FUZZY SET THEORY

Applications in the Social Sciences

Michael Smithson
Jay Verkuilen

Series: Quantitative Applications in the Social Sciences

147

SAGE Publications

Quantitative Applications in the Social Sciences

A SAGE PUBLICATIONS SERIES

Quantitative Applications in the Social Sciences

A SAGE PUBLICATIONS SERIES

87. **Analytic Mapping and Geographic Databases** *Garson/Biggs*
88. **Working With Archival Data** *Elder/Pavalko/Clipp*
89. **Multiple Comparison Procedures** *Toothaker*
90. **Nonparametric Statistics** *Gibbons*
91. **Nonparametric Measures of Association** *Gibbons*
92. **Understanding Regression Assumptions** *Berry*
93. **Regression With Dummy Variables** *Hardy*
94. **Loglinear Models With Latent Variables** *Hagenaars*
95. **Bootstrapping** *Mooney/Duval*
96. **Maximum Likelihood Estimation** *Eliason*
97. **Ordinal Log-Linear Models** *Ishii-Kuntz*
98. **Random Factors in ANOVA** *Jackson/Brashers*
99. **Univariate Tests for Time Series Models** *Cromwell/Labys/Terraza*
100. **Multivariate Tests for Time Series Models** *Cromwell/Hannan/Labys/Terraza*
101. **Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models** *Liao*
102. **Typologies and Taxonomies** *Bailey*
103. **Data Analysis: An Introduction** *Lewis-Beck*
104. **Multiple Attribute Decision Making** *Yoon/Hwang*
105. **Causal Analysis With Panel Data** *Finkel*
106. **Applied Logistic Regression Analysis, 2nd Edition** *Menard*
107. **Chaos and Catastrophe Theories** *Brown*
108. **Basic Math for Social Scientists: Concepts** *Hagle*
109. **Basic Math for Social Scientists: Problems and Solutions** *Hagle*
110. **Calculus** *Iversen*
111. **Regression Models: Censored, Sample Selected, or Truncated Data** *Breen*
112. **Tree Models of Similarity and Association** *James E. Corter*
113. **Computational Modeling** *Taber/Timpone*
114. **LISREL Approaches to Interaction Effects in Multiple Regression** *Jaccard/Wan*
115. **Analyzing Repeated Surveys** *Firebaugh*
116. **Monte Carlo Simulation** *Mooney*
117. **Statistical Graphics for Univariate and Bivariate Data** *Jacoby*
118. **Interaction Effects in Factorial Analysis of Variance** *Jaccard*
119. **Odds Ratios in the Analysis of Contingency Tables** *Rudas*
120. **Statistical Graphics for Visualizing Multivariate Data** *Jacoby*
121. **Applied Correspondence Analysis** *Clausen*
122. **Game Theory Topics** *Fink/Gates/Humes*
123. **Social Choice: Theory and Research** *Johnson*
124. **Neural Networks** *Abdi/Valentin/Edelman*
125. **Relating Statistics and Experimental Design: An Introduction** *Levin*
126. **Latent Class Scaling Analysis** *Dayton*
127. **Sorting Data: Collection and Analysis** *Coxon*
128. **Analyzing Documentary Accounts** *Hodson*
129. **Effect Size for ANOVA Designs** *Cortina/Nouri*
130. **Nonparametric Simple Regression: Smoothing Scatterplots** *Fox*
131. **Multiple and Generalized Nonparametric Regression** *Fox*
132. **Logistic Regression: A Primer** *Pampel*
133. **Translating Questionnaires and Other Research Instruments: Problems and Solutions** *Behling/Law*
134. **Generalized Linear Models: A United Approach** *Gill*
135. **Interaction Effects in Logistic Regression** *Jaccard*
136. **Missing Data** *Allison*
137. **Spline Regression Models** *Marsh/Cormier*
138. **Logit and Probit: Ordered and Multinomial Models** *Borooah*
139. **Correlation: Parametric and Nonparametric Measures** *Chen/Popovich*
140. **Confidence Intervals** *Smithson*
141. **Internet Data Collection** *Best/Krueger*
142. **Probability Theory** *Rudas*
143. **Multilevel Modeling** *Luke*
144. **Polytomous Item Response Theory Models** *Ostini/Nering*
145. **An Introduction to Generalized Linear Models** *Dunteman/Ho*
146. **Logistic Regression Models for Ordinal Response Variables** *O'Connell*
147. **Fuzzy Set Theory: Applications in the Social Sciences** *Smithson/Verkuilen*

Series/Number 07–147

FUZZY SET THEORY

Applications in the Social Sciences

MICHAEL SMITHSON
The Australian National University

JAY VERKUILEN
University of Illinois at Urbana–Champaign

SAGE PUBLICATIONS
International Educational and Professional Publisher
Thousand Oaks London New Delhi

For information:

Sage Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

Sage Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

Sage Publications India Pvt. Ltd.
B-42, Panchsheel Enclave
Post Box 4109
New Delhi 110 017 India

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Smithson, Michael.
Fuzzy set theory: Applications in the social sciences / Michael Smithson, Jay Verkuilen.
p. cm.—(Quantitative applications in the social sciences ; no. 07–147)
Includes bibliographical references and index.
ISBN 0-7619-2986-X (pbk. : alk. paper)
1. Social sciences—Mathematics. 2. Fuzzy sets. I. Verkuilen, Jay. II. Title. III. Sage university papers series. Quantitative applications in the social sciences ; no. 07–147.
H61.25.S652 2006
300.1′5113223—dc22
2005028893

This book is printed on acid-free paper.

06 07 08 09 10 10 9 8 7 6 5 4 3 2 1

Acquisitions Editor:	Lisa Cuevas Shaw
Editorial Assistant:	Karen Gia Wong
Production Editor:	Melanie Birdsall
Copy Editor:	Liann Lech
Typesetter:	C&M Digitals (P) Ltd.
Indexer:	Ellen Slavitz

From Michael Smithson to Susan.

From Jay Verkuilen to Helen Jacob and the memory of Max Jacob.

CONTENTS

SERIES EDITOR'S INTRODUCTION

The idea of sets must have come easily to humans; one may find the beginning of the logic of sets in Aristotle, when he declared that a man and an ox were both "animals" and that footed, two-footed, winged, aquatic, and so on are differentiae of "animal" in the first chapter of his classical book on logic. Such logic was not expressed with mathematical precision until 1874, when Georg Cantor, born in Denmark of Russian parents, published a paper that gave mathematical formalism and rigor to the notion of sets and marked the birth of set theory, a branch of mathematics.

Set theory, or classical set theory for our consideration in the book, has dominated the teaching and learning of mathematics, and still is a staple in a typical high school curriculum of mathematics. According to the theory, the membership function, or m_x, of any object belonging to the set X takes on only two values, 0 or 1, and the mapping of the function is given by m_x: $X \to \{0, 1\}$. However, this simple representation of reality was changed forever when Lotfi Zadeh published his revolutionary paper on fuzzy sets in 1965, and the world has never been the same place. According to fuzzy set theory, the mapping of the function is denoted by m_x: $X \to [0, 1]$, allowing for values from the entire unit interval. Although the foundation of fuzzy logic can be arguably traced back to Plato, it is Zadeh's work that paved the foundation for further theoretical research, followed by applications first in computer science and engineering and later on in basic sciences, including the social sciences. The applications of fuzzy logic today are embodied by how bullet trains run and how washing machines and video cameras operate.

What can fuzzy set theory contribute to the social sciences? Vagueness is such a common thing in social science research, and fuzzy set theory provides us with a proper way to handle vagueness systematically and constructively, instead of simply sweeping it under the rug. Smithson and Verkuilen's *Fuzzy Set Theory* is a badly needed addition to our series. My predecessor, Michael S. Lewis-Beck, who initiated the project with the first author, who is a seasoned author in the subject, deserves much of the credit for making the book a reality. This book will serve as a valuable introduction to fuzzy set theory and will take the reader on an informative walk into the wonder world beyond the familiar territory governed by bivalent, crisp logic.

—*Tim F. Liao*
Series Editor

ACKNOWLEDGMENTS

We have been fortunate to receive very useful advice and feedback on drafts of this book. We are especially indebted to John Beale, Derek Bopping, David Budescu, Gianfranco Giuntoli, Beryl Hesketh, Bernd Heubeck, Lawrence Hubert, James Kuklinski, Gerardo Munck, Chantal Mveh, James Neill, Carol Nickerson, and Charles Ragin, University of Arizona, and Kenneth G. Manton, Duke University, for their valuable suggestions, criticism, error detection, data, and insights. Of course, we are solely responsible for any remaining errors or shortcomings in this work. This research was supported by a NIMH National Research Service Award, No. MH14257, to the University of Illinois, while the second author was a predoctoral trainee in the Quantitative Methods Program of the Department of Psychology, University of Illinois at Urbana–Champaign.

FUZZY SET THEORY

Applications in the Social Sciences

Michael Smithson
The Australian National University

Jay Verkuilen
University of Illinois at Urbana–Champaign

1. INTRODUCTION

Researchers in the social sciences have long observed that although human beings break up their world into categories, they often use categories with blurred edges and gradations of membership. This is also true of many concepts employed in the social sciences themselves. This book introduces fuzzy set theory, an extension of classical set theory first proposed by Lotfi Zadeh (1965) that provides a mathematical framework for handling categories that permit partial membership (or membership in degree).

Beyond the initial intuitive appeal of permitting cases to belong partly to a set, fuzzy set theory offers generalizations of set-theoretic concepts such as intersection and union. Thus, it brings categorical concepts into the dimensional realm. If economic poverty and psychological depression are thought to be matters of degree, then fuzzy set theory claims that we can still meaningfully address the issue of how much the sets of poor people and depressed people intersect. There are five reasons for adding fuzzy sets to the social science toolbox:

- They are able to handle vagueness systematically.
- Many constructs in the social sciences have both a categorical and a dimensional character. Even apparently categorical concepts often turn out to be a matter of degree.
- They are able to analyze multivariate relationships beyond conditional means and the general linear model, via generalizations of set-theoretic operations.

- They have theoretical fidelity. Theories frequently are expressed in logical or set-wise terms, but most statistical models for continuous variables are not.
- Fuzzy set theory combines set-wise thinking and continuous variables in a rigorous fashion.

Sufficiently many applications of fuzzy set theory have accumulated during the four decades since its introduction to suggest that a book on fuzzy sets would be timely. In psychology, for instance, fuzzy set-based theories of perception (e.g., Oden & Massaro, 1978, and sequelae) and memory (Massaro, Weldon, & Kitzis, 1991) have appeared, and fuzzy sets have been used to solve measurement problems and provide novel data analysis tools (e.g., Hesketh, Pryor, Gleitzman, & Hesketh, 1988; Parasuraman, Masalonis, & Hancock, 2000; Smithson, 1987; Wallsten, Budescu, Rappoport, Zwick, & Forsyth, 1986; Zwick, Budescu, & Wallsten, 1988; and the survey of fuzzy set applications in psychology presented in Smithson & Oden, 1999). Likewise, in sociology and political science, fuzzy sets have been advocated by Ragin (2000) as enabling what he terms "diversity-oriented" research and strengthening the connection between theory and data analysis. A recent special issue of *Sociological Methods & Research* included articles on methodological aspects of fuzzy set theory as well as empirical applications (Ragin & Pennings, 2005).

Accordingly, this book is oriented toward acquainting researchers in the social sciences with fuzzy sets and some methodological tools for applying them. Chapter 2 introduces the basic concepts in fuzzy set theory, including grade of membership, set-theoretic operations, fuzzy numbers, and fuzzy variables. Chapter 3 addresses the issues involved in assigning graded membership in a fuzzy set and surveys methods for constructing membership functions. Chapter 4 explores the univariate properties of a fuzzy set, namely its size or cardinality, appropriate probability distributions, and fuzziness. Bivariate relations among sets (intersection, union, and inclusion) are developed in Chapter 5. Finally, Chapter 6 introduces multiset relations and concepts, including compound set indexes, conditional membership functions, and multiple and partial intersection and inclusion. Throughout the book, we draw examples from a variety of disciplines in the social sciences, and wherever possible, we establish connections between the fuzzy set approach and more traditional data-analytic techniques.

Unlike most textbooks on fuzzy set theory, this book emphasizes the combination of fuzzy set concepts with fairly straightforward statistical techniques, many of which can be implemented using standard statistical packages. We believe this combination is essential. Where fuzzy sets have been applied in the social sciences, most researchers simply use the notion of a fuzzy set to invoke gradations of membership and overlapping categories. Some applications use explicit grades of membership, perhaps in conjunction with prototypes and measures of similarity to prototypes. Few use fuzzy intersection and union, fuzzy logic, or other kinds of fuzzy inference. A primary reason for not going beyond programmatic statements and rather unsophisticated uses of fuzzy set theory has been the lack of practical methods for combining fuzzy set concepts with statistical methods. This book takes that topic as its major focus and provides explicit guides for researchers who would like to harness fuzzy set concepts while being able to make statistical inferences and test their models.

Several aspects of fuzzy sets are *not* covered in this work. This is not because we think they are unimportant, but because doing justice to them would easily double the book's length. Rather, we have chosen to concentrate on aspects we believe will be most accessible and interesting to a broad readership of social scientists who have only heard of fuzzy set theory but do not know much about it. However, we single out two areas not covered here that are relevant to social scientists and provide a few citations of each kind of research. *Fuzzy logic* is a direct outgrowth of fuzzy set theory and underpins many applications in fuzzy inference and control systems, ranging from simple static database structures to complex dynamic systems. Bárdossy and Duckstein (1995) discuss a number of systems that have been used for regional planning. Seitz and colleagues have employed dynamic inference systems to model foreign policy decision making and organizational behavior (e.g., Seitz, 1994, and Seitz, Hulin, & Hanisch, 2001, respectively). Taber (1992) provides an elementary introduction to computational models, some of which use fuzzy reasoning. Fuzzy data reduction techniques, initially motivated by research on pattern recognition, also have been developed. Much of this work is in fuzzy clustering methods (cf. Smithson, 1987, Ch. 5 for a review of work in the 1970s and 1980s, and Steenkamp & Wedel's 1991 fuzzy clusterwise regression technique for an example of a more recent development). An example from another approach is Manton, Woodbury, and Tolley's (1994) "grade of membership" extension of latent class analysis, which permits partial membership in the latent classes. A computer program implementing their techniques, DSIGoM, is commercially available (Decision Systems Inc., 1998), and the GoM model has been used heavily in health studies and demography. Likewise, Goldstein, Rasbash, Browne, Woodhouse, and Poulain (2000) employ multilevel models to develop fuzzy sets for modeling household data structures where households change composition over time.

2. AN OVERVIEW OF FUZZY SET MATHEMATICS

In this chapter, we provide a nontechnical introduction to fuzzy set mathematics. Rather than focusing on mathematical details, we will concentrate on making the concepts as clear as possible. There are several useful technical introductions in engineering textbooks, the most comprehensive being Zimmerman (1993) and Klir and Yuan (1995). Readers interested in more information about the topics covered in this chapter should consult these texts. Fuzzy set theory is a generalization of set theory. Although set theory is the foundation of the modern approach to mathematics and would be familiar to anyone with knowledge of, say, game theory or probability theory, we cannot assume that everyone is familiar with it. Thus, we will first start with a very brief overview of set theory and operations on sets. Then we will proceed to consider fuzzy sets as a particular extension of standard "crisp" set theory. Although "fuzzy" often carries a pejorative connotation, the mathematics of fuzzy set theory is precise. Its purpose is to allow us to better model phenomena that exhibit a certain kind of uncertainty, degree-vagueness.

2.1 Set Theory

The books mentioned above have reasonable introductions to set theory. Any introductory text on probability theory, real analysis, graph theory, logic, mathematical statistics, or linear algebra should also contain an introduction. Classical set theory is a mathematical calculus for dealing with collections of objects and certain relationships among these objects. At its most basic, a *set* is simply a list of objects, such as $A = \{\text{a, b, c, d, e}\}$ or $B = \{\text{orange, lemon, lime, grapefruit, tangerine}\}$. But sets generally become interesting by being connected with a *rule* that determines membership or nonmembership in the set. For instance, Set A can also be specified as the rule "first five letters of the alphabet" and Set B can be specified by the rule "commonly available citrus fruits," provided, of course, that "common" can be given precise meaning. Clearly, in any use of sets for modeling empirical reality or for testing real data, the rule connecting objects to each other is of utmost importance and must be specified clearly. In a location where kumquats and mandarins are grown in abundance but limes and grapefruit are unknown, the rule "common citrus fruits" would not specify Set B.

There are four common operations on sets: union, intersection, negation, and inclusion, commonly denoted by the symbols $\cup$, $\cap$, ~, and $\subset$, respectively (different notation for these operators is sometimes used by other authors). With these operations, it is possible to piece together quite complicated sets.

Union and intersection are known as connectives because they create a new set from two (or more) other sets according to a specified procedure. *Union* glues two sets together and corresponds to "or" in the inclusive sense, often expressed in natural language as and/or. Using the sets above, $A \cup B =$ {a, b, c, d, e, orange, lemon, lime, grapefruit, tangerine}. *Intersection* is the overlap between two sets and corresponds to "and." The two sets above have no elements in common and so their intersection is empty. We write $A \cap B = \varnothing$, the symbol for the null or empty set, the set that has no elements at all.

Negation, corresponding to "not," creates the complement of the set, which contains all elements in the universal set that are not in the set. Its definition requires that we define our universe of discourse, represented by the "universal" set *U*. Without *U*, we cannot meaningfully find a complement, and it is quite unclear whether we can make substantively meaningful statements about the sets at all. Assume for the moment that for Set *A* defined above, $U =$ {all letters of the English alphabet}. Then $\sim A =$ {f, g, . . . , z}. Note that $A \cup \sim A = U$, which says in words, "everything that is *A*, and everything that is not *A*, is everything." Also note that $A \cap \sim A = \varnothing$: "Nothing is in both *A* and not *A* at the same time." This statement is known as the Law of the Excluded Middle; it plays an important role in understanding fuzzy set theory because fuzzy intersections do not generally obey the Law.

Inclusion concerns whether a set has elements in common with another set. Set *P* is included in another set, *Q*, if all elements in *P* also are in *Q*. In the case of *A* and *B*, it is clear that neither set includes the other. However, given Set $T =$ {a, b, . . . j}, $A \subset T$ is read as "*A* is contained in *T*" or "*T* includes *A*." As we shall see in Chapter 5, the asymmetry of inclusion is exceptionally useful for examining relationships between empirical cases that are quite different from the correlations typically employed by social scientists. Inclusion and intersection have a special relationship. When $P \subset Q$, then $P \cap Q = P$. When $P \subset Q$ and $Q \subset P$, then $P = Q$. See Table 2.1 for more on key set-theoretic operations.

2.2 Why Fuzzy Sets?

We define Set $V =$ {a, e, i, o, u}, the set of vowels. Logically, $C = \sim V$, the set of consonants, because a letter is either a consonant or a vowel.

TABLE 2.1
Key Set-Theoretic Operations

Operation	*Symbol*	*Notation*	*Verbal Translation*
Union	$\cup$	$A \cup B$	All elements in either A or B, or both
Intersection	$\cap$	$A \cap B$	Elements that are in both A and B only
Complement	~	$\sim A$	Elements in U that are not in A
Inclusion	$\subset$	$A \subset B$	All elements in A are also in B

However, we know that in English, the letter *y* is sometimes a vowel and sometimes a consonant. For example, in the word "my," *y* is a vowel, but in the word "yours," it is not. Does *y* belong in Set *V*, or does it belong in *C*, the set of consonants? The answer is unclear because *y* does not fit neatly into either *V* or *C*, but rather into both. This means, of course, that the rule separating vowels and consonants does not lead to a mutually exclusive classification of letters suggested by the dichotomy between vowels and consonants. The letter *y* violates the Law of the Excluded Middle that is assumed when we define $C = \sim V$.

It is difficult to think precisely about even this example familiar to children, but the problem resembles those faced every day in the process of constructing data sets and making inferences about objects in the data set. Classical set theory is often not adequate for dealing with uncertainty in the rule that assigns objects to sets. Mathematical objects generally can be defined precisely; empirical objects often cannot be so defined.

Fuzzy sets are designed to handle a particular kind of uncertainty—namely *degree-vagueness*—which results when we have a property that can be possessed by objects to varying degrees. Vagueness is easiest to see by referring to a classical paradox, the Sorites, an example of which we describe now. Consider a truckload of sand. Clearly, this constitutes a heap of sand. If we remove one grain of sand, the resulting pile is still a heap. Arguing by a possibly fallacious appeal to mathematical induction, we can remove another grain of sand and still have a heap. And so forth. Eventually, however, we have so little sand that no one would be willing to call whatever is left a heap. Thus, the definition of heap is not precise. It is subject to vagueness because nowhere in the process is there a point that divides things into two distinguishable states: heap and not-heap.

Many concepts in the social sciences contain essential vagueness in the sense that while we can define prototypical cases that fit the definition, it is not possible to provide crisp boundaries between sets. Consider poverty. Given a context, such as "single and lives in a college town in Midwestern U.S.A." (which provides an understanding of cost of living), we can define a poverty line relatively simply: “made less than $20,000 per year in 2003.” Classical set theory would lead us to declare that a Midwesterner who made *exactly* $20,000 per year is, therefore, *not* poor, even though everyone would recognize that adding one extra dollar of income makes no material difference in the life of the person in question. However, adding $10,000 a year to the person’s income would probably make her not poor. Thus, somewhere between an annual income of $20,000 and $30,000, the person would cease to be poor. Where exactly? Given a proposed boundary, we can almost always play the same game, noting that one more dollar does not make one go from being poor to not poor. Fuzzy set theory provides a mathematical toolbox for analyzing situations like this with precision, not via a definite cutoff, but by defining a *degree* of membership between the qualitatively different states of definitely poor and definitely not poor.

2.3 The Membership Function

A fuzzy set is based on a classical set, but it adds one more element: a numerical degree of membership of an object in the set, ranging from 0 to 1. Formally, the *membership function* m_A is a function over some space of objects Ξ mapping to the unit interval [0, 1], and the mapping is denoted by

$$m_A(x) : \Xi \rightarrow [0, 1].$$

This generates fuzzy set A. Note that a domain may refer to a “universal” set, but it also can be defined in terms of some mathematical region such as the real line or an interval representing the range of a scale.

The membership function is an index of “sethood” that measures the degree to which an object x is a member of a particular set. Unlike probability theory, degrees of membership do not have to add up to 1 across all objects, so many or few objects in the set could have high membership. However, an object’s membership in a set and the set’s complement must still sum to 1. The main difference between classical set theory and fuzzy set theory is that the latter admits to partial set membership. A classical or crisp set, then, is a fuzzy set that restricts its membership values to {0, 1}, the endpoints of the unit interval. Fuzzy set theory models vague phenomena by assigning any object a weight given by the value of the membership function, measuring the extent to which the rule “this object is in Set A” is judged to be true.

We use two simple examples to illustrate a number of points. First, we will construct a set of “common citrus fruits,” assigning membership values subjectively by picking numbers from {0, .25, .50, .75, 1} based on our own

TABLE 2.2
Common Citrus Fruits

Fruit	*Membership*
Navel orange	1.00
Lemon	1.00
Red grapefruit	0.75
Lime	0.75
Tangerine	0.50
Kumquat	0.00
Mandarin	0.25

"expertise." These assignments are displayed in Table 2.2. Assignment of membership is a difficult problem that requires a lot of thought, and a great deal more will be said about the task in Chapter 3. The procedure just adopted is not all that different from the coding of objects as conducted by many social scientists working outside the context of fuzzy sets, however.

