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 $\nsubseteq$ for "is not a subset of." For nonempty sets $\mathcal{E}$, $\bigcup \mathcal{E}$ and $\bigcap \mathcal{E}$ have the following definitions:

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

"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. $\square$

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 $r = s$.* $\square$

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 $\mathcal{Y}$, $f(A) \in A$.* $\square$

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 $A = B$. $\square$

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 $B \subseteq A$. $\square$

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, $R \upharpoonright A$, 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 $\langle \mathbb{R}, + \rangle$. □