the prefixes "finitely additive" or "σ-additive", it refers to a finitely additive probability function. When σ-additivity is needed, the prefix "σ-additive" is added. □

In the Kolmogorov theory the important probabilistic concepts of conditional probability and independence are defined in terms of $\mathbb{P}$:

Definition 2.4 For all A and B in $\mathcal{A}$ such that $\mathbb{P}(B) \neq 0$, the *conditional probability of A given B*, in symbols, $\mathbb{P}(A|B)$, is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} . \quad \square$$

Definition 2.5 For all A and B in $\mathcal{A}$, A and B are said to be *independent,* in symbols, $A \perp B$, if and only if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) . \quad \square$$

Probability theory is enormously applicable. Part of the reason is due to its rich mathematical calculus for manipulating probabilistic quantities. The richness comes from the following correspondence: addition corresponds to forming disjoint unions of events, multiplication to the intersection of independent events, and division to conditioning one event on another.

Many alternatives to the Kolmogorov theory have been proposed in the literature. Some are generalizations; others are about a different kind of "probability." With few exceptions, the numerical assignments of the alternatives form very weak calculi of quantities. In contrast, this book presents alternatives that have calculi that rival the finitely additive version of the Kolmogorov theory in terms of mathematical richness, and in some cases exceed it.

The main competitor in the literature to Kolmogorov's (1933) theory has been the relative frequency approach of Richard von Mises (1936), where probabilities are defined as limits of sequences of relative frequencies of random events. Although many textbooks give limiting relative frequencies as the definition of probability, its rigorous development is almost never attempted in those books, which in addition often fail to mention that probability functions that are produced in this manner are finitely additive, and not σ-additive.[1]

[1]Descriptions and critical evaluations of prominent approaches to probability theory can be found in Terrence Fine's excellent book, *Theories of Probability* (Fine, 1973).

This book pursues very different foundational approaches to probability theory than those of Kolmogorov and von Mises. One is based on a strategy developed by various behavioral and economic scientists and philosophers. It assumes an ordering, $\precsim$, on a set of events $\mathcal{E}$. "$A \precsim B$" is usually read as "the event A is less or equally likely to occur than the event B." The measurement problem for this kind of situation is showing $\langle \mathcal{E}, \precsim \rangle$ has a *probability representation,* that is, showing the existence of a finitely additive probability function $\mathbb{P}$ on $\mathcal{E}$ such that

$$\text{if } A \precsim B \text{ then } \mathbb{P}(A) \leq \mathbb{P}(B)\,. \tag{2.1}$$

When Equation 2.1 holds, it is often said that "$\mathbb{P}$ represents $\precsim$." At this level of analysis, the qualitative theory is more general than the finitely additive version of the Kolmogorov theory, because it does not necessarily produce a unique probability function for representing $\precsim$. Nevertheless, as is shown in Chapter 4, it is still a mathematical rich probability theory. Some researchers, including the author, consider the lack of uniqueness to be an important generalization of the Kolmogorov theory.

Many researchers of probability have developed theories to represent strengths of personal belief as Kolmogorov probabilities. They often provide arguments that claim the rational assignment of numbers to beliefs must obey the Kolmogorov axioms. I and others consider the Kolmogorov theory to be overspecific for many belief situations. We believe it needs to be extended. (An extension that encompasses additional rational phenomena is proposed in Chapter 9; extensions that encompass human judgments of probability are presented in Chapters 7 and 10.)

Chapter 3

Infinitesimals

3.1 Introduction

Probability theory is one of several interpretations of measure theory. In it the set X is interpreted as a sample space consisting of the set of possible states world, the boolean algebra $\mathcal{E}$ of subsets of X as a set of events, and the measure $\mathbb{P}$ as a function that assigns to each event A in $\mathcal{E}$ the probability that the actual state of the world is in A. Although each x in X is considered to have some chance of occurring, many probabilistic situations are modeled in a manner such that $\mathbb{P}(\{x\}) = 0$. In such cases, the possibility of the occurrence of an element of X is not distinguishable in terms of probability from the impossibility of the occurrence of the impossible event $\varnothing$. The inability to make this distinction rules out many natural concepts for dealing with events of probability 0. Conditioning on events of probability 0 is an example: Consider the case where $\mathbb{P}$ arises from a uniform distribution on an infinite set X and x and y are elements of X. Then one would want $\mathbb{P}(\{x\} \mid \{x, y\}) = .5$. The obvious and natural way of extending probability theory to provide for this, and more generally for a more structured approach to events of probability 0, is to have the co-domain of $\mathbb{P}$ include infinitesimal quantities. As is shown in Chapter 4, such an inclusion not only provides a closer match to the intuitive concept of "chance of occurring," but also provides methods that often make the mathematics of probabilistic situations much easier to deal with—even when infinitesimals are not mentioned as part of the final theorems. It also provides a more encompassing theory: Results of Chapters 4 and 11 show that the inclusion of infinitesimals provide for sharper qualitative axiomatizations, a better fit with techniques of mathematical logic, and a better foundation for philosophical issues concerning probabilities.