A useful notation is to generalize the list of elements notation for standard sets to a list of ordered pairs: {(Navel Orange, 1), (Lemon, 1), (Red Grapefruit, .75), (Lime, .75), (Tangerine, .5), (Kumquat, 0), (Mandarin, .25)}. The list of ordered pairs notation is compact and useful for relatively small sets.

Our second example illustrates a rule that takes the domain into membership. Typically, this is done when the domain is defined in terms of a quantitative construct. For instance, using the poverty example given above, we might decide to use a *linear filter* over income. In the definition below, membership in the set of poor people is 0 if annual income exceeds \$30,000, increases linearly for incomes ranging from \$30,000 down to \$20,000, and equals 1 for any income below \$20,000.

$$Poor(x) = \begin{cases} 0, & x > 30{,}000 \\ \dfrac{30{,}000 - x}{30{,}000 - 20{,}000}, & 20{,}000 \le x \le 30{,}000. \\ 1, & 0 \le x < 20{,}000 \end{cases}$$

This seems simple. However, one dilemma we must face immediately when constructing a fuzzy set (or indeed any set) is the definition of the universal set U. The observant reader may note that the fruit set defined above is confounded in that it includes at least two properties, "common" and "citrus." (This was intentional.) What constitutes the universe of discourse? Is it citrus fruits specifically, fruits in general, things found at a grocery store, or something else? A membership value takes on different meanings for different universal sets. In the case of $U = \{\text{citrus fruits}\}$, the kumquat's zero membership value indicates it is quite uncommon, although it is certainly a citrus fruit. If, however, $U = \{\text{fruits in general}\}$, many other cases would have zero membership by virtue of not being citrus fruits. Apples are certainly common, but they are not citrus fruits at all and thus fail on that criterion entirely. Even this seemingly trivial set is, in fact, rather complicated. If a reader walks away with nothing else, it should be a reminder that when constituting a population, clarity is essential.

2.4 Operations of Fuzzy Set Theory

Like classical set theory, fuzzy set theory includes operations union, intersection, complement, and inclusion, but also includes operations that have no classical counterpart, such as the modifiers concentration and dilation, and the connective fuzzy aggregation. In this section, all formulas are written with the assumption that only two sets are considered, but it is possible to extend all to three or more sets fairly easily by mathematical induction. To illustrate the fuzzy operations, we elaborate the fruits example. We have constructed four fuzzy sets over the universe of discourse "fruits"; this is not an exhaustive list. *Common* represents a subjective assessment of the degree of availability of fruits in an American supermarket. *Citrus* is true if the fruit in question is classified botanically as a citrus. *Rose* is true if the fruit in question is classified botanically as a rose. Finally, *Sour* represents the subjective degree of sour taste. *Citrus* and *Rose* are crisp sets because all membership values are either 0 or 1 (see Table 2.3).

Membership in the *fuzzy union* is defined as the maximum degree of membership in the sets. Membership in the union $X \cup Y$ may be written

$$m_{X \cup Y} = \max(m_X, m_Y).$$

Thus, the membership of an orange in the set *Common* $\cup$*Citrus* would be max(1.00, 1.00) = 1.00, and its membership in *Rose* $\cup$*Sour* would be max(.00, .25) = .25. Membership in the *fuzzy intersection* is defined as the minimum degree of membership in the sets, that is,

$$m_{X \cap Y} = \min(m_X, m_Y).$$

Thus, an orange's membership in the set "common and sour fruit" would be min(1.00, .25) = .25. The *fuzzy complement* is defined as $m_{\sim X} = 1 - m_X$. Thus an orange's membership in the set ~*Sour* is 1 – .25 = .75.

TABLE 2.3
Fruits Example With Membership of Two Derived Sets

Fruit	*Common*	*Citrus*	*Rose*	*Sour*	*Citrus∪Rose*	*Common∩Sour*
Navel orange	1.00	1.00	0.00	0.25	1.00	0.25
Lemon	1.00	1.00	0.00	1.00	1.00	1.00
Red grapefruit	0.75	1.00	0.00	0.75	1.00	0.75
Lime	0.75	1.00	0.00	0.75	1.00	0.75
Tangerine	0.50	1.00	0.00	0.25	1.00	0.25
Kumquat	0.00	1.00	0.00	0.00	1.00	0.00
Mandarin	0.25	1.00	0.00	0.00	1.00	0.00
Delicious apple	1.00	0.00	1.00	0.00	1.00	0.00
Star fruit	0.00	0.00	0.00	0.25	0.00	0.00
Banana	1.00	0.00	0.00	0.00	0.00	0.00
Red raspberry	0.75	0.00	1.00	0.75	1.00	0.75
Bing cherry	0.25	0.00	1.00	0.25	1.00	0.25
Strawberry	0.75	0.00	1.00	0.00	1.00	0.00
Coconut	0.50	0.00	0.00	0.00	0.00	0.00
Pineapple	0.50	0.00	0.00	0.50	0.00	0.50
Green grape	1.00	0.00	0.00	0.50	0.00	0.50

Unless otherwise noted, we use the max and min operators throughout this book for fuzzy union and fuzzy intersection, respectively. However, it is important to note that these are not the only definitions of the union and intersection suited to fuzzy set theory. Smithson (1987, Chapter 1) discusses this issue extensively, although most of the other books cited also have useful discussions in their consideration of t-norms and co-norms. In some contexts, alternative definitions of the operators are better able to meet the needs of particular applications. For example, the *product operators* are $m_{X \cup Y} = m_X + m_Y - m_X m_Y$ and $m_{X \cap Y} = m_X m_Y$. These formulas are, in fact, the same as the rules for compound independent events in probability theory. Unlike the max-min operations, they are continuous; changes in membership in one set are always reflected in the membership of the union or intersection. By contrast, this is not true for the max-min operators. This more continuous change may better reflect the underlying conceptual space.

Despite discontinuity, max-min operations remain the “industry standard.” They are very easy to calculate, which is a virtue in some cases. Perhaps most importantly, they are relatively resistant to perturbations in the input membership values—which are often due more to measurement error than real variation—and demand only ordinal measurement. The multiplicity of operators is both a strength and a weakness of fuzzy set theory. As a strength, many different operators provide options for modeling different concepts. As a weakness, there are many choices to make, and it is not always clear which alternative is best. Of course, these operators all reduce to the classical ones when membership is restricted to just 0 and 1.

It is possible to chain operators together, thereby constructing quite complicated sets. In fact, much of the power of fuzzy set theory comes from this, as it is possible to derive many interesting sets from chains of rules built up from simple operators. The orange's membership in the set ~*Common*∩*Sour* would be min(1 – 1.00, .25) = 0. Indeed, as the orange is nearly prototypical of a sweet, common fruit, it should make sense that the membership in the set is low.

Fuzzy inclusion is somewhat more complicated. We introduce the Classical Inclusion Ratio (CIR) here, deferring a more complete discussion until Chapter 5. For crisp sets, inclusion is an all-or-nothing matter. Either Set *A* is included in Set *B* or it is not, and all it takes is one element in *A* not in *B* for inclusion to fail. This is decidedly unfuzzy and furthermore does not make sense from a data analytic standpoint, where we would expect to see some errors from the general pattern simply due to chance. Because crisp sets are just fuzzy sets with no membership values on the interior of the unit interval, they also have membership functions. Therefore, inclusion can be translated into a statement about membership: For *B* to include *A*, objects in *A* must have membership no greater than objects in *B*. We can easily extend this to continuous membership. Thus, the CIR simply counts the number of such objects relative to the total number of objects. If there are *n* objects,

$$\mathrm{CIR}_{A \subset B} = \#(m_A \leq m_B)/n. \qquad [2.1]$$

Because this is a proportion, it is possible to use the standard statistics for proportions to form the basis of statistical tests about the CIR, which is one of its main selling points. Another useful benchmark for inclusion of *A* in *B* is to consider how similar $m_{A \cap B}$ is to m_A, which can easily be seen by plotting $m_{A \cap B}$ on m_A. If identical, then they should form a straight line with intercept 0 and slope 1. In the fruits example, $\mathrm{CIR}_{Sour \subset Common} = 15/16 = .9375$, indicating that *Sour* is fuzzily included in *Common*.

Recall that fuzzy sets do not obey the Law of the Excluded Middle. Consider the lime, which has .75 membership in *Sour* and therefore a membership in ~*Sour* of .25. The membership in *Sour*∩~*Sour* is min(.75, .25) = .25. Given that we are considering a situation of vagueness, this seems sensible. But genetic engineering aside, a plant cannot be both citrus and not citrus at the same time, and the membership of the lime in "citrus and not citrus" is min(1, 0) = 0, as it should be.

We mentioned three other operations on fuzzy sets that are important: concentration, dilation, and aggregation. None of these operators has

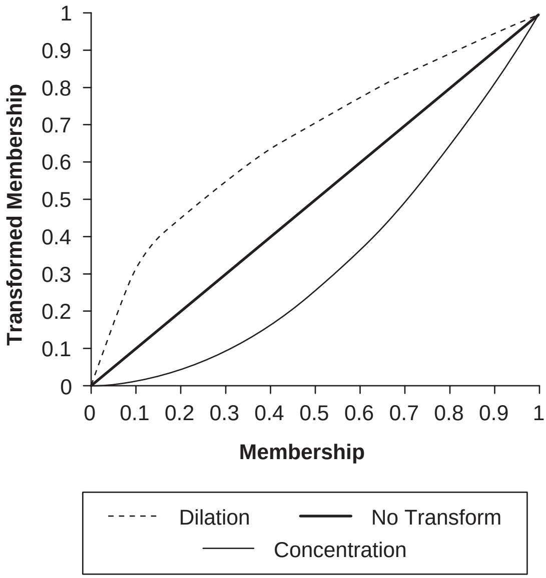

Figure 2.1 Concentration and Dilation

classical set parallels because all of them depend on membership values between 0 and 1. Concentration and dilation modify one set, similar to the complement, whereas aggregation is another connective between sets, similar to union and intersection. Concentration and dilation modify membership. Zadeh (1965) suggested that concentration corresponds to the phrase "very *X*," where *X* is the defining property, whereas dilation often is associated with the phrase "sort of *X*." The original *concentration operator* was $m_{Cx} = m_x^{\,2}$, and the original *dilation operator* was $m_{Dx} = m_x^{\,1/2}$. Generalization to using powers greater than one for concentration and powers less than one but greater than zero for dilation are straightforward.

The sense of these operators comes from the properties of power transformations of the unit interval: Power transformations in the unit interval map into the unit interval, which means that they can be interpreted as membership values. Concentration reduces all the values except 0 and 1 by squaring them, but the effect is weakest on those that are already small. Conversely, dilation increases all membership values except 0 and 1, but the effect is weakest on those that are already large. Figure 2.1 illustrates this,

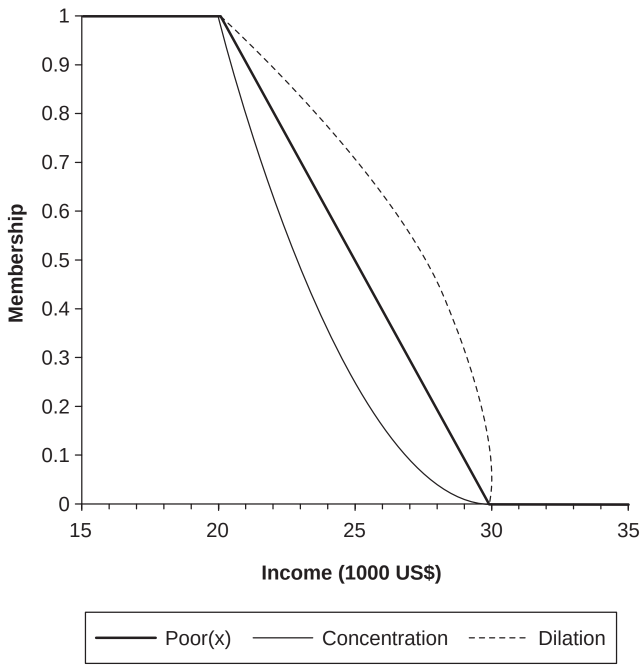

Figure 2.2 Concentration and Dilation for Membership Function *Poor*(x)

and Figure 2.2 shows concentration and dilation applied to the membership function *Poor*(x) given above.

The suitability of fuzzy set theory to model natural language usages of these terms, which are called linguistic hedges, has been questioned most forcefully by Lakoff (1973), particularly the use of dilation to model the hedge "sort of." Smithson (1987, Chapters 1–2) offers an extensive discussion drawing on the literatures in philosophy and cognitive science. However, we do not propose fuzzy set theory as a good model for natural language, but as a formal language for scientists operating in a domain of systematized logical reconstructions. The test of fuzzy set theory is whether it provides useful results. We discuss some further issues regarding transformations of fuzzy set memberships in Chapter 3 (Section 3.5 on sensitivity analysis) and present an example application of concentration and dilation in Chapter 6. One argument against their broad use is that they demand a high level of measurement, higher than most users will wish to assume.

The final fuzzy set operator we will discuss here is *fuzzy aggregation*, denoted by the symbol Γ (Gamma). Thus, the aggregation of two sets *X* and *Y* is denoted *X* Γ*Y*, and the aggregation of sets *X*, *Y*, and *Z* would be *X* Γ*Y* ΓZ. Classical sets have two connectives, unions and intersections, and these have extensions to fuzzy sets. As already discussed, membership in a union is determined by the maximum of memberships in the input sets, but membership in an intersection is determined by the minimum of memberships in the input sets. These are often described in terms of the strongest link/weakest link metaphor, because membership in the union is determined by the strongest link in the chain whereas membership in the intersection is determined by the weakest link in the chain. In this sense, fuzzy union is fully compensatory in that low membership values in Sets *A*, *B*, and *C* are completely compensated by a high membership in Set *D*. Fuzzy intersection is not at all compensatory, in that high membership values in Sets *A*, *B*, and *C* cannot compensate at all for a low membership in Set *D*. Alternatively, fuzzy union models redundant causation, whereas fuzzy intersection models conjoint causation.

However, in many cases, theory says that several properties contribute to overall membership in the aggregate, but that low values in one property are not fully compensated for by high values in another, invalidating fuzzy union. In fact, this is very similar to the assumption often used in scale construction, where different components make up the whole by summing together. There are many different aggregation operators, but we will discuss two simple ones. The first is the geometric mean of membership functions

$$m_{x\Gamma y} = \sqrt{m_x m_y}.$$

The geometric mean acts like an average for membership values near each other but like the intersection when one of the membership values is close to zero. The second is the arithmetic mean of the union and intersection

$$m_{x\Gamma y} = \frac{\max(m_x, m_y) + \min(m_x, m_y)}{2}.$$

In the two-set case, this is just the arithmetic mean of the membership values, but when three or more sets are considered, that is not necessarily true. More sophisticated aggregation operators—of which the ones listed above are special cases—are discussed in great detail in Zimmerman (1993). If we wanted to aggregate *Common* and *Sour* via the geometric mean, for the orange we would get

$$\sqrt{1.00 \times 0.25} = .50,$$

whereas the arithmetic mean would be $(1.00 + .25)/2 = .675$. Interpretation of aggregations is, of course, a matter for substantive theory.

2.4.1 Level Sets

Level sets provide a useful connection between crisp sets and fuzzy sets. Starting with a fuzzy set X, we introduce a level parameter, $\lambda \in [0,1]$, and define a set as $Y_\lambda = \{x \in X \mid m_x \geq \lambda\}$. Translating into words, Y_λ is the classical (dichotomous) set made from fuzzy set X with elements that have membership greater than λ. For instance, if $X = \{(a,0), (b,.2), (c,.3), (d,.6), (e,.8), (f,1)\}$, then $Y_0 = \{a, b, c, d, e, f\}$, $Y_{.5} = \{d, e, f\}$, and $Y_1 = \{f\}$. Notice that if $\lambda > \theta$, then $Y_\lambda \subseteq Y_\theta$, as can be seen from the example as $Y_1 \subset Y_{.5} \subset Y_0$. One use of level sets is to generate contingency tables. Using the fruit example, cross-tabulating $Common_{.5}$ and $Sour_{.5}$ produces Table 2.4. We will use level sets extensively in Chapter 5.

TABLE 2.4
Cross-Classification Generated by Level Sets

	$Common_{.5} = 0$	$Common_{.5} = 1$
$Sour_{.5} = 1$	0	6
$Sour_{.5} = 0$	4	6

2.5 Fuzzy Numbers and Fuzzy Variables

Is “several” a number? It clearly specifies numerical information, but it is vague. Instead, it specifies a range of possible integers, some of which are more plausible than others as referring to “several.” “Several” does not really apply to just one integer, instead it is vague. Verbal quantifiers such as “several” can be made precise using fuzzy set theory by creating a *fuzzy number*. Smithson (1987) provided the following data taken from a survey of 23 undergraduates used to specify meaning for the term “several,” shown in Figure 2.3, which plots the mean membership value and ±2SE (truncated to lie in the unit interval). Subjects were asked to give a numerical membership rating for each number on a bounded response scale that was subsequently mapped into the unit interval. Clearly, the peak is over 6, and subjects judged that the numbers 5 to 8 are the most consistent with “several.” Similarly, Budescu and his colleagues (e.g., Budescu, Karelitz, & Wallsten, 2003) have used fuzzy numbers to elicit probability assessments. This process is important in risk analysis and other areas where data from

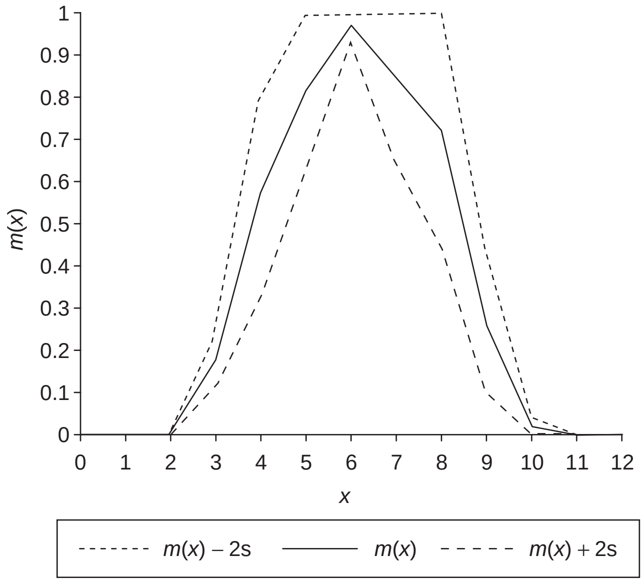

Figure 2.3 Fuzzy Number “Several”

judges are frequently stated in natural language, but it is desirable to have a numerical answer with an understood degree of uncertainty. By making more precise the typical understanding of a term like “highly unlikely,” fuzzy variables provide a means to translate qualitative language into quantitative statements.

The idea can be generalized to the notion of a *fuzzy variable.* For example, a survey question about sexual activity in teens might ask, “How many days in the past month did you have sex?” The response options might be {none, a few times, several times, many times}. The usual approach would be to break these categories into disjoint intervals, as in {0, [1,4], [5,8], [9,30]}. Many others are possible, and the results of the analysis could depend in nontrivial ways on the assignment of numerical values to the qualitative responses. Instead, we could use an approach similar to the one above for creating the fuzzy number “several” for the other responses, which would give us a better idea of exactly how many days the subjects had sex. Figure 2.4 shows a fuzzy variable constructed for this response set by mapping to the integers 0 to 30. Note that the intervals overlap, indicating degrees of uncertainty about exactly what the terms mean.

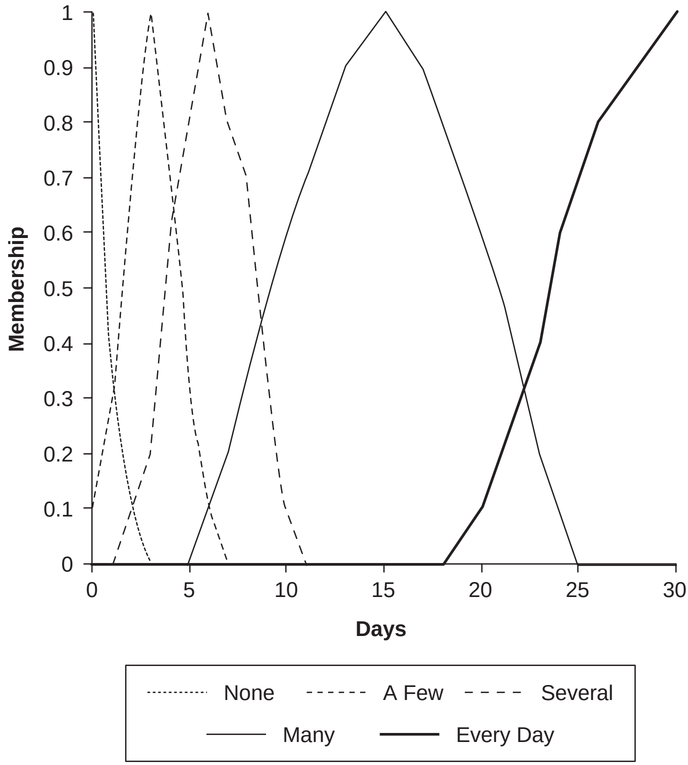

Figure 2.4 Fuzzy Variable "Number of Days in the Past Month"

2.6 Graphical Representations of Fuzzy Sets

Visualization is a key component of any data analysis, and data analysis with fuzzy sets is no exception. The first step in any analysis should be to graph the data. We will discuss graphs that consider only one fuzzy set at a time, focusing on it and its domain, and then discuss bivariate graphs, where we examine the membership of a domain of objects that are members of two fuzzy sets. The usual caveats and guidelines for creating good graphics apply. We refer readers to Jacoby (1997, 1998) or Cleveland (1993) for useful discussions. Because a membership function is a numerical value in the unit interval, we can graph it over its domain. Obviously, if the domain set has more structure—for example, it is numerical—the plot will have more structure, as in the plot for "several" (Figure 2.3) in the previous section. Often, it is sensible to examine how two fuzzy sets relate to each other. In Chapter 5, scatterplots will be used to reveal when one fuzzy set includes another.

3. MEASURING MEMBERSHIP

3.1 Introduction

As we noted in Chapter 2, applying fuzzy set theory requires that we

1. Precisely specify the domain X. What is the universe of objects under consideration? This could be a set that can be listed easily, such as all countries in the world, or Fortune 500 companies, or it could be a set such as all persons under the age of 100, where no such listing is practical.
2. Assign degrees of membership in fuzzy sets to objects in X. What properties do the fuzzy sets represent? What does a degree of membership mean?

This chapter will be concerned primarily with the second task, although specification of the first has important implications for the second, and so the separation of the two is difficult. We begin with the question of what a "degree of membership" means. Then we review the requirements of membership functions for fuzzy set theory operations. We discuss the measurement properties of the membership function and relate them to the literature on social science measurement. Finally, we discuss strategies for membership assignment and the construction of membership functions, including the largely ignored topic of assessing measurement errors in membership assignments. We use examples throughout to illustrate major points. Verkuilen (2005) contains a longer discussion of many of the points in this chapter, and also addresses some additional matters for which there was insufficient space.

We make one point up front: *Careful and clear conceptualization of the sets to be used is essential*. Unfortunately, as Adcock and Collier (2001) note, many social science concepts are essentially contestable in that they have no unique, correct definition. Here are three examples from economics, political science, and clinical psychology illustrating the persistence of such debates despite careful attention to conceptualization and measurement: (a) Ravallion (2003) identifies several different notions of poverty and inequality tapping different aspects of the background concept that comes out of ordinary discourse. Choosing different aspects leads to different measures. (b) The literature on democracy has often been confusing despite the careful attention that has been paid to theoretical and conceptual development, because definitions differ about what "real" democracy is (Munck & Verkuilen, 2002). (c) In the five decades since the publication of the first edition of the *Diagnostic and Statistical Manual*, massive research efforts and debates still have not resolved crucial diagnostic issues regarding disorders such as depression, anxiety, or schizophrenia. As the *DSM-IV* research coordinators recently remarked, "There might not in fact be one sentence within DSM-IV for which well-meaning clinicians, theorists, and researchers could not find some basis for fault" (Widiger & Clark, 2000, p. 946). Statistical techniques are no substitute for careful thinking about measurement issues, although some techniques are more susceptible to measurement decisions than others. Accordingly, our primary concerns in this chapter are the requirements for, and issues that must be resolved in, systematically assigning membership in fuzzy sets.

3.2 Methods for Constructing Membership Functions

What is a membership function? As mentioned in Chapter 2, formally it is a function for an attribute A over some space of objects (which may or may not be numerical) Ξ mapping to the unit interval [0, 1]:

$$m_A(x) : \Xi \to [0, 1]. \tag{3.1}$$

It is an index of "sethood" that measures the degree to which an object x with property A is a member of a particular defined set. It measures the fractional truth-value of the proposition "x is an element of A." A fuzzy set allows for partial membership, so the variable can have partial membership. For example, when scoring a test item, we could assign no credit, half credit, or full credit, representing membership values 0, .5, or 1 in the fuzzy set "correct answers for this item."

Because a membership function is only one number for a given object x, it can represent only one dimension at a time; more dimensions require more sets. In general, membership is latent, that is, not directly observable. It also embodies interpretations tied to a particular context. Although elapsed time certainly has bearing on the fuzzy set "long waits," the specification of this set depends on the domain. A long wait for a package sent by parcel post in the United States might be 3 weeks, whereas a package sent by overnight delivery is late if it arrives in 2 days. Thus, context should be specified as clearly as possible.

Degrees of membership also require an interpretive foundation. This foundation will, in turn, depend on the process of membership assignment. For example, it is not difficult to construct a collection of verbal phrases representing degrees of membership on whose order judges agree, although this task becomes more difficult with finer and more numerous gradations in membership. Somewhat more difficult is specifying or ascertaining what judges mean by a phrase like “sort of a member” or “neither out nor in,” and ensuring that they make membership assignments consistently. The most tenuous and vexatious link, however, is between the ordered membership phrases and numerical values. In many situations, the 0- and 1-valued membership assignments are fairly defensibly connected with phrases such as “not at all” and “fully” or “prototypical,” by reference to properties given in a theoretical literature or supplied by expert judges. Some phrases such as “halfway in and halfway out” or “neither in nor out” may arguably denote a value of 1/2. Finer-graded distinctions than this, however, seem arbitrary unless supported by specific operational criteria. Contrast this situation with the standard gambler or decision maker’s definition of a subjective probability. A probability rating p of an event means that the judge is willing to pay \$$p$ to receive \$1 if the event occurs and nothing otherwise. Given this operational definition, assigning a probability of .4 vs. .5 has clear implications for the decision maker in terms of the expected return.

Even when judges can be shown to be internally consistent, this consistency may not hold across judges, leading to the problem of calibration. Wallsten et al. (1986) established ratio scales for probability words within judges, but note that there was substantial between-subjects variability and thus they could not recommend averaging to generate a consensual scale because the resulting confidence intervals would be very wide. This meant that subjects’ membership values were not comparable, and thus they were not well-calibrated with an agreed standard meaning for the words being scaled.

Although the concepts of fuzziness and degree of membership are intuitively appealing, both caused confusion for some years after the publication of Zadeh’s classic 1965 paper, and only recently have matters become clearer. Indeed, it turns out that there are several distinct and viable meanings of degree of membership. As Bilgiç and Türkšen (2000, p. 195) point out, this state of affairs is “neither bizarre nor unsound.” Using typologies provided by Smithson (1987, pp. 78–79) and Bilgiç and Türkšen (2000), we group these interpretations into four clusters, each of which is suited for specific research purposes.

The first cluster consists of the *formalist interpretation,* which assigns membership functions solely in mathematical terms by mapping an underlying support variable into the membership scale. This variable can come from many different sources: subjective assessments by judges, indirect scaling/measurement models, or an objectively measured variable. Many fuzzy set theorists themselves are formalists insofar as they begin by assuming that we have agreed-upon locations for 0 and 1 or some other criterial membership value, and then define all intermediate membership values by a (usually) smooth function of the underlying support variable. Kaufmann's (1975) catalog of membership functions is an extreme example of this position. A more moderate exemplar is Kochen and Badre (1974), who begin with directly elicited degrees of belief and then derive their membership function from plausible and mathematically tractable criteria.

➢ Example 3.1: Human Development Index

The United Nations Development Program (UNDP) Human Development Index (HDI) was devised to make a broader, more conceptually rich measure of development than traditional national-level indicators, such as GDP per capita or energy expenditure per capita, both of which are common (UNDP, 1999). Starting from normative theory developed by Sen (1999) and others, the authors of the HDI disaggregated the top-level concept development into three components, Economic, Health, and Education. One aspect of the HDI and its relatives that is not immediately apparent is that they can be thought of as fuzzy sets.

To combine these HDI components, each needed to be put on a common scale. The unit interval was chosen. In addition, the authors believed that certain lower and upper goalposts represented important key points on the continuum of development from no development to full development. A country with a value above the upper goalpost could be considered fully developed on that component. Conversely, a country with a value below a lower goalpost could be considered fully undeveloped on that component. Variation between the goalposts was important, but outside it was not. These are exactly the sorts of conditions discussed in the poverty example from Chapter 2. Therefore, the basic strategy was to use a linear filter to assign membership. Table 3.1 shows the components, the indicators chosen to measure them, the goalposts, and the equation used to assign membership for each component.

Formalist approaches do not address the question of how any numerical scale for "degree of membership" or any other pretransformation construct

TABLE 3.1
HDI Example Component Membership Assignments

Component	*Indicator*	*Lower Goal Post*	*Upper Goal Post*	*Membership Between Goal Posts*
Economic: Decent standard of living?	GDP per capita ($PPP)	$100	$40,000	$econ = \frac{\log(GDPpc) - \log(100)}{\log(40,000) - \log(100)}$
Health: Long and healthy life?	Life expectancy at birth	25 years	85 years	$health = \frac{LE - 25}{85 - 25}$
Education: Knowledge?	Adult literacy rate and gross enrollment	0%	100%	$educ = \frac{2}{3}AL + \frac{1}{3}GE$

could be obtained. In the HDI example, linear filters were chosen, but they could just as readily have used some other smooth monotonic function. Instead, the life expectancy function could use a logistic formula:

$$m_H(x) = 1/[1 + e^{-a(x-b)}], \tag{3.2}$$

where a is the slope and b is the life expectancy corresponding to $m_H(x) = 1/2$. Setting $\alpha = 0.1$ and $b = 55$ yields the curve shown in Figure 3.1, with both the linear filter and logistic membership functions agreeing that $m_H(55) = 1/2$.

Note that the assignments by linear filter and logistic function are quite similar. Indeed, they correlate quite strongly; *any* reasonable monotonic functions will correlate strongly. Researchers taking a formalist approach therefore would need some empirical or theoretical criterion by which to choose a transformation from a class of strongly correlated functions. As Verkuilen (2005) notes, the major problem with the formalist transformation approach is that the number of plausible transformations is limitless, and consequences of a particular choice may not be immediately obvious. Nevertheless, transformation is frequently an important component of the assignment process, no matter how the base scores are obtained. For example, the Wallsten et al. (1986) study established a ratio scale for membership by axiomatic methods, thus rigorously establishing the zero point, but it required a subject-specific transformation to establish the point of full membership (and, implicitly, the neutral point).

The second cluster is the *probabilist interpretation,* which bases degree of membership on probability theory. The most direct grounding is simply

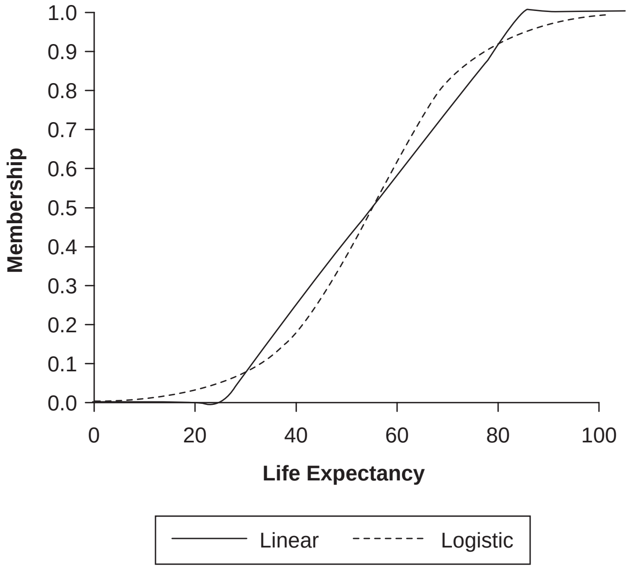

Figure 3.1 Linear Filter and Logistic Membership Functions

to assert that a degree of membership of Object *x* in Set *A* is the probability that *x* belongs in *A*. This probability may be a subjective rating provided by a single judge. The probability could also arise from a poll in two ways. The first identifies the degree of membership with the proportion of a sample who say that *x* belongs in *A*. Black's (1937) paper on vagueness was perhaps the first to suggest this approach, but others have followed suit (e.g., Hersh & Caramazza, 1976). The second way identifies degree of membership with the proportion of a sample who say that all *x* contained in a particular class belong to *A*.

This interpretation is sometimes called the *random set* view of fuzzy membership. Suppose the integer 4 has been assigned a degree of membership of .7 in the fuzzy set "several." The pollster interpretation of the integer 4's membership grade of .7 in "several" would be that 70% of people polled said that 4 is a member of "several." The random set interpretation of the grade .7 would be that 70% of the intervals over the integers that people provided when asked which integers were members of "several" included the integers from 4 to 7. Although many fuzzy set proponents have rejected the probabilist version of membership, the random set version has gained a number of adherents and has interpretive advantages in some cases. The constant inclusion path method for determining subset relations discussed in Chapter 5 has a natural interpretation in random sets.

It is, of course, possible to combine formalist and probabilist notions by devising a transformation that converts a probability distribution into a membership function. Cheli and Lemmi (1995) propose a fuzzy membership function for assessing poverty that is based on an existing population with a cumulative distribution function (CDF) for each relevant support variable (e.g., income). They identify 0-membership with a threshold on the original scale (x_0). The membership function is defined by

$$m_P(x) = \max[0, (F(x) - F(x_0))/(1 - F(x_0))], \qquad [3.3]$$

where $F(x)$ is the CDF at x. This formulation applies to any variable that has a CDF, and in fact is just a truncated CDF. Elaborations of this approach are easy to imagine.

Proponents of the probabilist interpretation (e.g., Hisdal, 1988; Thomas, 1995) argue that, like subjective probabilities, grades of membership reflect imperfect knowledge and/or errors in classification. The implication is that with perfect knowledge and error-free classification, degrees of membership would not exist. A counterargument is that judgments about degrees of membership need not arise from imperfect knowledge or error; in fact, they might be predicated on highly reliable expertise. For example, an artist who can distinguish a "warm" from a "cold" green, unlike the novice, knows that warm greens have a tinge of red and is not taking a gamble in that assessment. Whatever degree of membership the artist would assign a warm green to "red" would not be translatable into a "fair betting price" against $1 that the color in question really is red. Similarly, a student who receives partial credit on a test item usually has partial knowledge of the item content that is not especially like a gamble. Without new information, the student is likely to get the same parts right and the same parts wrong if the item is administered again.

The third cluster consists of those who regard the assignment of membership from a *decision-theoretic viewpoint*. In this approach, the degree of membership corresponds to the utility (payoff) of asserting that x is in A, which is related to the degree of truth in asserting that x belongs in A (Giles, 1988). An older version of the decision-theoretic approach that combines it with probability is signal detection theory (SDT), in which the expected utility for asserting that x is A versus ~A covaries with x's value or state on the underlying support variable(s). Both the utility and SDT frameworks treat labels such as "a few" or "several" as if they are chosen from a (usually finite) collection of labels. Thus, these frameworks apply most naturally in contexts where decisions must be made (e.g., whether to sound an alarm, or whether to describe the objects as "a few" vs. "several"). The implicit assumption here is not that we have limited knowledge, but limited choices instead. Like the formalists, the decision-theoretic view begs the question of where a utility scale comes from, but methods for the construction of utility scales exist.

The fourth cluster is composed of those who consider degree of membership as a problem in *axiomatic measurement theory* (Krantz, Luce, Suppes, & Tversky, 1971; for a more accessible treatment, see Michell, 1990). In fuzzy set theory, the earliest papers along these lines were Yager (1979) and Norwich and Türkšen (1982). A detailed examination of axiomatic measurement theory and its application to degree of membership is beyond the scope of this book, but Bilgiç and Türkšen (2000) provide a good technical overview. According to an axiomatic approach, we should be able to demonstrate that numerical membership assignments are quantitative in the sense that they behave just like fractional counts.

The key point in this perspective is that the quantitative structure of membership boils down to a set of qualitative axiomatic conditions that can and should be demonstrated empirically. Here are a few examples. The Wallsten et al. (1986) study is arguably the gold standard in fuzzy set theory because axiomatic methods were applied to show that the membership values elicited from judges satisfy the properties of a ratio scale. Verkuilen (2005) presents a simple example using the Bradley-Terry-Luce (BTL) model to convert judges' pairwise choices between various medical occupations based on their prestige into a membership function in the fuzzy set "prestigious medical occupations." The preference scale generated by the BTL model has an axiomatic basis: Provided the model fits, it satisfies the axioms of a strong utility scale, and it generates an interval scale for the objects. Recent work by Marchant (2004a, 2004b) illustrates the axiomatic approach by using comparisons and subjective ratio scaling to generate membership values, respectively, in a fuzzy set context. Finally, we should mention the work of Crowther, Batchelder, and Hu (1995), who examine the fuzzy logic model of perception (FLMP) of Massaro (1987) from the perspective of axiomatic measurement theory. In the FLMP, subjects provide direct ratings of membership in a variety of sets. This, in turn, is used to generate choice predictions. Crowther et al. demonstrate that the FLMP is equivalent to the BTL model, but one where subjects provide interval ratings rather than choices.

Connections between axiomatic measurement and psychometrics seem to be growing stronger at the time of writing. On one hand, as computational capacity catches up with the often extremely demanding requirements of testing measurement axioms, it becomes possible to provide a rational, probabilistic basis for testing the usually algebraic/deterministic measurement models in the presence of noise. On the other hand, axiomatic methods often provide sharper indications of model misfit than do the usual goodness-of-fit tests, which are dependent on the data. For instance, the well-known Rasch, or one-parameter logistic, model of IRT (which is, in turn, mathematically equivalent to the BTL model) also satisfies the axioms of conjoint measurement and thus can generate interval-scale information. Karabatsos has shown that inference about both subjects and items in the context of Rasch modeling can be improved by using axiomatic conditions (Karabatsos, 2001; Karabatsos & Ullrich, 2002). It seems clear that additional work in this area beyond the earlier, fairly simplistic studies would be useful.

So which of the four approaches—formalist, probabilist, decision-theoretic, or axiomatic—is the right one? We argue that none of these views is the sole correct one. If one's problem is similar to a decision-theoretic problem, then the tools of decision theory become relevant. Likewise, if the membership function we want amounts to a special kind of rating scale, then axiomatic measurement would be the best perspective. In short, we feel that a judicious choice of methods drawn from each view combined with a general skepticism is the healthiest attitude to take. It should also be noted that there are many opportunities for combining approaches.

3.3 Measurement Properties Required for Fuzzy Sets

Given the variety of interpretations of membership in fuzzy sets and the likelihood that membership functions may vary considerably in their measurement properties, it is worthwhile to consider how weak we would make our assumptions and still make use of fuzzy sets. Fewer and/or weaker assumptions are desirable, although there is always a tradeoff with statistical power and the clarity of results we can present on one hand, and the strength of assumptions we need to make on the other.

A "minimalist" membership assignment might consist of {0 = definite nonmember, possible member, 1 = definite member}. The case for intermediate degrees of membership (and therefore fuzziness) hinges on comparisons between objects (x and y, say) regarding whether x belongs to A more than y does. If such comparisons yield at least three objects for which the strict inequality $m_A(x) > m_A(y) > m_A(z)$ holds, then the case has been made for membership values between 0 and 1, and, thus, A being a fuzzy set.

Perhaps surprisingly, most fuzzy set concepts could be used effectively with a minimalist assignment. We would still be able to utilize fuzzy intersection and union, providing that min and max operators are used. The probabilistic viewpoint leads to a rejection of the min and max operators (Hisdal, 1988), and so does the decision-theoretic framework. An axiomatic measurement framework may have qualitative conditions on membership values (the axioms) that admit or even require min and max (e.g., Bollman-Sdorra, Wong, & Yao, 1993; Yager, 1979). However, as Bilgiç and Türkšen (2000) note, measurement that is strong enough to yield interval or ratio scales does not generally privilege the min-max aggregators over other aggregation operators (e.g., addition). Instead, the min-max pair emerges as the best aggregator for ordinal scales.

Negation is slightly more problematic. A probabilistic or decision-theoretic stance on fuzzy membership requires the standard definition of $m_{\sim A}(x) = 1 - m_A(x)$. Moreover, some measurement theorists, such as Bilgiç and Türkšen (2000), incorrectly assert that without a bounded ratio scale in the [0,1] interval, negation cannot be used. However, Smithson (1987, pp. 86–88) points out that an interval scale with an agreed-upon neutral point (q, such that $m_A(q) = 1/2$) is sufficient to justify a "mirror image" definition of negation that can be used even for ordinal membership scales. The mirror image of x is then $2q - x$, so $m_{\sim A}(2q - x) = m_A(x)$. The minimalist assignment is {0 = definite nonmember, possible member, 1 = definite member}, where negation works without assigning a numerical value for "possible member."

Comparisons between fuzzy sets with identical membership functions may be unproblematic (provided one is willing to assume or can establish comparability), but comparisons between fuzzy sets using different membership scales entail additional difficulties. Bollman-Sdorra et al. (1993) draw an important distinction between membership measurement and property ranking. Membership measurement hinges on comparisons between objects in the same set, regarding whether *x* belongs to *A* more than *y* does. In contrast, property ranking is based on comparisons between sets on the same object, that is, whether *x* belongs more to Set *A* than it does to Set *B*. If we cannot establish property ranking, then degrees of membership in Sets *A* and *B* are not directly comparable, regardless of the level of measurement each of the membership functions for *A* and *B* has.

If the same scales are used, then property ranking usually may be assumed simply by equating identical membership values with each other—although this assumption could be debatable in some circumstances. Otherwise (and more generally), we must specify a joint ordering of the membership levels of the sets being compared. A joint ordering might seem difficult if we compare the "apples" in one scale with the "oranges" in another. This issue is discussed in Chapters 4 and 5. The central point of this section, however, is that the fuzzy set framework compels researchers to make decisions about measurement properties. At the very least, we must decide what is in the set, what is excluded from it, and what is neither in nor out. If there is an underlying scale on which membership assignments are based, then scale-based criteria for full membership, partial membership, and nonmembership must be established.

3.4 Measurement Properties of Membership Functions

How do we determine the measurement level of a membership function? Researchers working with fuzzy sets have claimed levels anywhere from ordinal (unique up to monotonic transformation), to absolute, or unique (Bilgiç & Türkšen, 2000). Fuzzy set research shares this diversity with all of the social sciences, where debate on the measurement properties of variables continues (e.g., Michell, 1997). We lack the space for a review of those issues, but any readers who familiarize themselves with the debates will be able to handle measurement issues in using fuzzy sets. Jacoby (1991) is an excellent introduction to the data-theoretic perspective. Michell (1990) offers an eloquent introduction to and defense of the classical perspective.

There is one central point of data theory deserving emphasis: *No variable comes with a measurement level obviously attached.* Instead, the measurement level must be justified in the context of a specific problem. Indeed, many variables used in social or psychological research that have perfectly well-defined *physical* meanings in terms of ratio-level measurement (e.g., reaction times or electric potentials) may not have so obvious a connection with *behavioral* constructs of interest. For example, the percentage of income collected by the state as tax revenue—a frequently used measure of state capacity—is sometimes offered as an example of a measure that is "obviously ratio." But it is far from clear that the difference between 0% (Somalia) and 10% (Paraguay) is equivalent to 30% (Spain) and 40% (Italy) in terms of the concept of interest: state capacity (Lieberman, 2000). The first difference is a huge jump in terms of state capacity from none to possibly substantial, whereas the latter is a relatively small shift at the margin in terms of tax policy.

In short, it is incumbent upon the investigator to specify the relationship between the observed data and conceptual variables. Verkuilen (2005) notes that many relations between data and conceptual variables can be captured by the notions "more (less) is better," meaning that the relation between the data and concept is monotonic, and "just right," which means that the relation between the data is one of an ideal point, with membership declining from a peak. Furthermore, the notion of diminishing returns is implicit in most fuzzy set applications. Values near the extremes of membership (0 or 1) should rise or fall relatively slowly.

What measurement properties characterize a membership function, beyond the bare necessities described earlier? For some fuzzy set A, any collection of objects $\{x_1, x_2, \ldots, x_k\}$ can be ordered in terms of degree of membership in A, so that

$$m_A(x_1) \leq m_A(x_2) \leq \cdots \leq m_A(x_{k-1}) \leq m_A(x_k). \qquad [3.4]$$

As pointed out earlier, we must be able to identify at least two strict inequalities, that is, there must be x_h, x_i, and x_j such that $m_A(x_h) < m_A(x_i) < m_A(x_j)$.

Moreover, membership functions have endpoints (at least in principle) representing full nonmembership and full membership. In saying this, we have moved to a higher level of measurement, albeit one that is largely ignored by the standard textbook classification: ordinal with natural 0 and 1. In this case, we can assert that

$$\begin{aligned} 0 &\leq m_A(x_1) \leq \cdots < m_A(x_i) \leq \cdots \leq m_A(x_j) < \cdots \leq m_A(x_{k-1}) \\ &\leq m_A(x_k) \leq 1. \end{aligned} \qquad [3.5]$$

For every x_i, if we can determine that either $m_A(x_i) < 1/2$ or $m_A(x_i) \geq 1/2$, then we have still more structure than a simple ordinal scale. In Chapter 4, we will make use of a crude but effective membership scale comprising nonmembers, those closer to nonmembership, those closer to full membership, and full members. If we can identify an object $x_{neutral}$ such that $m_A(x_{neutral}) = 1/2$ and the fuzzy set is *normal* (i.e., has an object with 0 membership and another with 1 membership), we have even more structure. In short, starting with a weak ordering of objects in terms of membership, as we identify more and more objects with points in the membership scale, we constrain the possible membership values to a greater degree.

The move from membership functions with these more or less constrained quasi-ordinal scales to truly quantified membership functions requires stronger assumptions, special elicitation methods, or empirically based scaling techniques. None of these is beyond the purview of traditional measurement or scale construction approaches in the social sciences; the chief difference lies in identifying benchmark scale points for full membership, nonmembership, and/or neutrality. This is not to imply that obtaining truly quantified membership is easy, however, and often it is the case that it is more straightforward in a given application to use sensitivity analysis over the plausible range of the variables to show that conclusions are invariant under perturbation of the assigned membership values. Indeed, this practice is widespread in control systems theory applications of fuzzy set theory. One disadvantage of this approach is that it is unclear whether the scale generalizes to other applications, because no additional validation was done.

The most straightforward, practical route to membership functions uses the properties of an existing support variable in combination with endpoint identification. If we are mapping a single variable to membership by identifying the endpoints, we can interpolate intermediate membership values. In fact, this is precisely what the linear filter does, using a linear function to interpolate between endpoints. The HDI example is based entirely on this approach. As demonstrated with the logistic function, we could choose different interpolating functions, although linear filters frequently do quite well and have the virtue of simplicity. If we are willing to identify the neutral point or other interior reference points, we could use a piecewise interpolating function such as a piecewise linear or cubic spline, depending on the degree of smoothness required. The reference points provide more control over the shape of the interpolating function.

Depending on the data-gathering method, we may be able to extract even more information from the sort order itself. As Coombs (1951) suggests, it is highly desirable to collect data in a way that provides substantial redundant information, which allows for extensive cross-checking of validity and model fit. Rather than assuming ordinality, a method that allows one to check to see if responses are ordinal provides useful leverage, and the same principle holds for higher levels of measurement. Indeed, this is precisely what axiomatic measurement theory is about. Measurement models allow for assumption checking and often can "promote" the ordinal information to a higher level. Methods for doing this range from the widely used Rasch (1980) item-response theory (IRT) approach in aptitude and ability testing to recent and more flexible models suited to attitude measurement (e.g., Rossi, Gilula, & Allenby, 2001). Virtually all of these methods use probit or logit models to estimate "threshold" values on an interval-level latent continuum corresponding to the ordinal categories. If there is justification for designating certain thresholds as the nonmembership, neutral, and/or full membership cutoffs, then we may interpolate the remaining membership values using linear filters or appropriate splines. In sum, the entire bag of tricks of psychometrics and scaling is open, ranging from direct membership assignments by judges, to indirect scaling methods (e.g., IRT models or optimal scaling), to fundamental measurement.

3.5 Uncertainty Estimates in Membership Assignment

"What should we put in parentheses? It is a basic principle of sound econometrics that *every serious estimate deserves a standard error*" (Koenker & Hallock, 2001, emphasis in original). Despite the confusion about the relationship between fuzziness and probability and general disagreements about the nature of the membership function, relatively little attention has been paid to providing uncertainty estimates for membership functions in a practical sense. This is a major deficiency. As in any sort of measurement, there is no reason to believe that membership assignments are without error, and it is incumbent on researchers to assess the degree of error. Fuzzy set theory has managed to develop largely without a formal theory of error because in engineering, it is usually possible to do a lot of testing to show that the device being created works satisfactorily. Unfortunately, the concerns of an empirical scientist are not so easily addressed.

Many techniques are available, and we lack the space for an exhaustive treatment. Even if the formal machinery of statistical analysis is not applicable in all circumstances, *any* assignment method is open to uncertainty estimates of some form. An analysis involving fuzzy set techniques is incomplete without an assessment of uncertainty. Instead, we will concentrate on providing two detailed examples. The first uses sensitivity analysis in the case of a compound scale created from a single judge's ratings on four scales. The general strategy is useful because it can be applied in any situation, even direct ratings by one judge. The second considers using bootstrapping to provide pointwise error bounds for membership assignments when scores from an additive scale are mapped into the unit interval. However, we should note that many assignment techniques come with error assessments built in. If a multiple indicator measurement model were used, such as maximum likelihood factor analysis or polytomous IRT, it is possible to compute a confidence interval for assigned scores, which can in turn be transformed into a confidence interval for assigned membership. Even some direct elicitation methods, such as the "staircase method," generate their own uncertainty estimates (Tversky & Koehler, 1994).

3.5.1 Sensitivity Analysis

One way of assessing the precision of measurement of a membership function is to use a *sensitivity analysis*, an experiment designed to show the likely differences in the conclusions due to perturbations in the input (Saltelli, Tarantola, & Campolongo, 2000). This is particularly useful when there are no other sources of uncertainty estimates, for instance, coming from multiple measurements or a data-gathering strategy that included redundant information. If the membership values are the result of, say, one expert judge, sensitivity analysis can give an idea of how uncertain these assessments might be if different judges were used. We will focus the discussion on membership provided by an expert judge, but the technique is not limited to such situations. Examining the sensitivity of a parametric function used for membership assignment to different parameter values also is valuable.

The basic idea in sensitivity analysis is to examine the effect of varying the inputs on a particular calculated quantity by considering different scenarios of inputs. Assuming a given judge represents the baseline, one judge might be systematically lower than the baseline judge, whereas another judge might be systematically higher. In the application to be presented, we do not have these judges, but we will simulate them. In most membership assignment tasks, there are four principal options for judgmental bias from a given baseline: (a) biased systematically toward 0 (the tough grader), (b) biased systematically toward 1 (the easy grader), (c) biased toward the endpoints (the extreme grader), and (d) biased toward the neutral point (the vague grader).

Table 3.2 lists several families of transformations that implement these four types by systematically modifying the assigned memberships. By applying these transformations to the baseline membership assignments, we can then use standard descriptive statistics to generate pointwise error bars. There are caveats, however. Real judges are never totally consistent, so including random error may be desirable. Also, we do not mean to imply that the transformations listed in Table 3.2 are the only ones necessary. Transformations should be tailored to specific problems. More importantly, the baseline judge might not be very reliable and/or valid, so sensitivity analysis is *not* a substitute for real replications. However, in the absence of replications, sensitivity analysis does provide a method of obtaining uncertainty estimates.

➢ Example 3.2: Electoral Democracy Index Sensitivity Analysis

There have been numerous proposals for a democracy index (Munck & Verkuilen, 2002). One, the Electoral Democracy Index (EDI) (Munck & Verkuilen, 2003; UNDP, 2004), is based on fuzzy sets. The EDI is a compound index built from four components, each of which is assessed by an expert judge. The four components are Suffrage (S), Offices (O), Free (F), and Clean (C). S refers to the right of all adults to vote. O refers to the condition where the decision-making offices (executive and legislative) are filled by elections. F refers to the right of party competition and

TABLE 3.2

Some Transformations Useful in Sensitivity Analysis

Transformation	*Function*	*Parameter*	*Effect on m*	*Compared to Baseline/Notes*
0. Identity	$\text{identity}(m) = m$	—	• For $k = 1$; all other transforms reduce to the identity	Exactly the same.
1. Concentration	$\text{conc}(m) = m^k$	$k > 1$, typically 2	• Lowers all values in (0,1) • Endpoints unchanged	Systematically closer to 0. A harder grader.
2. Dilation	$\text{dil}(m) = \sqrt[k]{m}$	$k > 1$, typically 2	• Raises all values in (0,1) • Endpoints unchanged	Systematically closer to 1. An easier grader.
3. Contrast Intensification	$\text{cintens}(m) = \begin{cases} km^k, & m < .5 \\ .5, & m = .5 \\ 1 - k(1-m)^k, & m > .5 \end{cases}$	$k > 1$, typically 2	• Values > .5 go up • Values < .5 go down • Endpoints and .5 unchanged	Systematically closer to the endpoints. A more extreme grader.
4. Contrast Diffusion	$\text{cdiff}(m) = 2m - \text{cintens}(m)$	$k > 1$, typically 2	• Values > .5 go down • Values < .5 go up • Endpoints and .5 unchanged	Systematically closer to the neutral point, .5. A more vague grader.
5. Interval Squash	$m' = \text{squash}(m) = .5u + (1-u)m$	$0 < u < 1$, typically .05	• Resets membership to be in $[u/2, 1 - u/2]$, symmetric around .5	Used to move endpoints into (0,1) so other transformations change those values.
6. Interval Expand	$\text{expand}(m') = (m' - .5u)/(1-u)$	As squash(·)	• Inverts squash(·).	Used to restore original interval. Can move points outside [0,1] so extreme points should be clipped.

organization. Finally, *C* refers to the right to have the votes counted fairly and not have the voting process manipulated. Each of these indexes is assigned a score by one judge according to a set of rigorously defined coding rules. In Chapter 6, we will discuss how these components are combined to form the index.

To allow the endpoints to move under transformation, the given scores were "squashed" into the interval [.025, .975] by choosing $u = .05$ in the squash(·) transformation from Table 3.2; upon completion, the values were expanded back to [0,1], with any inadmissible values, such as 1.05 or –.1, clipped to fit. We use the transformations found in Table 3.2 to alter the baseline ratings on all four components given by the judge according to a 5^4 factorial design crossing transformations with components to simulate 625 different judges (one of which is the baseline judge). This design generates judges who are biased in different ways on different components. For example, a judge might be a tougher grader on *S*, be an easier grader on *O*, and agree with the baseline judge on *F* and *C*. We use order statistics to generate error bands. Figure 3.2 shows error bands using the 5% and 95% quantiles, with the actual EDI score for Brazil in years 1960, 1977, 1985, and 1990–2002. These bands encompass 90% of the simulated values. Note that because the error bands are based on order statistics, they are not always symmetric, unlike a confidence interval for the mean based on the standard error.

3.5.2 Test Inversion and Bootstrapping

In situations where degrees of membership are based on sample estimates (e.g., transformations of quantiles), confidence bands around the membership function can be estimated. Otherwise, bootstrapping may be employed instead (Efron & Tibshirani, 1994). Bootstrapping uses sampling with replacement from the original data set to generate as many replications of the data set as desired. Then, standard quantities such as quantiles or standard deviations for the statistic in question can be calculated using the usual procedures. Either way, the confidence interval approach is compatible with treating membership functions as random variables.

➢ Example 3.3: Confidence Bands for the Fuzzy Set "Violent Crime Prone" State

This example is based on the violent crime statistics in the data set "USArrests," one of the sample data sets included with the *R* statistical package. We lack the space to discuss the example in detail but encourage readers to

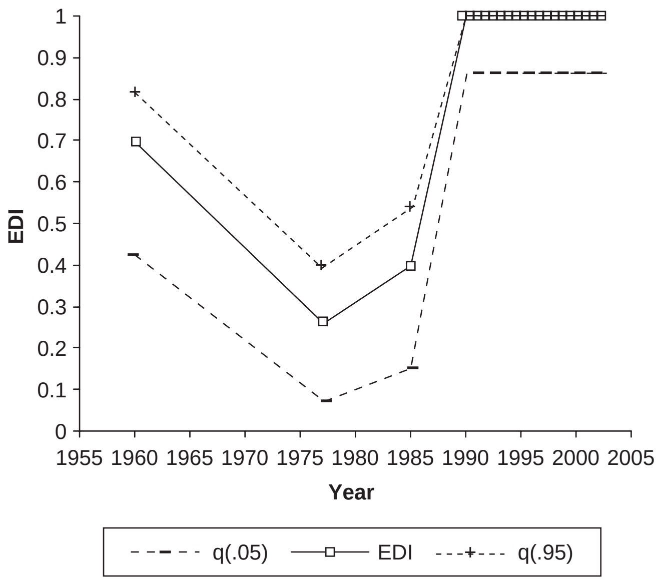

Figure 3.2 Error Bands by Perturbation for EDI Scores for Brazil

try their own coding. The relevant data comprise the number of arrests reported to the FBI for three violent crimes—murder, rape, and assault—for the 50 U.S. states in 1975. We want to create a fuzzy set "violent crime prone," denoted *VCP.*

First, we define *vcscore* to be the mean of the standard scores of the three crime rates, murder, rape, and assault. We justify this procedure on two grounds. First, although there are far fewer murders than rapes and fewer rapes than assaults, the severity of the crimes is such that this weighting of them makes sense. Standard scores adjust for the variations in scale between the components, unlike the FBI's unweighted Crime Index, which simply sums the arrests, ignoring the severity of the crime and thus swamping murders with assaults. Second, the intercorrelations are all .56 or higher, so summation is reasonable based on reliability theory conventions. The mean of *vcscore* is 0 by construction, and we further standardize it to have a standard deviation of 1. To create membership values, we use the CDF in Equation 3.3, as suggested by Cheli and Lemmi (1995). Note that this assignment generates a subnormal fuzzy set because there are no states with 0 membership, but then again there are no states without violent crime. (Choosing a lower cutoff would normalize the set.)

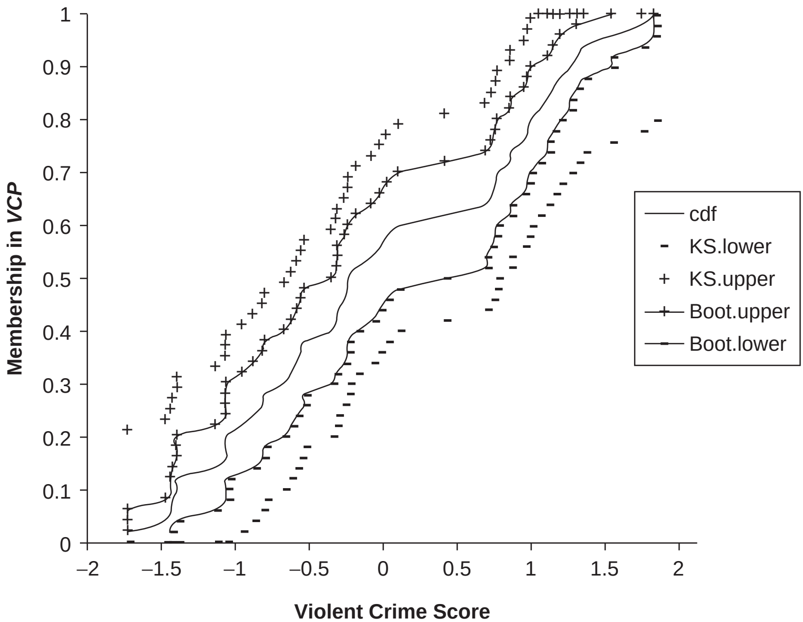

Figure 3.3 Error Bands by Inverting the Kolmogorov-Smirnov Test Versus Bootstrapping

To assess the uncertainty due to crime statistics, we employ two different techniques, one based on a classical statistical test and the other on bootstrapping. First, we inverted the Kolmogorov-Smirnov test to create confidence bands for the CDF (Conover, 1980). This test is known to lack power and therefore generates rather wide confidence bounds. Second, we generated 1,000 bootstrap samples from the 50 original values. Each sample was sorted from least to greatest; scores for $q_{.025}$ and $q_{.975}$ were taken for each of the 50 states' membership scores; and finally, a CDF was computed for each set of scores, $q_{.025}$ corresponding to the upper interval and $q_{.975}$ to the lower.

Figure 3.3 shows both sets of confidence bands. Note that, as expected, the bootstrap generates narrower intervals than does the inverted K-S test. One way to gauge the impact of sampling error on subsequent analyses would be to substitute the lower and upper membership estimates in the calculations and examine how conclusions change.

4. INTERNAL STRUCTURE AND PROPERTIES OF A FUZZY SET

The potential utility of analyzing internal category structure is indicated by developments over the past 30 years or so in several disciplines. From the 1970s onward, cognitive psychologists increasingly recognized that natural cognitive categories often have complex internal structures. Some psychologists have made similar claims about categories in psychological theories. Broughton (1990) claims that fuzzy sets are useful for organizing research and assessment in personality psychology, particularly for refining personality assessment instruments and improving abnormal diagnosis. Horowitz and Malle (1993) argue that depression should be thought of as a fuzzy concept, as do Waterhouse, Wing, and Fein (1989) regarding autism, and Burisch (1993) concerning burnout. Likewise, social scientists have increasingly emphasized the complexity as well as mutability of categories in their theoretical and research frameworks. Ragin (2000), for instance, presents extensive arguments for generalizing Lazarsfeld's (1937) concept of "property space" from its original crisp set to a fuzzy set version.

Because degree of membership in crisp sets is restricted to 0 and 1, little can be said about their internal structure. Usually, we are limited to evaluating the *cardinality* (the size) of the set, whether absolutely (i.e., counting the members) or relatively (e.g., in comparison with another set). However, intermediate degrees of membership enable several additional aspects of categorical structure to be analyzed and even quantified. These aspects, in turn, can be related to numerous useful concepts in the social sciences. In this chapter, we will begin with a brief overview of cardinality for fuzzy sets. We then consider what are appropriate probability distributions for fuzzy sets. Finally, we examine how to measure the fuzziness of a set.

4.1 Cardinality: The Size of a Fuzzy Set

For fuzzy sets, the concept of set size or cardinality is both richer and more problematic than it is for crisp sets. It is richer because, as we shall see, we may use more than one kind of cardinality. The problems arise primarily from the measurement status of the membership scale.

Scalar cardinality is the generalization of classical (crisp set) cardinality. For a quantifiable membership scale (i.e., either an absolute or a ratio scale), scalar cardinality is the sum of the degrees of membership over all elements. This is identical to a sum of scores for any quantitative variable. The setwise interpretation of cardinality, however, makes the link between this sum and the notion of the size of a set more direct than it is for most quantitative variables. We can see this by examining the properties of scalar cardinality. Let $|A|$ denote the scalar cardinality of Set A, and define this as

$$|A| = \sum_{i=1}^{N} m_A(X_i) \qquad [4.1]$$

where $m_A(X_i)$ is the degree of membership that the score X_i has in Set A. Assuming that membership in the complement of A is $m_{\sim A}(X_i) = 1 - m_A(X_i)$, $|A|$ obeys the properties considered essential for cardinality:

1. For any Sets A and B, if for all i, $m_B(X_i) \leq m_A(X_i)$, then $|B| \leq |A|$.
2. $|\sim A| = N - |A|$; that is, the cardinality of A and its complement sum to N.
3. For any Sets A and B, $|A \cup B| + |A \cap B| = |A| + |B|$. That is, the cardinalities of the intersection and union of two sets add up to the sum of the cardinalities of each set.
4. For Sets A and B, let the scalar cardinality of their Cartesian product be defined as $|A \times B| = \Sigma m_A(X_i) m_B(X_i)$. If A and B are statistically independent, then $|A \times B| = |A| \times |B|$ (but not the converse).

Scalar cardinality can be converted to *proportional cardinality,* that is, $||A|| = |A|/N$, whose properties are similar to those of probabilities. The first three are identical to the rules for probability, but the fourth lacks the converse that applies to probabilities:

1. $0 \leq ||A|| \leq 1$.
2. $||A|| = 0$ iff $m_A(X_i) = 0$ for all i, and $||A|| = 1$ iff $m_A(X_i) = 1$ for all i.
3. Defining $||A\&B|| = |A \times B|/N$ and the "conditional" cardinality $||A|B|| = ||A\&B||/||B||$, we have $||\sim A|B|| + ||A|B|| = 1$.
4. If A and B are statistically independent, then $||A\&B|| = ||A|| \times ||B||$ (but not the converse).

There are, however, important respects in which proportional scalar cardinality does not behave like probability. For example, $||A|A|| \leq 1$, whereas $P(A|A) = 1$, and $||A|\sim A|| \geq 0$, whereas $P(A|\sim A) = 0$. Both of these deviations from the rules for probability arise because the intermediate membership values need not sum to 1 across all alternatives, but probabilities do. For crisp sets, of course, proportional scalar cardinality is identical to probability.

4.1.1 Cardinality for Interval-Level Measurement

When the membership scale is not an absolute or a ratio scale, then scalar cardinality becomes more problematic. Let us begin by considering an interval-level membership scale for which we have no benchmark values corresponding to full membership or nonmembership. We can still compare the "size" of one set to another on this scale in the same way we compare means for any interval scale. The meaning of cardinality itself is largely lost, however, because scalar cardinality properties 2 and 4 and proportional scalar cardinality properties 1, 3, and 4 do not hold under linear transformations of the membership values.

Most of those properties can be recovered for an interval scale, however, if we have thresholds for full membership and nonmembership. A typical example is when a cutoff score is used to determine the minimum criterion for membership in a category on an interval-level scale that also has an upper bound. This amounts to imposing a linear filter on the original scale.

Let the nonmembership threshold be denoted by X_n and the full membership threshold by X_f. If, for any X between X_n and X_f, $m_A(X)$ is a linear function of X, then we obtain $||A||$ by defining a new membership function:

$$\begin{aligned} m_A^*(X) &= (X - X_n)/(X_f - X_n) \text{ if } X_n \le X \le X_f, \\ &= 0 \text{ if } X < X_n, \text{ and} \\ &= 1 \text{ if } X > X_f. \end{aligned} \qquad [4.2]$$

Then we have $||A|| = \Sigma m_A^*(X_i)/N$ as before.

➢ Example 4.1: Aggressive Behavior Scale

Heubeck (2001) compared a sample of 89 children ages 8 to 14 with another sample of children who had been referred to a child psychiatric clinic on the Aggressive Behavior subscale from the Child Behavior Checklist (Achenbach, 1991). The CBCL Aggressive Behavior scale is constructed from parents' ratings of how often the child has displayed fighting, arguing, and the like during the past 6 months and has a range from 0 to 40. This subscale has a "Clinical Borderline" cutoff score of 20.

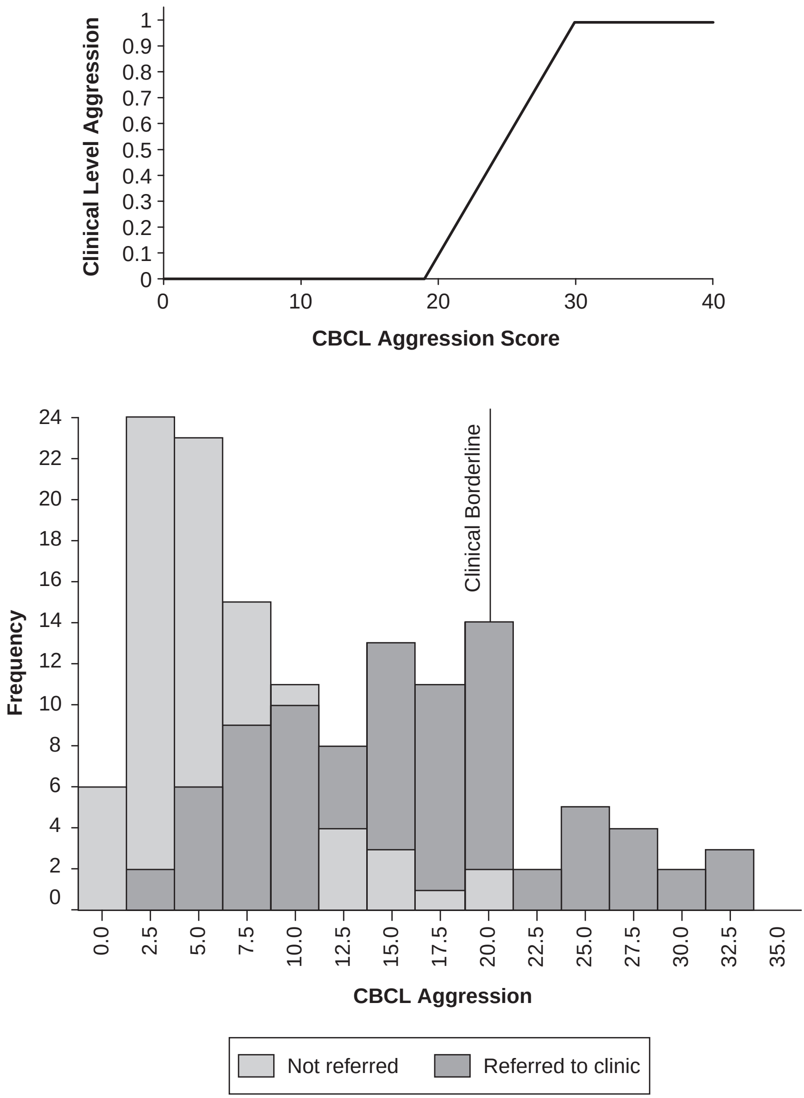

Figure 4.1 CBCL Aggression Distribution for Clinical and Nonreferred Samples

The upper half of Figure 4.1 shows a linear filter using 19 as the borderline cutoff for complete nonmembership (so that 20 has a small degree of membership) and 30 as the cutoff for full membership in the fuzzy set of "clinical level aggression." The lower half shows the distribution of CBCL scores for the children from both samples.

Using the clinical borderline of 20 to define "clinical level aggression" as a crisp set, 2 out of 89 nonreferred children are in this set whereas 22 out of 89 referred children are in the set. On the other hand, if we use the fuzzy membership function defined by the linear filter, the two nonreferred children who do not have 0 membership are marginally over the clinical borderline, with .09 and .18 membership, for a cardinality of only 0.27. The clinically referred children, however, include some very high scores that sum to a cardinality of 11.93. The ratio of these two set sizes is $11.93/0.27 = 44.19$, compared with $22/2 = 11$ for the crisp-set version. Taking into account the degree of membership in the set of "clinically aggressive" children yields a more realistic picture of the magnitude of the difference between the two samples while retaining the notion of counting.

4.1.2 Cardinality for Ordinal-Level Measurement

The interval-level approach to cardinality may be used for any ordinal scale from which interval-level information about a latent variable has been derived. For instance, as mentioned in Chapter 3, we might have a Rasch scale constructed from an ordinal one. If we are willing to convert that scale into a corresponding membership scale, then scalar cardinality may be established by the approach described above.

Even for a purely ordinal scale, we can often obtain meaningful statements about cardinality. The crudest distinctions involved in any fuzzy set are between nonmembers, possible members, and full members of a set. The possible members constitute the *fuzzy core* of the set. For example, if we know that 120 people describe themselves as "nonsmokers," 20 as "having quit," 40 as "trying to quit," and 30 as "smokers," then the fuzzy core contains $20 + 40 = 60$ people. The lower and upper bounds on the number of people who may be smokers are 30 and $20 + 40 + 30 = 90$.

If we can make one more distinction between those fuzzy core categories closer to nonmembership and those closer to full membership, then we can narrow the bounds on cardinality. Suppose we designate "having quit" as closer to nonmembership (i.e., $< 1/2$) and "trying to quit" as closer to full membership (i.e., $> 1/2$). Then the lower and upper bounds on the number of smokers become $(1/2) \times 40 + 30 = 50$ and $(1/2) \times 20 + 40 + 30 = 80$.

4.2 Probability Distributions for Fuzzy Sets

In this section, we survey probability distribution functions (PDFs) for statistical models in which a fuzzy set plays the role of a dependent variable. We focus on those PDFs for which multivariate analytical techniques are available. Based on Verkuilen (2005), a general way of describing a PDF for a fuzzy set is

$$f(m) = p_0\Delta(0) + p_1\Delta(1) + (1 - p_0 - p_1)g(m), \qquad [4.3]$$

where p_0 is the probability of nonmembership, p_1 is the probability of full membership, $\Delta(m)$ is the Dirac delta with unit mass at m, and $g(m)$ is a continuous PDF whose entire density is confined to the (0,1) interval. When full membership and nonmembership are important in themselves, this mixture distribution can lead to useful statistical models of fuzzy sets. Even when these are not particularly important, modelers need to be careful about the 0-1 endpoints of the unit interval. Clearly, default assumptions such as normality are not tenable for fuzzy sets. The most accessible alternatives include censored distributions, truncated distributions, and beta distributions.

A *censored distribution* is one in which the density over one or more subranges of the distribution has been collapsed to a single point. As Long (1997, p. 187) points out, censored distributions often arise because we lack knowledge about the variable concerned. However, a censored distribution can result when a fuzzy membership function is a filter imposed on a variable whose distribution is known or estimable. Consider a standardized IQ test that has a normal distribution with a mean of 100 and standard deviation of 15. If we defined the fuzzy set of people "capable of learning skill X" as a linear filter with 0 membership up to a score of 85 and full membership from 100 upwards, then $f(m)$ would have the form

$$f(m) = (.1587)\Delta(0) + (.5)\Delta(1) + (1 - p_0 - p_1)\phi(m),$$

where $\phi(m)$ is the normal PDF and $m = \max(0, \min(1, (\mathrm{IQ} - 85)/15))$. The resulting PDF would have a spike of .1587 at 0 and another spike of .5 at 1 with the appropriate section of the normal PDF for m between 0 and 1.

A *truncated distribution* has part of the PDF excised from it, and the distribution is then renormalized by dividing the PDF by the area under the remaining part of its curve. Truncated distributions are applicable to fuzzy sets when the researcher is studying a subpopulation selected in such a way that nonmembers and/or full members of a set are excluded. Consider the fuzzy set of "young adults" with 0 membership assigned to anyone below the age of 16. The PDF of this fuzzy set in a sample for which no one under the age of 16 has been selected from the population could be modeled with a distribution truncated at 0 membership.

Well-established models are available for modeling censored and truncated dependent variables along the lines of multiple regression. Most of these models are variations on or extensions of the tobit model, and an accessible introduction to them is provided in Long (1997, Ch. 7). We provide an example applying a tobit model in Chapter 6.

Finally, there are many situations in which neither censored nor truncated distributions are appropriate for fuzzy sets. If researchers are willing to represent 0 and 1 membership with values that are arbitrarily close to 0 or 1 but inside the unit interval (to avoid undefined quantities), then the beta distribution is both a sensible and a flexible candidate. We will briefly describe the beta distribution here because it is not commonly used in the social sciences despite its potential applicability to virtually any scale bounded at both ends. For some fuzzy set we say that m has a Beta(ω,τ) distribution when

$$f(m) = m^{\omega-1}(1-m)^{\tau-1}\Gamma(\omega+\tau)/[\Gamma(\omega)\Gamma(\tau)],$$

where $\omega,\tau > 0$, and $\Gamma(\cdot)$ denotes the gamma function. Both ω and τ are shape parameters, with ω pulling density toward 0 and τ pulling density toward 1. The beta family includes a wide variety of shapes. It includes the uniform distribution as a special case. The beta distribution is symmetric around .5 in the sense that $f(m;\omega,\tau) = f(1-m;\tau,\omega)$, so that fuzzy negation does not affect the distribution in any important way. Further details about the beta distribution may be found in Gupta and Nadarajah (2004) and Johnson, Kotz, and Balakrishnan (1995). The beta distribution defined over the unit interval sometimes is known as the "standard beta" to distinguish it from beta distributions defined over intervals with other bounds.

The usual parameterization of the beta distribution is inconvenient for many purposes because ω and τ are shape parameters whose interpretation is difficult. However, a well-known reparameterization translates ω and τ into location and dispersion parameters, thereby enabling the construction of multivariate generalized linear models (GLMs) for beta-distributed dependent variables with separate vectors of predictors for location and dispersion (Paolino, 2001; Smithson & Verkuilen, in press). An exegesis of these models is beyond the scope of this book, but the main point here is the availability of multivariate GLM techniques with beta-distributed fuzzy sets in the role of dependent variables.

4.3 Defining and Measuring Fuzziness

Broadly speaking, a fuzzy set is fuzzier if many cases have intermediate membership values, and it is less fuzzy if many cases have membership values of 0 or 1. The standard fuzzy set-theoretic definition of fuzziness holds that a set is maximally fuzzy if all its elements have membership of 1/2 in it, and completely crisp if its elements all have membership of 0 or 1. In the social sciences, fuzziness has connections with concepts such as polarization, consensus, relative variation, taxonicity, concentration, and inequality. The literature on attempts to construct measures of these properties has applications to the measurement of fuzziness, and, conversely, considerations about how to measure fuzziness could enrich debates about these properties. Thus, in several disciplines, the notion of measuring fuzziness is connected with important concepts, and researchers are likely to find a fuzziness measure useful.

A visual analogy may help here. For fixed luminance, a photo has maximal contrast if all of its pixels are black or white; it is maximally fuzzy if all of its pixels are a shade of grey in the middle of the greyscale. Fuzziness is the opposite of contrast. Thinking about fuzziness in the more conventional terms of PDFs, a crisp set's membership PDF would be bimodal, with two spikes at 0 and 1. A maximally fuzzy set's PDF would have a single spike at 1/2.

De Luca and Termini (1972) specified the following criteria for any measure of fuzziness $fuz(A)$ of Set A:

1. $fuz(A) = 0$ iff $m_A(X) = 0$ or 1 for all X;
2. $fuz(A) =$ maximum iff $m_A(X) = 1/2$ for all X;
3. $fuz(A) \leq fuz(B)$ iff for all X either $m_A(X) \leq m_B(X)$ whenever $m_B(X) < 1/2$, or $m_A(X) \geq m_B(X)$ whenever $m_B(X) > 1/2$; and
4. $fuz(A) = fuz(\sim A)$.

Several measures of fuzziness have been proposed in the fuzzy set literature. Most of these are described and compared in Smithson's (1987, 1994) overviews. As with cardinality, the extent to which we can effectively quantify fuzziness depends on the measurement level of the membership scale. As we have already seen, one of the shortcomings of the fuzzy set literature is the pervasive assumption that membership scales are absolute (or at least ratio-level), and this is reflected in the measures of fuzziness that have been developed. We will briefly survey the fuzziness measures that have been used in the fuzzy set literature and then investigate more generally useful alternatives for social science data.

The simplest fuzziness measure was proposed by Kaufmann (1975) and normed for sample size in Smithson (1987, p. 112):

$$FK = (1/NH) \sum |m_{Ai} - m^*_{Ai}|, \quad [4.5]$$

where m_{Ai} is the membership of the ith element in Set A, $m^*_{Ai} = 1$ iff $m_{Ai} \geq 1/2$, and $m^*_{Ai} = 0$; otherwise, N is the sample size, and $H = 1/2$ is the maximum possible value of the expression being summed (i.e., when $m_{Ai} = 1/2$ for all i so that the set is maximally fuzzy). This coefficient is just the average absolute difference between the ith membership value and its nearest “unfuzzy” counterpart resulting from rounding off the fractional membership values to either 1 or 0, divided by the maximum possible difference. So, it is a measure of the difference between the fuzzy set and the nearest crisp set. If there is no difference, then of course the fuzzy set is not fuzzy and $FK = 0$.

The analog between fuzziness and lack of contrast suggests that we could measure fuzziness by comparing membership in a set with membership in its complement. One of the earliest fuzziness coefficients does exactly that (De Luca & Termini, 1972, normed for sample size in Smithson, 1987, p. 112):

$$FD = (1/NH) \sum [-m_{Ai} \ln(m_{Ai}) - (1 - m_{Ai}) \ln(1 - m_{Ai})], \qquad [4.6]$$

where $H = -\ln(1/2)$, the largest value that the summation term can have. De Luca and Termini claim that it measures the amount of uncertain information conveyed by fuzzy membership values. This measures fuzziness by bit rate and is related to information-theoretic measures of entropy. If all $m_{Ai} =$ either 0 or 1, then $FD = 0$. If they all equal 1/2, $FD = 1$. Kosko (1992) also proposed an entropy-like measure of fuzziness.

Finally, the connection between fuzziness and inequality led Smithson (1982a) to suggest that measures of fuzziness could be based on *relative variation* around the mean. The main idea is that a fuzzy set fails to discriminate among its members if it assigns all of them the same degree of membership, regardless of whether that degree is 1/2 or some other value. Conversely, membership values restricted to 0 or 1 have greater variance than any other membership assignments with the same mean. Therefore, fuzziness ought to be inversely related to variance. Of course, not all researchers using fuzzy sets will agree with this claim; they may want to insist on using 1/2 as a fixed benchmark.

Allison’s (1978) review of relative variation measures specified two primary criteria that they should fulfill. First, they should be invariant under scalar multiplication (i.e., suitable for a ratio-scaled variable). Second, they should be “sensitive to transfers,” which means that they should increase in value when some amount is transferred from a low-valued element and added to a high-valued one. The Gini coefficient, the coefficient of variation, and Theil’s (1967) information-theoretic variation coefficient all satisfy Allison's criteria. Smithson (1982b; 1987, pp. 113-116) has argued that these coefficients should be normed by their maximum attainable values.

The relative variation measures are invariant under scalar multiplication, and the fuzziness measures such as those defined in Formulas 4.5 and 4.6 may also be generalized to handle ratio scales under certain conditions. Unfortunately, none of them is suited to interval or ordinal membership scales. However, some simple techniques enable us to measure fuzziness for these kinds of scale.

4.3.1 A Cumulative Distribution Approach to Fuzziness

We introduce a method of measuring fuzziness that may be applied to any membership scale that distinguishes among nonmembers, those closer to nonmembership, those closer to full membership, and full members. As in the earlier material on comparing set sizes, this approach is very widely applicable and nearly scale-free.

The empirical cumulative distribution function (CDF) may be compared with its nearest crisp-set CDF to provide an indication of fuzziness, along the lines suggested by Kaufmann (1975). The more similar these two CDFs are, the less fuzzy the set. By using the CDF, we avoid the need to quantify degrees of membership, and so this approach can be applied to an ordinal membership scale.

Now, we need a measure of the dissimilarity between the empirical and nearest-crisp CDFs. There are several alternatives, but for the sake of simplicity and general applicability, we will use the Kolmogorov goodness-of-fit statistic (cf. Conover, 1980; for alternatives, see D'Agostino & Stephens, 1986). Let $F(m_{Ai})$ denote the empirical CDF of a sample of membership values, m_{Ai}, and $F(c_{Ai})$ the nearest crisp CDF. Then the Kolmogorov statistic is the largest absolute difference between $F(m_{Ai})$ and $F(c_{Ai})$:

$$T(A) = \sup_i |F(m_{Ai}) - F(c_{Ai})|. \qquad [4.7]$$

Let s denote the cutoff score on the original scale for nonmembership, h denote the cutoff below which intermediate membership is closer to nonmembership and above which it is closer to full membership, and b denote the cutoff score for full membership. Because $F(c_{Ai})$ is a step function with

$$\begin{aligned} F(c_{Ai}) &= F(h) \text{ for all } m_{Ai} < \text{full membership and} \\ &= 1 \text{ for all } m_{Ai} = \text{full membership,} \end{aligned} \qquad [4.8]$$

$T(A)$ occurs either at s or at b. Thus, $T(A)$ depends only on the CDF's values at s, h, and b.

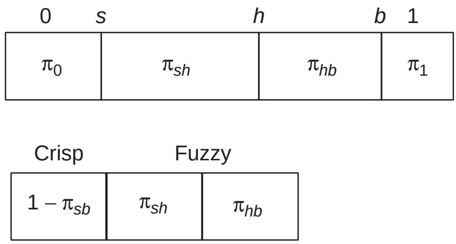

Figure 4.2 The PDF Basis for $T(A)$

Let $F(s)$ denote the value of the empirical CDF including only the nonmembers, $F(h)$ the CDF including nonmembers and intermediate members below h, and $F(b)$ the CDF including all cases except full members. Then

$$T(A) = \max[F(h) - F(s), F(b) - F(h)] = \max[p_{sh}, p_{hb}], \qquad [4.9]$$

where p_{sh} is the proportion of intermediate membership cases closer to nonmembership and p_{hb} is the proportion of intermediate membership cases closer to membership.

Figure 4.2 illustrates this scheme. The basis for $T(A)$ can be considered in terms of a four-category PDF induced by the comparisons between $F(m_{Ai})$ and $F(c_{Ai})$. The top half of this figure displays the population proportions of nonmembers (π_0), intermediate members (π_{sh} and π_{hb}), and full members (π_1). The dividing points on the membership scale (s, h, and b) are represented schematically as the borders separating each proportion of cases from its neighbor.

Clearly, if $T(A) = 0$, then A is crisp. $T(A)$ has a maximum value of 1, indicating complete concentration of the CDF on a single value or category that lies in between nonmembership and full membership. This corresponds to Smithson's (1982a) definition of maximal fuzziness. Also, $T(A) = T(\sim A)$. Finally, $T(A) \leq T(B)$ iff for all X either $m_A(X) \leq m_B(X)$ whenever $m_B(X) <$ 1/2, or $m_A(X) \geq m_B(X)$ whenever $m_B(X) > 1/2$. So $T(A)$ satisfies three of the four properties De Luca and Termini considered essential for a measure of fuzziness.

As mentioned earlier, this measure of fuzziness is nearly scale-free. If any two membership scales have criteria for nonmembership, full membership, and deciding which intermediate membership values are closer to nonmembership or to full membership, then their $T(A)$ values may be directly compared without regard for any other measurement characteristics of the underlying scales.

➢ Example 4.2: Beck's Depression Inventory

The Beck Depression Inventory II (BDI-II) (Beck & Steer, 1996) is a 21-item instrument used to assess depression severity, with a range from 0 to 63. The cutoff points used as referents for levels of depression are 0 to 13 for minimal, 14 to 19 for mild, 20 to 28 for moderate, and 29 to 63 for severe depression. We will set $s = 13.5$, $h = 19.5$, and $b = 28.5$.

Suppose we have BDI-II scores for 128 people. Using our values for s, h, and b, we compare $F(m_{Ai})$ and $F(c_{Ai})$, as shown in Figure 4.3. It turns out that

$$F(s) = 59/128 = .461,\ F(h) = 92/128 = .719,\ \text{and}$$
$$F(b) = 110/128 = .859.$$

Therefore, we have

$$p_0 = 59/128 = .461,\ p_{sh} = F(h) - F(s) = (92 - 59)/128 = .258,$$
$$p_{hb} = F(b) - F(s) = (110 - 92)/128 = .141,\ \text{and}$$
$$p_1 = (128 - 110)/128 = .141.$$

The sample estimate of $T(A)$ is

$$T(A) = \max[p_{sh}, p_{hb}] = \max(.258, .141) = .258.$$

4.3.2 Confidence Interval and Significance Tests for *T*(*A*)

We may obtain a confidence interval and significance tests for $T(A)$ by rearranging the PDF in the top half of Figure 4.2 into a three-category table as shown in the bottom half of that figure. The cells distinguish the full/nonmembership or crisp cases from the intermediate membership or fuzzy ones. The size of the *fuzzy core* of the set is the sum of intermediate membership cells, $\pi_{sb} = \pi_{sh} + \pi_{hb}$.

To begin with, we may obtain a confidence interval (CI) for π_{sb}, the size of the fuzzy core, in the usual fashion for proportions, whether exact or appropriate approximations. In our example, we will use the method developed by Wilson (1927) and recommended by Agresti and Coull (1998). Moreover, the conventional methods for multiway frequency analysis can be applied to comparing fuzzy core sizes for different sets or modeling it across collections of sets. We denote the CI for π_{sb} by $[\pi_{sbL}, \pi_{sbU}]$.

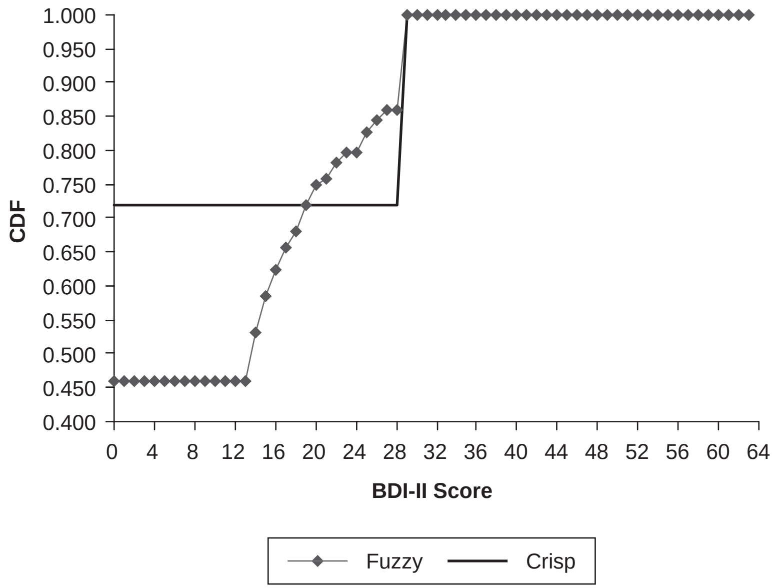

Figure 4.3 $F(m_{Ai})$ and $F(c_{Ai})$ for the BDI-II Data

Likewise, we may obtain simultaneous independent $(1 - \alpha)^{1/2}100\%$ CIs for π_{sh}, say, and π_{sb}. Then we can define a $(1 - \alpha)100\%$ CI for $T(A)$ as follows. The lower limit is

$$\pi_{shL} \text{ if } \pi_{shL}/\pi_{sbL} \geq 1/2,\ \pi_{hbL} \text{ if } \pi_{hbL}/\pi_{sbL} \geq 1/2,\ \text{and} \quad \pi_{sbL}/2 \text{ otherwise.} \qquad [4.10]$$

The upper limit is

$$\pi_{shU} \text{ if } \pi_{shU}/\pi_{sbU} \geq 1/2,\ \pi_{hbU} \text{ if } \pi_{hbU}/\pi_{sbU} \geq 1/2,\ \text{and} \quad \max(\pi_{hbU}, 1 - \pi_{shL}) \text{ otherwise.}$$

➢ Example 4.3: Beck's Depression Inventory Continued

Recall that in Example 4.2, $p_{sh} = .258$ and $p_{hb} =.141$, so the sample estimate of $T(A) = \max[p_{sh}, p_{hb}] = .258$. Starting with the fuzzy core, we have $p_{sb} = (33 + 18)/128 = .398$, and the 97.5% confidence interval for π_{sb} turns out to be [.3071, .4975]. The 97.5% CI for π_{sh} is [.1815, .3525] and the CI for π_{hb} is [.0853, .2231]. From Formula 4.8, the lower limit of the CI for $T(A)$ is .1815 and the upper limit is .3525. Whereas the lower limit differs substantially from 0, the upper limit suggests that this set is at most moderately fuzzy.

5. SIMPLE RELATIONS BETWEEN FUZZY SETS

5.1 Intersection, Union, and Inclusion

This chapter focuses on three of the elementary relationships offered by fuzzy set theory that are arguably distinctive but unfamiliar members of the family of bivariate associations (e.g., correlation, odds-ratio). These are fuzzy intersection, union, and inclusion. As explained in Chapter 2, the conventional rules for evaluating the membership of x in the intersection and union of fuzzy sets A and B are

$$\begin{aligned} m_{A\cap B}(x) &= \min(m_A(x), m_B(x)) \text{ and} \\ m_{A\cup B}(x) &= \max(m_A(x), m_B(x)). \end{aligned} \qquad [5.1]$$

Likewise, the rule stipulating that the fuzzy set A includes B ($A \supset B$) is that for all x,

$$m_A(x) \geq m_B(x). \qquad [5.2]$$

Intersection and union are distinct from addition because they are not compensatory, whereas addition is. For example, for any x from Formula 5.1, we can see that a high degree of membership $m_A(x)$ in A will not compensate a low membership $m_B(x)$ in B regarding membership $m_{A\cap B}(x)$ in $A\cap B$. Inclusion is distinct from correlation both because of its asymmetry (i.e., the extent to which A includes B tells us little about the extent to which B includes A) and its direct relationship with the logical concepts of necessity and sufficiency.

As mentioned in Chapter 3, one of the chief requirements for evaluating intersection, union, or inclusion empirically is *property ranking* (e.g., does Japan have higher membership in the set of "Asian countries" than in the set of "capitalist economies"?). Accordingly, we shall be concerned not only with the level of measurement but also with property ranking when considering techniques for evaluating fuzzy intersection or inclusion. For the remainder of this section, however, we will explore two illustrative examples in which property ranking may reasonably be assumed. Smithson (2005) goes over additional examples in detail.

➢ **Example 5.1: Attitudes Toward Immigrants**

Fuzzy set inclusion is a generalization of crisp set inclusion and thereby conceptually related to Guttman, Mokken, and Rasch scaling. Although the inequality in Formula 5.2 seldom is perfectly satisfied, real examples may be found where it holds to quite a high degree. Figure 5.1 shows one such instance, in which 84 second-year psychology students at the Australian National University rated their degree of agreement with the propositions A = "Australia should permit immigrants to enter the country" and B = "Boat people should be allowed to enter Australia and have their claims processed." There are only three exceptions to the inclusion relationship $A \supset B$, in the upper right-hand corner of the scatterplot.

This example illustrates an important connection between intersection, union, and inclusion. If $A \supset B$, then $A \cap B$ is identical to the smallest set A or B, and $A \cup B$ is identical to the larger of the two sets. In Figure 5.1, we can see that the membership assignments for $A \cap B$ will be $m_{A \cap B}(x) = m_B(x)$ in all but the three cases where $m_B(x) = 1$ and $m_A(x) = 5/6$. Likewise, we can see that the membership assignments for $A \cup B$ will be $m_{A \cup B}(x) = m_A(x)$ in all but the same three cases. The nearer the distribution of membership values $m_{A \cap B}(x)$ to those of the smallest of A or B, the closer the relationship between A and B to a true inclusion relationship. The same is true regarding the distribution of $m_{A \cup B}(x)$ and those of the largest of A or B. Of course, computing intersections and unions depends on property ranking. Without good reason, it is unwise simply to assume that property ranking holds. We feel comfortable with it here because the response scale for each item is the same and the items are of the same form.

The example in Figure 5.1 also highlights a connection between fuzzy inclusion and the logical concepts of *necessity* and *sufficiency*. A predictive interpretation of the scatterplot in Figure 5.1 is that a high membership in Set A is necessary but not sufficient to predict high membership in Set B (or conversely, high membership in B is sufficient but not necessary to predict high membership in A). These asymmetric logical or predictive relations are not assessable by symmetric measures of association such as correlation. To characterize the pattern in Figure 5.1 by saying that the two

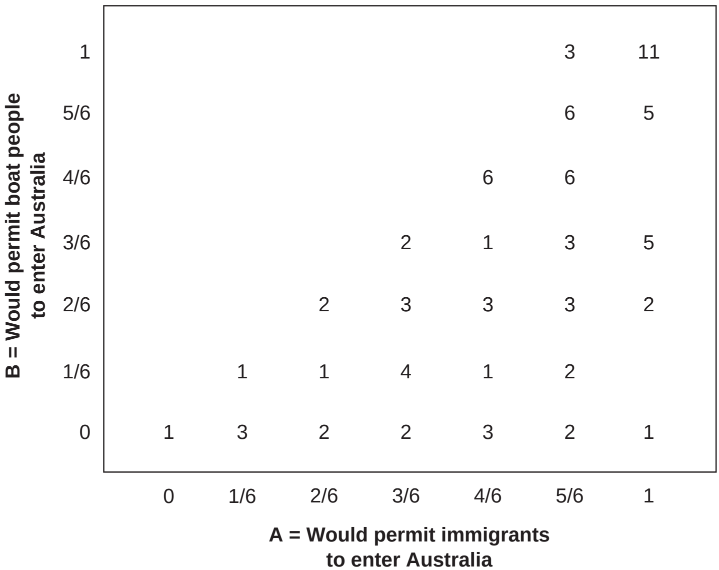

Figure 5.1 Example of Fuzzy Set Inclusion

variables have a correlation of .299 would surely miss the point. Even pointing out heteroscedasticity would not be specific enough.

Finally, it is noteworthy that inclusion, necessity, and sufficiency are special cases of a very useful and broad class of relations called *fuzzy restrictions*. A third interpretation of Figure 5.1 is that the joint distribution of A and B almost strictly satisfies the inequality $m_A(x) - m_B(x) \geq 0$. Fuzzy restrictions are generalizations of such inequalities.

➢ Example 5.2: Decisions to Disclose or Withhold Information

In an occupational survey (Bopping, 2003), 229 respondents were presented with a dilemma over whether to disclose information confided to them by a colleague. The respondents rated each of these statements on two identical 7-point scales:

I = "It is important to provide information to others"
T = "It is important to maintain the trust of a confidant"

TABLE 5.1
Cross-Tabulation of *I* and *T*

		m_I							
		0	*1/6*	*2/6*	*3/6*	*4/6*	*5/6*	*1*	*Total*
m_T	*1*	9	4	5	8	4	6	23	59
	5/6		8	9	7	11	22	5	62
	4/6		2	14		4	15	1	36
	3/6			1	7	6	10	9	33
	2/6		1	1		4	8	3	17
	1/6		1			1	5	4	11
	0	1				1		9	11
	Total	10	16	30	22	31	66	54	229

		I∪T							
		0	*1/6*	*2/6*	*3/6*	*4/6*	*5/6*	*1*	*Total*
observed	*pdf*	1	1	2	8	32	95	90	229
expected	*pdf*	0.48	2.02	7.04	15	26.9	78.5	99.1	229
sq. std.	*resid.*	0.56	0.51	3.61	3.26	0.97	3.47	0.83	13.2

We will demonstrate the application of fuzzy sets to investigating the hypothesis that respondents produced high ratings on *I* or *T* or both.

This hypothesis may be interpreted as saying that membership in the fuzzy union *I*∪*T* should be strongly skewed toward 1. A "stronger" version would predict that $T \supset \sim I$ (or equivalently, $I \supset \sim T$), i.e., $m_I(x) \geq 1 - m_T(x)$. An equivalent fuzzy restriction is $m_I(x) + m_T(x) \geq 1$. The scales for *I* and *T* have identical response formats, so for the sake of illustration, we will assume that the property ranking issue is resolved. The data in the upper part of Table 5.1 show that the strong version of this hypothesis is true for all but nine cases.

A comparison of the observed *I*∪*T* distribution with its expected-values counterpart if *I* and *T* are assumed independent (bottom part of Table 5.1) suggests that it is more strongly negatively skewed than would be expected under independence. A chi-square test may be used to compare the two distributions, and the squared standardized residuals constituting the chi-square statistic are shown in the third row of the bottom table. The chi-square test yields $\chi^2(6) = 13.2$, $p = .04$, thereby supporting the skew hypothesis. As in Example 5.1, the bivariate hypothesis investigated here would be rather difficult to evaluate using the usual concepts and measures of association but is readily accessible via fuzzy sets.

5.2 Detecting and Evaluating Fuzzy Inclusion

The task of detecting and evaluating fuzzy inclusion raises three questions. First, how do we assess the degree to which the fuzzy inclusion $m_A(x) \geq m_B(x)$ rule is satisfied? Second, how can we distinguish fuzzy inclusion from "impostors" such as the bivariate distribution of two independent skewed variables? And third, when do we have grounds for preferring a fuzzy set interpretation of our findings to rival interpretations?

Beginning with the first question, a number of fuzzy set theorists (e.g., Dubois & Prade, 1980, p. 22) have criticized the $m_A(x) \geq m_B(x)$ rule as too inflexible, and not sufficiently fuzzy. Smithson's (1987, pp. 31–32, 101–104) review of alternative proposals for evaluating fuzzy inclusion finds that they fall into two groups. One approach is to "fuzzify" the $m_A(x) \geq m_B(x)$ rule (e.g., Dubois & Prade, 1980; Ragin, 2000). The other is to construct an index of the degree of inclusion based on fuzzy set operators or other appropriate concepts. Both approaches hinge on the level of measurement possessed by the membership scales. We defer the discussion of this issue to the next section. Instead, we turn to the questions of distinguishing inclusion from impostors and deciding whether a bivariate relationship is better described as inclusion or some other kind of association. Table 5.2 makes this point by showing three impostors (the first, third, and fourth tables) and a genuine inclusion relationship (the second table).

➢ EXAMPLE 5.3: Realistic Job-Seeking/Avoiding Example

The second table is taken from real data (Smithson & Hesketh, 1998), namely 360 respondents' responses to two items on the Holland vocational

TABLE 5.2
Inclusion Relation and Impostors

Independence + Skew								
	Seek							
Not Avoid	*0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*	*Total*
0	4	3	1	1	1	0	0	10
$m_{NA}(2)$	5	4	2	2	1	1	0	15
$m_{NA}(3)$	6	4	3	2	1	1	0	17
$m_{NA}(4)$	7	5	3	2	1	1	1	20
$m_{NA}(5)$	20	13	8	6	4	2	2	55
$m_{NA}(6)$	33	23	12	10	6	4	3	91
1	55	38	21	17	11	6	4	152
Total	130	90	50	40	25	15	10	360

Inclusion Relationship

Not Avoid	*Seek* *0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*	*Total*
0	8	3	2	1	1	1	0	16
$m_{NA}(2)$	8	4	3	2	0	0	0	17
$m_{NA}(3)$	17	11	7	10	1	0	0	46
$m_{NA}(4)$	13	11	22	30	6	0	1	83
$m_{NA}(5)$	5	7	12	23	3	2	0	52
$m_{NA}(6)$	1	5	10	25	19	9	1	70
1	3	3	2	16	13	13	26	76
Total	55	44	58	107	43	25	28	360

Positive Correlation

Not Avoid	*Seek* *0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*	*Total*
0	47	0	0	0	0	0	0	47
$m_{NA}(2)$	0	47	5	0	0	0	0	52
$m_{NA}(3)$	0	0	47	10	4	0	0	61
$m_{NA}(4)$	0	0	0	45	10	0	0	55
$m_{NA}(5)$	0	0	0	0	46	5	0	51
$m_{NA}(6)$	0	0	0	0	0	47	0	47
1	0	0	0	0	0	0	47	47
Total	47	47	52	55	60	52	47	360

Negative Correlation

Not Avoid	*Seek* *0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*	*Total*
0							16	16
$m_{NA}(2)$						4		4
$m_{NA}(3)$					14			14
$m_{NA}(4)$				87				87
$m_{NA}(5)$			92					92
$m_{NA}(6)$		92						92
1	55							55
Total	55	92	92	87	14	4	16	360

interest inventory. One item has them rate the extent to which they would *seek* a job that involves "realistic" tasks, and another asks them to rate the extent to which they would *avoid* this kind of job. Both scales were identical (ranging from *not at all* to *very strongly*), and the "avoid" scale has been reverse-scored to convert it into a "not avoid" scale. The hypothesized relationship is that seeking this kind of job is sufficient but not necessary to also not avoid it, because one could decide not to avoid it for other reasons as well. So "seeking" is included in "not avoiding."

All four two-way tables have a very similar proportion of cases obeying the fuzzy inclusion rule $m_A(x) \geq m_B(x)$. Excluding the zero-membership cases on the included set, the proportions are .887, .889, .891, and .889 for the first, second, third, and fourth tables, respectively. However, the uppermost table was generated by cross-tabulating two independent skewed distributions. A chi-square test for this table yields $\chi^2(36) = 3.669$, which is a very good fit with the independence model. The apparently strong inclusion relationship in this table is due solely to the skew in both distributions.

Moving now to the other three tables in Table 5.2, a chi-square test yields $\chi^2(36) = 234.036$ for the second table, $\chi^2(36) = 1781.344$ for the third table, and $\chi^2(36) = 3625.220$ for the fourth table, all indicating large departures from independence. However, the third and fourth tables show very strong correlation patterns rather than an inclusion relationship, even though the proportion of cases obeying the fuzzy inclusion rule is nearly identical to the second table. Many researchers would prefer to describe the third and fourth tables in terms of this correlation, which measures the strength of a one-to-one association between two variables, as opposed to the one-to-many of necessity. We could readily imagine (and find) "intermediate" situations in which there is both a moderately strong correlation and a reasonably strong inclusion relationship.

Which description should we prefer, and why? This problem is more difficult than merely detecting independence, because more judgments are required. For instance, even if correlation provides a "good" description of the relationship (i.e., all assumptions and requirements such as homoscedasticity are satisfied), the inclusion interpretation might still be the more theoretically relevant. On the other hand, inclusion is a one-to-many relation and thus is a less precise proposition than a one-to-one relationship such as correlation or stronger measures of association.

Let us dispense with independence + skew first. Independence + skew cannot be a genuine inclusion relation because there *is* no association between two statistically independent random variables. Nevertheless, it is easy to make examples of skewed statistically independent random variables that seem to satisfy fuzzy inclusion, as our example demonstrates.

When two variables are statistically independent, their joint distribution is *completely* determined by the marginal distributions because the joint distribution is just the product of the marginals. The marginals, in turn, depend on the assignment of membership. As we have already seen in Chapter 3, assignment of membership is a very difficult task. It is wise to rest one's conclusions on the assignment process as little as possible because it is almost always possible to argue that a given assignment is wrong. For discrete membership scales, as in this example, the conventional chi-square test of independence usually will suffice. For continuous membership scales, the Kolmogorov-Smirnov test is the most well-known, and it compares the observed joint cumulative distribution function (JCDF) against the expected JCDF under independence. For alternative approaches, see D'Agostino and Stephens (1986).

Association + skew raises other issues. We take the view that if the bivariate distribution satisfies the relevant assumptions and the researcher is primarily interested in predicting one variable from the other, then a correlation-regression description may be preferable to a fuzzy set perspective. Better still would be a GLM that models location and dispersion simultaneously. On the other hand, particular kinds of heteroscedasticity, a strong inclusion rate combined with a marked difference in the sizes of the two sets, and/or research questions that are expressed in set-like terms should motivate a serious consideration of fuzzy inclusion as a description of the patterns. The next two sections present techniques for investigating inclusion relations in detail.

5.3 Quantifying and Modeling Inclusion: Ordinal Membership Scales

In many circumstances, we may wish to evaluate how robust a claim about an inclusion relationship is against alternative membership value assignments. For both the $m_A(x) \geq m_B(x)$ rule and any inclusion index, the joint ordering of membership values for the two sets crucially determines the result, so it is essential to explore what happens to inclusion rates and index values when the joint ordering is modified. A reasonable approach to assessing how dependent our results are on the joint ordering of membership values is to stipulate a benchmark inclusion rate before seeing the data, and then ascertain the collection of paths whose confidence intervals (CIs) include that rate or higher. One way to determine the relevant "collection" is to begin with a specific joint ordering of the values that yields a path whose inclusion CI contains the prescribed rate, and then ascertain which neighboring paths' CIs also include that rate.

TABLE 5.3
Paths With Inclusion Rate CI Containing 0.9

	Seek							
Not Avoid	*0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*	*Total*
0	8	3	2	1	1	1	0	16
$m_{NA}(2)$	8	4	3	2	0	0	0	17
$m_{NA}(3)$	17	11	7	10	1	0	0	46
$m_{NA}(4)$	13	11	22	30	6	0	1	83
$m_{NA}(5)$	5	7	12	23	3	2	0	52
$m_{NA}(6)$	1	5	10	25	19	9	1	70
1	3	3	2	16	13	13	26	76
Total	55	44	58	107	43	25	28	360

To see how this works, let us return to the job-seeking example using a criterion inclusion rate of .9. As mentioned earlier, the proportion of cases obeying the $m_A(x) \geq m_B(x)$ rule for the diagonal path is 271/305 = .889. A 95% CI for this path is [.848, .922], so it is compatible with an inclusion rate of .9. In fact, it can be shown that any path with a proportion as low as 264/305 has a CI that includes .9.

The second table from Table 5.2 is reproduced in Table 5.3. The shaded region denotes the collection of paths involving one alteration in the original values' joint ordering whose inclusion proportions are at least 264/305. The joint ordering of memberships corresponding to the diagonal path is $0 < m_S(2) = m_{NA}(2) < m_S(3) = m_{NA}(3) < m_S(4) = m_{NA}(4) < m_S(5) = m_{NA}(5) < m_S(6) = m_{NA}(6) < 1$. The collection of paths forms a region that begins only slightly beneath the diagonal path. For example, the path deviating from the diagonal once by following the cells with frequencies {8, 4, 22, 30, 3, 9, 26} corresponds to the joint ordering $0 < m_S(2) = m_{NA}(2) < m_{NA}(3) < m_S(3) < m_S(4) = m_{NA}(4) < m_S(5) = m_{NA}(5) < m_S(6) = m_{NA}(6) < 1$.

In the absence of an inclusion rate criterion, we may use the single alteration in the joint-ordering criterion for exploring the sensitivity of inclusion rates to membership assignments. Again starting with the inclusion rate of .889 for the diagonal path, the largest possible change in this rate incurred by one alteration of the joint ordering is the exclusion of the 30 cases in the $\{m_S(4) = m_{NA}(4)\}$ cell. Excluding them by "lowering" the path decreases the inclusion rate from .889 to (271 – 30)/305 = .790. The biggest increase of the inclusion rate by one alteration in joint ordering is the inclusion of the 10 cases in the $\{m_S(4) = m_{NA}(3)\}$ cell, resulting in a rate of (217 + 10)/305 = .921. Neither the proportion of cases obeying the $m_A(x) \geq m_B(x)$

rule nor the inclusion indexes do a good job of distinguishing between negative correlation and genuine inclusion because they are strongly influenced by the marginal distributions. The tests proposed in Ragin (2000) based on the $m_A(x) \geq m_B(x)$ rule implicitly assume that the marginal distributions are uniform. To avoid making strong assumptions about the marginal distributions, we must turn to models of inclusion based on localized inclusion relations in tables and scatterplots.

One way of modeling inclusion throughout a scatterplot or table is via level sets, which were introduced in Chapter 2. We may establish the inclusion rate for any cell in a table (or point on a scatterplot) by constructing the joint cumulative distribution function (JCDF). The first table in Table 5.4 shows the JCDF from Table 5.3, which accumulates frequencies starting in the {1,1} cell at the lower right and moving upward and to the left. That cell contains 26 cases, so moving up one cell accumulates 1 more to give $26 + 1 = 27$, whereas moving one cell to the left accumulates 13 cases to give $26 + 13 = 39$, moving one cell up and to the left accumulates $1 + 13 + 9$ cases to give $26 + 1 + 13 + 9 = 49$, and so on.

The second table shows the local inclusion rate for each cell. These are determined by dividing the cumulative frequency in that cell by the column cumulative total, located in the first row of the table. For the lower-right cell, we have $26/28 = .929$, for the next cell to the left we have $39/53 = .736$, and so on. We may regard these proportions as local inclusion rates because each of them is the proportion of cases that obeys the $m_A(x) \geq m_B(x)$ rule for the level set that corresponds to its cell. Consider the {.83, .83} cell, which has a JCDF of 49 cases. There are 53 cases for which membership in Seek is .83 or above, and 49 of those obey the $m_A(x) \geq m_B(x)$ rule because their membership in Not Avoid also is .83 or above. The proportion is therefore $49/53 = .925$, the entry for that cell in the second table.

The level set and JCDF approach enables researchers to examine patterns of local inclusion rates. Notice that the inclusion rates for the cells on the diagonal path are very similar to one another. This path arguably has a constant inclusion rate along it, and we shall see shortly how to test a constant inclusion model for this path. The inclusion rate pattern in Table 5.4 contrasts vividly with that for the negative correlation example, shown in Table 5.5. The inclusion rates along the diagonal path clearly are not constant, but instead jump suddenly from 0 to quite high levels as we move upward and to the left along that path. This comparison demonstrates that local inclusion models can distinguish between relationships that a global inclusion rate or inclusion index cannot.

Now let us test a constant inclusion model for the diagonal path in the job-seeking example. It turns out that the average inclusion rate along this path is .947. We will test whether the local inclusion pattern along the

TABLE 5.4
JCDF and Local Inclusion Rates for Example 5.3

Joint Cumulative Distribution

	Seek						
Not Avoid	*0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*
0	360	305	261	203	96	53	28
$m_{NA}(2)$	344	297	256	200	94	52	28
$m_{NA}(3)$	327	288	251	198	94	52	28
$m_{NA}(4)$	281	259	233	187	93	52	28
$m_{NA}(5)$	198	189	174	150	86	51	27
$m_{NA}(6)$	146	142	134	122	81	49	27
1	76	73	70	68	52	39	26

Local Inclusion Rates

	Seek						
Not Avoid	*0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*
0							
$m_{NA}(2)$	0.956	0.974	0.981	0.985	0.979	0.981	1.000
$m_{NA}(3)$	0.908	0.944	0.962	0.975	0.979	0.981	1.000
$m_{NA}(4)$	0.781	0.849	0.893	0.921	0.969	0.981	1.000
$m_{NA}(5)$	0.550	0.620	0.667	0.739	0.896	0.962	0.964
$m_{NA}(6)$	0.406	0.466	0.513	0.601	0.844	0.925	0.964
1	0.211	0.239	0.268	0.335	0.542	0.736	0.929

diagonal path is consistent with a constant inclusion rate of .947. There are several methods for doing this, but the most familiar and perhaps simplest is to use a chi-square test. The principle is to generate expected frequencies for the JCDF along the diagonal path, obtain expected frequencies for that path by taking the differences between adjacent cells, and then compare those with the observed frequencies for the same path using a one-way chi-square test.

The first table in Table 5.6 shows how to obtain the observed frequencies for each cell in the diagonal path by taking differences between adjacent cells, starting at the lower right. The second table shows how the expected frequencies are computed, using an inclusion rate of .947 and the marginal observed frequencies in the first row of the first table. The expected frequency for the upper-leftmost cell, 71.165, is computed by subtracting the sum of the other expected frequencies from the total sample size, 360.

TABLE 5.5
JCDF and Local Inclusion Rates for Negative Correlation Example

Joint Cumulative Distribution

	Seek						
Not Avoid	*0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*
0	360	305	213	121	34	20	16
$m_{NA}(2)$	344	289	197	105	18	4	0
$m_{NA}(3)$	340	285	193	101	14	0	0
$m_{NA}(4)$	326	271	179	87	0	0	0
$m_{NA}(5)$	239	184	92	0	0	0	0
$m_{NA}(6)$	147	92	0	0	0	0	0
1	55	0	0	0	0	0	0

Local Inclusion Rates

	Seek						
Not Avoid	*0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*
0							
$m_{NA}(2)$	0.956	0.948	0.925	0.868	0.529	0.200	0.000
$m_{NA}(3)$	0.944	0.934	0.906	0.835	0.412	0.000	0.000
$m_{NA}(4)$	0.906	0.889	0.840	0.719	0.000	0.000	0.000
$m_{NA}(5)$	0.664	0.603	0.432	0.000	0.000	0.000	0.000
$m_{NA}(6)$	0.408	0.302	0.000	0.000	0.000	0.000	0.000
1	0.153	0.000	0.000	0.000	0.000	0.000	0.000

Because there are seven cells, we have 6 degrees of freedom and so, using a significance criterion of .05, the critical chi-square value is 12.592. The observed chi-square turns out to be $\chi^2(6) = 3.257$, which is well below the critical value and indicates a rather good fit between the constant inclusion model and the data. There are inclusion rates other than .947 that we could not reject using the chi-square test. It is not difficult to obtain a 95% confidence interval for the constant inclusion rates, although it should be borne in mind that the chi-square version is conservative, and the resulting CI is [.888, 1]. Likewise, it is possible to find the collection of all paths that are compatible with a constant inclusion model, whether for a prespecified rate or in general. However, a full exploration of this topic is beyond the scope of this chapter.

Now let us test a constant inclusion model for the diagonal path in the negative correlation example. Table 5.7 shows the observed frequencies. It turns out that no matter what inclusion rate is used, the chi-square test

TABLE 5.6
Constant Inclusion Model

Observed Frequencies

Not Avoid	*Seek* 0	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	1
0	360 − 305 = 55	44	58	107	43	25	28
$m_{NA}(2)$		297 − 251 = 46					
$m_{NA}(3)$			251 − 187 = 64				
$m_{NA}(4)$				187 − 86 = 101			
$m_{NA}(5)$					86 − 49 = 37		
$m_{NA}(6)$						49 − 26 = 23	
1							26

Expected Frequencies

Not Avoid	*Seek* 0	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	1
0	71.165	44	58	107	43	25	28
$m_{NA}(2)$		$.947 \times 44 =$ 41.668					
$m_{NA}(3)$			$.947 \times 58 =$ 54.926				
$m_{NA}(4)$				$.947 \times 107 =$ 101.329			
$m_{NA}(5)$					$.947 \times 43 =$ 40.721		
$m_{NA}(6)$						$.947 \times 25 =$ 23.675	
1							$.947 \times 28 =$ 26.516

rejects a constant inclusion model for this table. The lowest chi-square obtainable (for an inclusion rate of 1) is $\chi^2(6) = 40.959$, substantially higher than the critical chi-square value of 12.592. The constant inclusion model successfully distinguishes between the job-seeking and negative correlation examples.

TABLE 5.7
Frequencies for the Negative Correlation Example

Observed Frequencies

	Seek						
Not Avoid	*0*	$m_S(2)$	$m_S(3)$	$m_S(4)$	$m_S(5)$	$m_S(6)$	*1*
0	360 – 289 = 71	92	92	87	14	4	16
$m_{NA}(2)$		289 – 193 = 96					
$m_{NA}(3)$			193 – 87 = 106				
$m_{NA}(4)$				87			
$m_{NA}(5)$					0		
$m_{NA}(6)$						0	
1							0

Likewise, it can be readily proved that the only constant inclusion paths for two statistically independent fuzzy sets are horizontal. For ordinal categorical membership functions and contingency tables, this property follows from the same argument behind the formula for computing expected frequencies when independence is assumed. The horizontal inclusion path result underscores our reasons for not considering the independence + skew inclusion pattern as a genuine inclusion relation.

5.4 Quantified and Comparable Membership Scales

When $m_A(x)$ and $m_B(x)$ are quantified and comparable, the fuzzy set tool chest opens up. This concluding section to Chapter 5 presents a brief survey of the possibilities that await the researcher in this situation.

➢ Example 5.4: Fear and Loathing in the Tropics

The data set that will be used for illustrations here was collected from 262 psychology undergraduates (from James Cook University in a tropical region in Australia). The data comprise their self-reported feelings about 31 noxious stimuli, such as snakes or vomit. They were asked to rate their degree of fear, disgust, and dislike for each stimulus, using a 4-point rating scale ranging from 0 = *not at all* to 3 = *very much.* The 31 ratings of fear, disgust, and dislike were summed and divided by 31 to obtain fuzzy membership scales for each. For our purposes here, we will treat these as quantified comparable membership scales. The chief object in this study was a hypothesis that the "phobic"-style responses fear and disgust are subsets of dislike, which is considered to be a much broader emotional response. A subsidiary question was the "comorbidity" issue raised often in clinical and health psychology; that is, to what extent respondents simultaneously fear and are disgusted by noxious stimuli.

5.4.1 Cardinality of Intersections and Unions

Starting with the comorbidity issue, a traditional approach would use correlation (we shall explore this issue in greater detail in Chapter 6). Table 5.8 shows that all three fuzzy sets are significantly and moderately correlated. However, the correlations cannot tell us whether one set strongly includes another, nor do they provide meaningful estimates of the relative sizes of these sets or their intersections.

We may measure the cardinality (size) of fuzzy set intersections and unions, thereby enabling us to directly address comorbidity. The upper subtable in Table 5.9 shows the mean memberships of *Fear*, *Dislike*, and *Disgust* on the diagonal and the mean memberships of their pairwise intersections in the off-diagonal cells. The lower subtable shows the proportion of each set accounted for by its intersection with another set. For instance, the intersection between *Fear* and *Dislike* has average membership .229. Because the average membership in *Fear* is .231 and in *Dislike* is .563, the proportion of *Fear* accounted for in the intersection is .229/.231 = .991 and the proportion of *Dislike* accounted for is .229/.563 = .407.

The comorbidity picture presented by fuzzy intersections differs vividly from the correlational perspective. It is evident that the comorbidity rate for *Fear* and *Disgust* is quite high (78.4% of *Fear* and 84.2% of *Disgust* are accounted for by their intersection). *Dislike* clearly includes most of *Fear* and *Disgust* (99.1% and 99.5% respectively), but only 38.0% of *Dislike* is accounted for by its intersection with *Disgust* and only 40.7% is accounted

TABLE 5.8
Correlations Among *Fear*, *Disgust*, and *Dislike*

Fear		
.434	*Dislike*	
.747	.410	*Disgust*

TABLE 5.9
Mean Membership of *Fear*, *Disgust*, *Dislike*, and Their Intersections

	Fear	*Dislike*	*Disgust*
	Mean Membership		
Fear	0.231		
Dislike	0.229	0.563	
Disgust	0.181	0.214	0.215
	Intersection Proportions		
	Fear	0.407	0.842
	0.991	*Dislike*	0.995
	0.784	0.380	*Disgust*

for by its intersection with *Fear*. The finding that *Dislike* strongly includes *Fear* and *Disgust* is supported by their scatterplots in Figure 5.2.

Even though *Dislike* subsumes most of *Fear* and *Disgust*, is the union of *Fear* and *Disgust* sufficient to include most of *Dislike*? The average membership in *Fear*∪*Disgust* turns out to be .266, which is less than half the size of *Dislike* (.563). In fact, the average membership in the intersection (*Fear*∪*Disgust*)∩*Dislike* is .263, so *Dislike* includes 100(.263/.266) = 98.7% of *Fear*∪*Disgust*. These findings indicate that *Dislike* is a much broader category than the union of *Fear* and *Disgust*. Even in this quick exploration of intersections and unions, we have gone far beyond anything that correlation or regression could tell us.

5.4.2 Inclusion Coefficients

We now turn to inclusion coefficients as a way of quantifying the degree to which one set includes another. The simplest index of inclusion is just the proportion of cases satisfying the $m_A(x) \geq m_B(x)$ rule, the "Classical Inclusion Ratio" presented in Chapter 2. While attractive for its simplicity, its main limitation is that a "near miss" is counted as strongly as a drastic counterexample. We present two coefficients that overcome this limitation, namely, the "Inclusion 1" and "Inclusion 5" indexes discussed in Smithson (1994). The first inclusion index is defined by

$$I_{AB} = \sum m_{A\cup B}(x_i) \Big/ \sum m_B(x_i). \qquad [5.3]$$

I_{AB} is the proportion of Set *B* in the intersection of Sets *A* and *B* (Sanchez, 1979). It is clearly based on fuzzy set-theoretic concepts. We already used

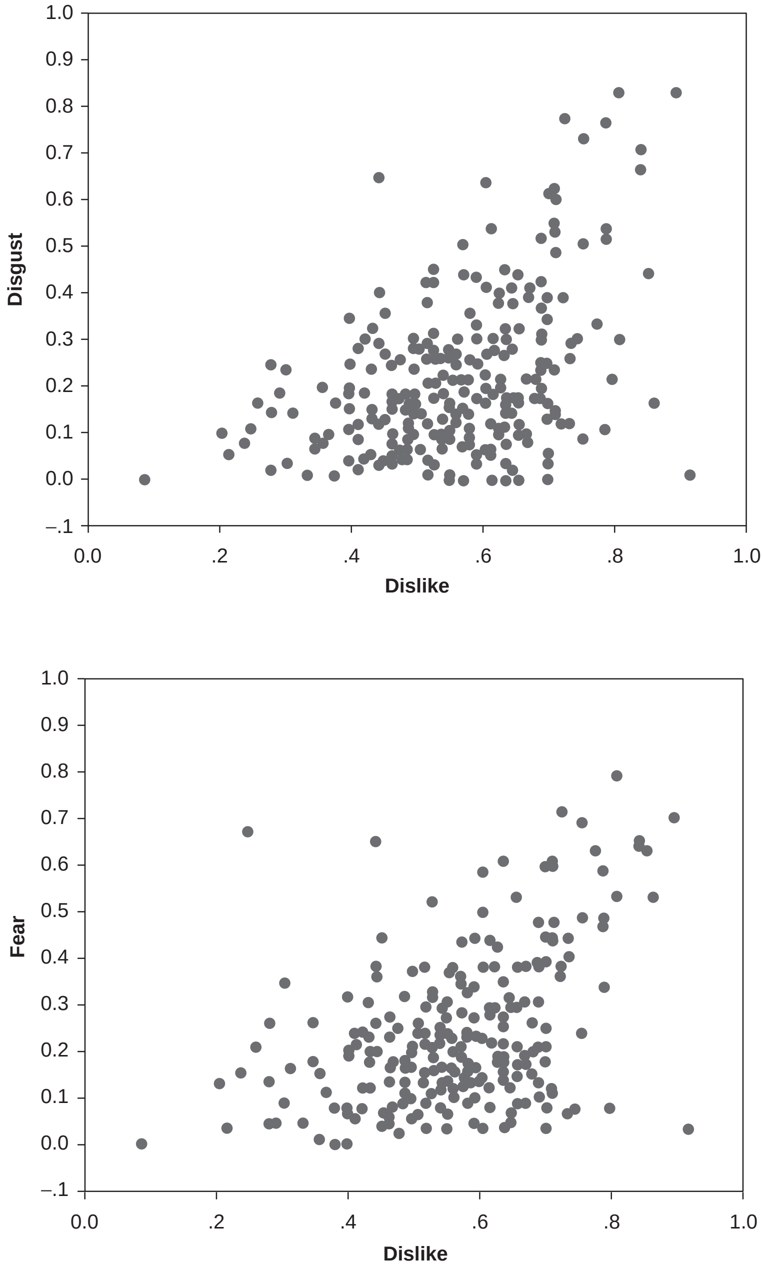

Figure 5.2 Scatterplots of *Disgust* × *Dislike* and *Fear* × *Dislike*

this coefficient in Table 5.9, where we discussed the proportion of one set accounted for by its intersection with another.

There is an important link between the JCDF and local inclusion approach from the previous section and *IAB*. Summing the JCDF values along the diagonal path in Table 5.4 and dividing by the number of cells gives (26 + 49 + 86 + 187 + 251 + 297)/6 = 149.333. This is the cardinality of the intersection between the *Seek* and *Not Avoid* sets, if we are willing to regard the membership scale as taking values of k/K, for $k = 0, 1, 2, \ldots, K$, where K is the number of nonzero membership levels. Summing the column CDF and dividing that by 6 yields (28 + 53 + 96 + 203 + 261 + 305)/6 = 157.667, the size of the *Seek* set. So the inclusion index $I_{AB} = 149.333/157.667 = .947$, and it may be thought of as the sum of the JCDF along the path divided by the sum of the CDF of the included set if we are willing to regard the membership values as taking values of k/K, where K is the number of cells in the path. This argument was the basis for choosing to test a constant inclusion model with a rate of .947.

The "Inclusion 5" coefficient is defined by

$$C_{AB} = \frac{\sum \max(0, m_A(x_i) - m_B(x_i))}{\sum |m_A(x_i) - m_B(x_i)|}. \qquad [5.4]$$

C_{AB} is the proportion of deviations from equality between $m_A(x)$ and $m_B(x)$ that are in the appropriate direction. It is actually a generalization of the proportion of observations with unequal membership obeying the strict inequality $m_A(x) > m_B(x)$.

Which of these indexes is preferable depends on the researcher's goals. To start with, cases of 0-valued membership for either set do not affect the value of I_{AB}, but they do affect C_{AB}. Second, cases where $m_A(x_i) = m_B(x_i)$ do not affect C_{AB}, but do affect I_{AB}. Third, $C_{AB} = 1 - C_{BA}$ but this does not hold for I_{AB}. Finally, neither coefficient is defined casewise, an attractive property for estimation purposes (Smithson, 1987, 1994 review others that are).

As with any coefficient designed to measure a particular kind of relationship and no other, inclusion coefficients have their limitations. First, neither of the inclusion coefficients tells us whether independence holds. Additionally, as mentioned previously, they are strongly influenced by the marginal distributions; an inclusion index that is free of the margins in the same way that the odds ratio is for 2 × 2 tables would be highly useful. Table 5.10 shows inclusion coefficients for the four tables in Table 5.2. For the independence example (the first table in Table 5.2), $I_{AB} = .914$ and $C_{AB} = .962$, both of which appear impressive unless we know about independence.

TABLE 5.10
Inclusion Coefficients for Table 5.2 Examples

	A = *Not Avoid*	*B =* *Seek*	*Intersection*	I_{AB}	I_{BA}	C_{AB}	C_{BA}
Indep.	282.667	90.833	83.000	0.914	0.294	0.962	0.038
Inclusion	228.667	157.667	149.333	0.947	0.653	0.905	0.095
Pos. Corr.	176.667	183.000	176.667	0.965	1.000	0.000	1.000
Neg. Corr.	241.833	118.167	94.833	0.803	0.392	0.863	0.137

Likewise, neither coefficient does a good job of distinguishing the skewed negative correlation case from the true inclusion case. Finally, for a high positive correlation, I_{AB} and C_{AB} behave quite differently because although I_{AB} and I_{BA} both may be high, $C_{AB} = 1 - C_{BA}$, so if one is high, the other will be low.

However, all that these limitations imply is that an inclusion coefficient on its own is insufficient to tell us whether we would prefer to describe a bivariate relationship in terms of inclusion or some other model. Currently, the best way to rigorously address this issue is by inspecting scatterplots and modeling local inclusion rates over the bivariate distribution. Thus, while an inclusion coefficient provides a useful omnibus statistic, it shares the weaknesses of all omnibus statistics. In our view, inclusion coefficients are useful as summaries of bivariate setwise relations or as a means to sift through many pairs of variables. Chapter 6 provides some examples of how inclusion coefficients may be generalized and used to gain insights into multiset relations.

6. MULTIVARIATE FUZZY SET RELATIONS

This chapter introduces multiset relations and concepts. We begin with compound set indexes and conditional membership functions, both of which highlight the advantages of combining fuzzy set tools with more traditional concepts of scale construction. The focus then shifts to multiset intersections and unions. We show how a fuzzy set approach can help resolve long-standing debates concerning the concept of comorbidity, by explicating the links between co-occurrence, fuzzy intersection, and covariation. An example illustrates how a very different and arguably more accurate view of comorbidity is obtained by using fuzzy intersections instead of correlations. The chapter concludes with an introduction to multiple and partial intersection and inclusion.

6.1 Compound Set Indexes

One of the powerful aspects of fuzzy set theory is that the mathematical operations can be used to make compound indexes that encapsulate a particular theoretical definition in a quantitative index. Of course, we could always do this in an ad hoc manner, but fuzzy sets are particularly natural because many theoretical definitions are specified in logical or set-theoretic fashions. For instance, theory often specifies that a property B is defined over objects X by having one or more necessary conditions, $A_1, A_2, \ldots A_k$. This is mathematically equivalent to saying that Set X must be included in the joint intersection $A_1 \cap A_2 \cap \ldots \cap A_k$. If some or all of the component sets are fuzzy, then the membership values can be computed easily:

$$m_B(x) = \min(m_1(x), m_2(x), \ldots, m_k(x)). \qquad [6.1]$$

Much more complex indexes are possible so long as the definitions can be translated into set-theoretic terms. As we have noted on several occasions, unlike weighted sums, the min-max operators are noncompensatory (in some areas, the term is “noninteractive”). In Formula 6.1, high membership in A_1, for example, will not compensate for low membership in A_2. Therefore, researchers may wish to consider whether to use the standard min-max operators or alternatives such as the product and probabilistic sum (Smithson, 1987, pp. 26–29).

➢ Example 6.1: Index of Electoral Democracy

In Chapter 3, we introduced the Electoral Democracy Index (Munck & Verkuilen, 2003; UNDP, 2004). As mentioned therein, the EDI consists of four components: Suffrage (S), Offices (O), Free (F), and Clean (C). S refers to the right of all adults to vote, O to the condition where the decision-making offices (executive and legislative) are filled by elections, F to the right of party competition and organization, and C to the right to fair voting and vote-counting procedures. Theoretical considerations suggest that these components are all necessary conditions, so the absence of any of the components is sufficient to make a country a nondemocracy. This implies that $EDI = S \cap O \cap F \cap C$. However, all of the components are matters of degree. It is possible for suffrage, for example, to be widespread but not all-encompassing, as was the case in the United States prior to the electoral reforms of the 1960s, when widespread restrictions on the right to vote existed throughout the South.

We therefore define a fuzzy set EDI as the intersection of the components. The standard approach would be $m_{ED}(x) = \min(m_S(x), m_O(x), m_F(x), m_C(x))$. Because the min operator is not compensatory, countries such as these two would have identical EDI memberships: min(.25,.25,.25,.25) = min(1,1,1,.25) = .25. However, these two countries would be considered quite different by most observers. The first one is a consistent poor performer, whereas the second one is an otherwise democratic system that had a stolen election. Instead of using the min operator for intersection, the product operator was used: $m_{ED}(x) = m_S(x) \cdot m_O(x) \cdot m_F(x) \cdot m_C(x)$. Using the product operator, the two countries in our example have substantially different EDI membership values. The first case has membership .004, whereas the second has membership .25.

6.1.1 Conditional Membership Functions

One early criticism of fuzzy set theory was its insensitivity to context (Amarel, 1984; Zeleny, 1984). It takes no great intellectual leap to posit that the fuzzy membership function of "tall" for men will differ from that for women in the same population (see Foddy & Smithson, 1989, for an application of this point to the social psychology of status characteristics). Although no explicit provision for conditioning membership is made in the basic fuzzy set framework, the general idea is simple enough. All that is required is one or more variables on which to condition the membership function and an explicit definition of the conditioning function.

However, the term *conditional* can mislead because in common use, it has two meanings. One is *predictive*, that is, the membership in *A* that is predicted given *B*. The other is *assigned*, that is, the membership in *A* that is assigned given *B*. In this section, we will focus principally on assigned conditional membership.

The simplest assigned conditional membership is the *normalization* of a fuzzy set, that is, dividing the raw membership values by the highest value in the set. Normalization is a special case of relativized membership, whereby membership is evaluated relative to subset prototypes. The "tall man" versus "tall woman" membership function may be interpreted as an instance of membership relativization.

A more interesting form of conditioning is on one or more variables, whereby the degree of membership in Set *A* depends by definition on the conditioning variables. Smithson (1987, pp. 281–282) uses this kind of conditioning to measure the popularity of transportation options for an elderly population. People who do not have a private vehicle cannot "choose" whether or not to use it, nor can people "choose" to take the bus if that is their only option. Thus, if 42% of a population have a private vehicle and all of them have other transportation options as well, then if 32% of them use private vehicles for errands, the conditional degree of popularity of this option is .32/.42 = .76. Likewise, if 72% of this population can use the bus system, but 12% have no other alternative, then if 32% of them use buses for errands, the conditional degree of popularity of this option is (.32 – .12)/(.72 – .12) = .33.

➢ Example 6.2: The Torino Scale

Assigned conditional scales are not commonly used in the social sciences, with most of the exceptions occurring in applied domains such as clinical psychology and risk assessment. A real-world example that can be reinterpreted as a conditional membership function is the Torino Scale (Binzel, 1999), developed in the wake of public confusion over false alarms concerning the likelihood of asteroid collisions with Earth. The scale's thresholds are determined by the estimated kinetic energy (MT) and probability of collision, as shown in Figure 6.1. The numerals from 0 to 10 are analogous to degrees of membership in the set of potentially destructive encounters with asteroids.

Perhaps the most instructive aspect of this scale is the manner in which collision probability thresholds distinguishing one degree of membership from another are conditioned by kinetic energy. Clearly, as MT increases, three of these thresholds decrease, indicating greater caution (loss aversion) regarding the likelihood of collision. In this example, events with identical membership values imply equivalent expected utilities. For instance, on the border between a rating of 1 and 2, events with MT = 100 and collision probability = 10^{-2} are equated with events with MT = 10^{8} and collision probability = 10^{-6}.

6.2 Multiset Relations: Comorbidity, Covariation, and Co-Occurrence

In this section and the next, we will generalize the fuzzy set concepts of intersection, union, and inclusion to handle more than two sets. To start with, we will make a case for the utility of multiset fuzzy intersection and related concepts by showing how they can shed light on long-standing debates concerning the concept of comorbidity.

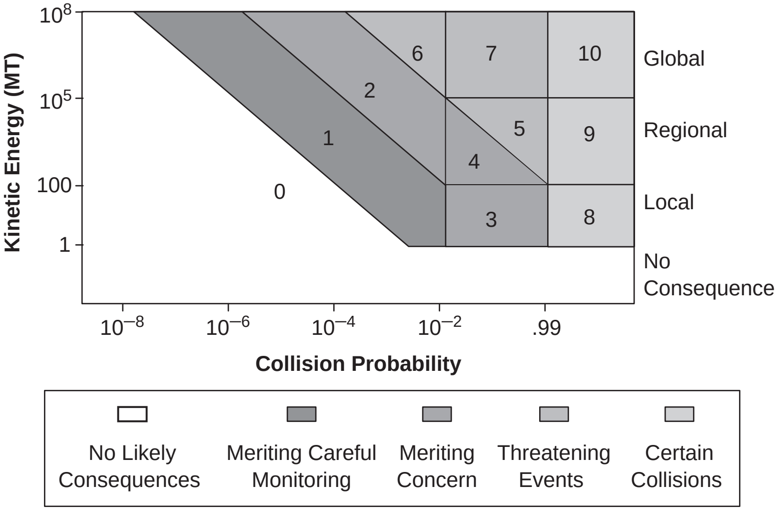

Figure 6.1 The Torino Scale

SOURCE: Adapted from Binzel (1999).

In health-oriented disciplines such as clinical psychology, comorbidity means the co-occurrence of two or more syndromes. Borrowed from medicine, this concept has been widely criticized and disputed, partly because it corresponds to a categorical view of phenomena and stems from a medical disease model where that view is (mostly) appropriate. Some authors now prefer the terms *covariation* and *co-occurrence*, but they see these as quite separate, with the former applying only to a dimensional viewpoint and the latter only to a categorical view.

This perspective is needlessly restrictive. Fuzzy sets offer a bridge between the concepts of covariation and co-occurrence. For the sake of simplicity, but without losing much generality, let us assume that we have interval-level scales on which we are willing to impose lower and upper bounds corresponding to nonmembership and full membership in fuzzy sets A and B. Then we may still use Pearson's R as a scale-free measure of the covariation between m_A and m_B. But now we can connect covariances with fuzzy set intersections (co-occurrence) in the following way.

The fundamental result needed to establish this connection is the familiar formulas for the covariance and variance:

$$\text{cov}(m_A, m_B) = \sum m_A m_B / N - \left(\sum m_A / N\right)\left(\sum m_B / N\right)$$

and

$$\mathrm{var}(m_A) = \sum m_A^2/N - \left(\sum m_A/N\right)^2.$$

These equalities hold for sample estimates if we multiply the right-hand side of each equation by $N/(N-1)$. The importance of the first equation lies in the $\Sigma m_A m_B/N$ term, which is the mean intersection of fuzzy sets A and B using the product operator (i.e., the proportional cardinality of that intersection). In other words, the covariance of two membership functions is the difference between their intersection (co-occurrence!) and the expected intersection if the sets were statistically independent.

Therefore, the co-occurrence of A and B as measured by $\Sigma m_A m_B/N$ may be written as

$$\mathrm{cov}(m_A, m_B) + \bar{m}_A \bar{m}_B,$$

or for sample estimates,

$$(N-1)\mathrm{cov}(m_A, m_B)/N + \bar{m}_A \bar{m}_B. \qquad [6.3]$$

Any linear transformation of the original scales for A and B into fuzzy set membership functions for A and B will preserve the original covariation between A and B, thereby providing measures of co-occurrence that are simple functions of the moments of the original distributions on A and B. That said, it is important to bear in mind that covariance and co-occurrence both are not dimensionless (scale-independent) measures in the sense that the correlation coefficient is. This state of affairs is consistent with the general characterization of fuzzy sets, which do indeed depend on how their membership functions are defined.

We have established a measure of co-occurrence that is congruent with covariation. A somewhat less obvious connection may be made between a generalized version of probability, odds (and odds-ratios), and covariance. To begin with, in Chapter 4, we saw that proportional cardinality shares some important properties with probabilities.

In the notation of Chapter 4, $\|A\| = \bar{m}_A$. We can therefore write the generalized odds as

$$\Omega_A = \bar{m}_A/(1 - \bar{m}_A). \qquad [6.4]$$

To simplify notation, let $\bar{m}_{AB} = \Sigma m_A m_B/N$. Then the conditional odds of A given B is

$$\Omega_{A|B} = \bar{m}_{AB}(\bar{m}_B - \bar{m}_{AB}). \qquad [6.5]$$

Note that the conditional odds may also be expressed in terms of covariance (although that is more cumbersome). Finally, the generalized cross-product ratio $\Omega_{A|B}/\Omega_{A|\sim B}$ is

$$\Omega_{A|B}/\Omega_{A|\sim B} = \frac{\bar{m}_{AB}(1 - \bar{m}_A - \bar{m}_B + \bar{m}_{AB})}{(\bar{m}_B - \bar{m}_{AB})(\bar{m}_A - \bar{m}_{AB})}. \qquad [6.6]$$

These definitions lend meanings to functions of covariances and means that heretofore have been unavailable.

Thus, we may address questions concerning co-occurrence that cannot be addressed via covariation on its own, but in ways that are based on functions of covariances if we use the product operator. These include the following:

- What is the average membership in one category but not any of the others (i.e., with co-occurrence in other categories removed)?
- What is the average co-occurrence of more than two categories?
- What is the average membership in compound categories (e.g., *A* or *B* and not *C*)?
- How does the distribution of co-occurrence in one sample compare with that in another?

We now investigate the co-occurrence of more than two fuzzy sets and its relation to covariance. Perhaps the most intuitive way to think of the intersection of three fuzzy sets using the product operator is in terms of the covariance between one fuzzy set and the product of the other two. For example, from Formulas 6.2 and 6.3, we obtain

$$\begin{aligned} \mathrm{cov}(m_A, m_{BC}) &= \bar{m}_{ABC} - \bar{m}_A\bar{m}_{BC} \\ &= \bar{m}_{ABC} - \bar{m}_A(\mathrm{cov}(m_B, m_C) + \bar{m}_B\bar{m}_C), \end{aligned}$$

so that

$$\begin{aligned} \bar{m}_{ABC} &= \mathrm{cov}(m_A, m_{BC}) + \bar{m}_A\bar{m}_{BC} \\ &= \mathrm{cov}(m_B, m_{BC}) + \bar{m}_A(\mathrm{cov}(m_B, m_C) + \bar{m}_B\bar{m}_C). \end{aligned} \qquad [6.7]$$

As before, for sample estimates, the right-hand side of the upper equation must be multiplied by $N/(N-1)$. We can, of course, obtain corresponding formulas for $\mathrm{cov}(m_B, m_{AC})$ and $\mathrm{cov}(m_C, m_{AB})$. It is easy to build up expressions for the product intersection of any number of fuzzy sets in terms of covariances and means. However, it is clear that multiple-set intersections cannot be reduced to strictly bivariate terms and may give inconsistent estimators.

There are two additional issues that should be addressed before moving on to illustrations. First, as noted before, we may choose between the min and product operators for measuring co-occurrence or comorbidity. The product operator has direct connections with covariance, and we have seen earlier that it has other advantages. Nonetheless, the min operator has an important interpretive advantage in relation to co-occurrence, namely, that it corresponds exactly to the notion that someone has all of the syndromes $A_1, A_2, \ldots, A_K$ to at least degree X. It also declines less inevitably than the product operator as the number of syndromes is increased, thereby making some kinds of comparisons easier to interpret. For instance, someone who has a membership of .8 in five syndromes would be accorded a comorbidity of .8 using the min operator but only $(.8)^5 = .328$ using the product operator. An obvious corrective measure would be to use the geometric mean instead of the raw product, although for the purposes of this chapter, we will use the product.

The second issue concerns the use of raw comorbidities versus "relative" comorbidities normed against the occurrence rates of the individual syndromes. Relative comorbidities are identical to the inclusion coefficient I_{AB} defined by Formula 5.3 in Chapter 5, that is, the proportion of Set B included in the intersection of A and B. Likewise, the covariance may be used as a measure of comorbidity relative to the co-occurrence expected if A and B are statistically independent. Consequently, both I_{AB} and the covariance are useful for comparisons of comorbidities between samples whose syndrome rates differ systematically.

➢ Example 6.3: Return of Fear and Loathing

We return to Example 5.4 from Chapter 5 and the dislike, disgust, and fear ratings of 31 noxious stimuli from 262 undergraduate students. We have already seen that comorbidity as measured using the min operator reveals quite a different pattern from the correlations, indicating that dislike strongly includes disgust and fear. Here we will use the same example to illustrate the relationships between covariance, product operator comorbidity, and generalized odds.

The top part of Table 6.1 displays the mean memberships in the three fuzzy sets along with the mean product intersection memberships. The bottom part displays the covariances. Equation 6.1 connects these quantities. For example, setting $A = Fear$ and $B = Dislike$,

$$\mathrm{cov}(m_A, m_B) = (262/261)[.1395 - (.2315)(.5633)] = .0092.$$

TABLE 6.1
Covariances Among *Fear*, *Disgust*, and *Dislike*

	Mean Product Membership (Comorbidity)		
	A = Fear	*B = Dislike*	*C = Disgust*
Fear	0.2315		
Dislike	0.1395	0.5633	
Disgust	0.0696	0.1302	0.2152
	Covariances		
	Fear		
	0.0092	*Dislike*	
	0.0199	0.0090	*Disgust*

Table 6.1 provides all that we need to assess generalized odds and odds-ratios. For example, from Formulas 6.5 and 6.6, we have

$$\begin{aligned}\Omega_{B|A} &= .1395/(.2315 - .1395) = 1.5181, \text{ and} \\ \Omega_{B|\sim A} &= (.5633 - .1395)/(1 - .2315 - .5633 + .1395) \\ &= 1.2290, \text{ and} \\ \Omega_{B|A}/\Omega_{B|\sim A} &= 1.2353.\end{aligned}$$

That is, the generalized odds of *Dislike* given *Fear* is about 1.24 times higher than that of *Dislike* given *~Fear.*

Because there is overlap among the three fuzzy sets, it is reasonable to assess the extent of exclusive membership in each of them. That is, we want to compare the cardinalities of *A*, *B*, and *C* with those of $A \cap \sim B \cap \sim C$, $B \cap \sim A \cap \sim C$, and $C \cap \sim A \cap \sim B$. In Chapter 5, we already saw that *Dislike* strongly includes both *Fear* and *Disgust*, so we should expect that $B \cap \sim A \cap \sim C$ will be closer to *B* in size than $A \cap \sim B \cap \sim C$ and $C \cap \sim A \cap \sim B$ will be to *A* and *C*, respectively. In Table 6.2, we can see that these expectations are confirmed.

➢ Example 6.4: Psychological Pathologies Among Children

Heubeck (2001) has developed a model of child psychological pathological syndromes constructed from Achenbach's (1991) widely used Child Behavior Checklist (CBCL), based on a large sample of 3,712 children who have been referred to a clinic for psychological and/or behavioral disturbances, and 3,400 children from the same regions who have not been

TABLE 6.2
Mean Exclusive Membership in *Fear*, *Dislike*, and *Disgust*

	Focal Set	*Joint Compl.*
Fear	0.2315	
~Dislike & ~Disgust	0.0674	0.3517
Dislike	0.5633	
~Fear & ~Disgust	0.3387	0.6629
Disgust	0.2152	
~Fear & ~Dislike	0.0605	0.3448

referred to a clinic. The seven syndromes in Heubeck's final model are as follows: A_1 = Withdrawn, A_2 = Somatic, A_3 = Anxious/Depressed, A_4 = Thought-Disorder, A_5 = Attentional Deficit, A_6 = Delinquency, and A_7 = Aggression.

The factor scores for the seven syndromes were converted into fuzzy membership functions in a way that combines clinical psychological conventions and diagnostic criteria. One conventional approach to categorical diagnostic cutoffs is to use a high percentile point (e.g., 95%) on the general sample distribution. An alternative is to use logistic regression to predict whether a case is referred or nonreferred, and use the scale point where p(referred) = 1/2 as the cutoff. The compromise used here is to assign the logistic regression cutoff as the upper limit for 0 membership and the 95th percentile in the nonreferred sample as the neutral point (where membership is 1/2). For example, the A_1 factor has a 95th percentile score of X_1 = 0.5726 in the general population. A logistic regression predicting clinical versus general sample yields

$$\ln(p/(1-p)) = 0.1 + 2.11X_1,$$

where $p = p$(referred), which implies that when $p = 1/2$, then $X_1^{(0)} = -0.0452$. The 95th percentile in the nonreferred sample is $X_1^{(N95)} = 0.5726$. A linear filter membership function based on these benchmarks is

$$m_1 = \max[0, \min((X_1 - X_1^{(0)})/2(X_1^{(N95)} - X_1^{(0)}), 1)].$$

The result in this instance is a "window" from approximately the 78th to the 95th percentiles of the general sample corresponding to membership increasing from 0 to 1/2, and then rising from 1/2 to 1 as we move from the 20th to the 79th percentiles of the clinical sample.

We first examine the correlations between these membership functions for the referred and nonreferred samples, which are displayed in Table 6.3.

TABLE 6.3

Correlations for Nonreferred and **Referred** Samples

	Withdrawn	Somatic	Anx./Dep.	Thought	Attention	Delinq.	Aggres.
Withdrawn	1	**0.3377**	**0.6117**	**0.4818**	**0.2873**	**0.1286**	**0.2474**
Somatic	0.3123	1	**0.5224**	**0.5077**	**0.1244**	**0.2084**	**0.2107**
Anx./Dep.	0.5861	0.4551	1	**0.5690**	**0.1422**	**0.1867**	**0.2090**
Thought	0.4455	0.4646	0.4804	1	**0.4406**	**0.2304**	**0.3759**
Attention	0.4632	0.2686	0.3161	0.5615	1	**0.1836**	**0.5316**
Delinq.	0.1949	0.2823	0.2474	0.3021	0.2534	1	**0.1757**
Aggres.	0.3953	0.3062	0.3075	0.4372	0.5279	0.2306	1

TABLE 6.4

Means for Nonreferred and **Referred** Samples

	Means	*Means*	*Ratios*
Withdrawn	0.0743	**0.4974**	6.6955
Somatic	0.0909	**0.3107**	3.4164
Anx./Dep.	0.0757	**0.5185**	6.8488
Thought	0.0686	**0.5279**	7.6994
Attention	0.0725	**0.5672**	7.8178
Delinq.	0.0594	**0.2761**	4.6441
Aggres.	0.0681	**0.5535**	8.1245

The majority of the correlations from the nonreferred sample are larger than the corresponding correlations from the referred sample (13 out of 21). A similar picture emerges if we use Kendall's tau correlations instead of Pearson correlations. Although most of these differences are not great, these correlations could lead us to conclude that pairwise comorbidity is somewhat higher in the nonreferred sample. This conclusion would be greatly mistaken.

Table 6.4 shows the means for the two samples and compares them with ratios. As should be expected, the mean membership in each syndrome is substantially higher for the referred than the nonreferred sample. Table 6.5 shows the pairwise comorbidities (using the product operator) for both samples and compares them with ratios as well. Each of the comorbidity ratios is higher than the ratios of either of its constituent syndromes, so clearly both the raw and relative comorbidity rates for the referred sample are higher than those for the nonreferred sample.

In fact, all three kinds of comorbidity measure are markedly higher in the referred sample. The average raw comorbidity ratio in Table 6.5 is 14.737 (they range from 8.183 to 20.527), and it turns out that the average relative

TABLE 6.5
Pairwise Comorbidities and Ratios for Nonreferred and **Referred** Samples

Comorbidities

	Withdrawn	Somatic	Anx./Dep.	Thought	Attention	Delinq.	Aggres.
Withdrawn	1	**0.1957**	**0.3494**	**0.3338**	**0.3247**	**0.1509**	**0.3124**
Somatic	0.0174	1	**0.2281**	**0.2284**	**0.1920**	**0.1046**	**0.1990**
Anx./Dep.	0.0262	0.0223	1	**0.3623**	**0.3162**	**0.1638**	**0.3199**
Thought	0.0205	0.0217	0.0216	1	**0.3674**	**0.1710**	**0.3508**
Attention	0.0220	0.0159	0.0167	0.0246	1	**0.1767**	**0.3969**
Delinq.	0.0097	0.0128	0.0111	0.0120	0.0112	1	**0.1723**
Aggres.	0.0186	0.0163	0.0156	0.0192	0.0232	0.0101	1

Comorbidity Ratios

	Withdrawn	Somatic	Anx./Dep.	Thought	Attention	Delinq.	Aggres.
Withdrawn	1	11.2181	13.3414	16.3212	14.7685	15.5718	16.7976
Somatic		1	10.2231	10.5201	12.0800	8.1825	12.1933
Anx./Dep.			1	16.7753	18.9229	14.7207	20.5270
Thought				1	14.9588	14.2156	18.2309
Attention					1	15.7146	17.0892
Delinq.						1	17.0961
Aggres.							1

comorbidity ratio is 2.394 (ranging from 1.366 to 3.682). Finally, the covariances also are higher for the referred sample, with an average ratio of 3.392 (ranging from 1.699 to 5.399). Thus, even taking into account its higher syndrome rates, the referred sample exhibits higher pairwise comorbidity rates than the nonreferred sample. This flatly and correctly contradicts the “comorbidity” comparison that relies on correlations instead of co-occurrence, and highlights the benefit of using the fuzzy set approach to assessing comorbidity.

The benefits of the fuzzy set approach become even more apparent when we ask about comorbidity involving more than two syndromes. For illustration, consider the six- and seven-syndrome comorbidities in the two samples displayed in Table 6.6. In this case (again for illustrative purposes), we have used the min operator for these comorbidities. This table reveals two striking trends. First, the tendency to greater comorbidity in the referred sample is upheld here and in fact appears somewhat more consistent than in the pairwise comorbidities. The average ratio of the raw comorbidities in this table is 18.068, and the average ratio of the relative comorbidities

TABLE 6.6
Six- and Seven-Syndrome Comorbidities
for Nonreferred and **Referred** Samples

Multi-syndrome	*Raw*	***Raw***	*Ratio*	*Relative*	***Relative***	*Ratio*
7-syndrome	0.0047	**0.0843**	17.9515	0.0516	**0.1486**	2.8782
All but *withd*	0.0051	**0.0935**	18.1670	0.0566	**0.1648**	2.9127
All but *som*	0.0051	**0.1110**	21.6250	0.0678	**0.1956**	2.8865
All but *anxdep*	0.0053	**0.0888**	16.6792	0.0585	**0.1565**	2.6742
All but *thought*	0.0049	**0.0869**	17.8615	0.0535	**0.1532**	2.8638
All but *att*	0.0054	**0.0935**	17.2009	0.0598	**0.1690**	2.8258
All but *delin*	0.0078	**0.1382**	17.6638	0.0861	**0.2437**	2.8320
All but *agg*	0.0054	**0.0940**	17.3960	0.0594	**0.1657**	2.7891

is 2.833. These averages are slightly higher and the variation around them considerably less.

The reason for this is that a large majority of cases with nonzero membership in any combination of six syndromes also has nonzero membership in seven syndromes, as indicated by the similarity between the means for the seven-syndrome membership and most of the six-syndrome combinations in Table 6.6. The major exception is delinquency, where there is a sizeable increase in mean comorbidity for the six remaining syndromes. And that leads us to the second clear trend revealed by the fuzzy set approach, namely, that despite the difference in comorbidity rates between the two samples, both of them contain a minority of cases with multiple syndromes. This tendency is especially marked for six of the seven syndromes other than delinquency.

Likewise, a valuable cross-check on comorbidity comparisons between samples is to compare "single" syndrome rates. For example, the degree to which a person has Syndrome A_1 but not any of the others may be measured by $\min(mA_1, 1 - \min(m_{A2}, m_{A3}, \ldots, m_{A7})$. In Table 6.7, we can see that the raw single-syndrome means are rather similar for the referred and nonreferred samples, with the referred sample means being a bit higher. However, norming these means by expressing each of them as a proportion of the corresponding raw syndrome mean from Table 6.4 reveals that relative to the raw syndrome rate, the nonreferred sample has a considerably higher proportion of single-syndrome membership. This finding accords with the picture already obtained of higher comorbidity rates in the referred sample, because that would imply lower single-syndrome rates.

Finally, we will briefly illustrate how comorbidities as constructed from fuzzy set intersections may be linked with standard statistical techniques for purposes of comparing samples or modeling parameters. Suppose we

TABLE 6.7
Single-Syndrome Means and Proportions for
Nonreferred and **Referred** Samples

	Means	*Proport.*	***Means***	***Proport.***
Withdrawn	0.0457	0.6158	**0.0575**	**0.1155**
Somatic	0.0682	0.7499	**0.0383**	**0.1233**
Anx./Dep.	0.0488	0.6452	**0.0641**	**0.1236**
Thought	0.0392	0.5711	**0.0495**	**0.0938**
Attention	0.0430	0.5930	**0.0818**	**0.1442**
Delinq.	0.0469	0.7896	**0.0502**	**0.1818**
Aggres.	0.0429	0.6303	**0.0786**	**0.1419**

wish to compare the comorbidity rates for the nonreferred and referred samples in the seven-syndrome set. As we already know, the referred sample has a much higher CDF than the nonreferred. If we consider degree of membership as a "hazard" variable and the reverse CDF as a "survival" rate, then a standard Kaplan-Meier analysis with a log-rank test for comparing the two distributions yields a chi-square value of 1101.61, indicating that the two reverse CDFs differ considerably.

However, we can model the two samples' distributions by the censored-distribution approach briefly described in Chapter 4. We will apply a two-limit tobit model (cf. Long, 1997, pp. 205–212) to these data by positing a latent variable y that is censored at 0 and 1. For censored observations,

$$\Pr(y_i \leq 0|x_i) = \Phi(-\beta x_i/\sigma_i) \text{ and}$$
$$\Pr(y_i \leq 1|x_i) = 1 - \Phi(1 - \beta x_i/\sigma_i),$$

where Φ is the standardized normal CDF, $x_i = 0$ if the ith case is nonreferred and 1 if referred, and σ_i is either the standard deviation for the nonreferred sample or for the referred sample. For uncensored observations,

$$y_i = \beta x_i + \varepsilon_i,$$

where ε_i is distributed $N(0, \sigma_i)$.

The null model is greatly improved by a homoscedastic model ($\chi^2(1) = 699.60$, $p < .0001$). Including heteroscedasticity does not significantly improve model fit ($\chi^2(1) = 0.106$, $p = .745$). The final model yields a censored distribution for the nonreferred sample whose mean is –0.684 and one for the referred sample whose mean is –0.140, both of which have a standard deviation of 0.351. The standard errors for these coefficients are

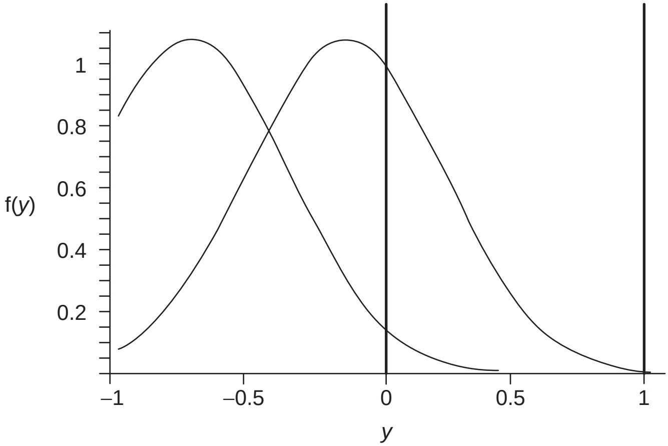

Figure 6.2 PDFs of Seven-Syndrome Membership for Referred and Nonreferred Samples

around 0.02 because of the large sample size. The censoring at 0 yields a spike comprising 97.43% of the cases in the nonreferred sample (compared with 97.50% in the data) and 65.47% of the cases in the referred sample (compared with 65.87% in the data). The censoring at 1 underestimates the spikes slightly, with a spike comprising 0.00008% in the nonreferred sample (compared with 0.147% in the data) and 0.059% of the cases in the referred sample (compared with 0.350% in the data). Figure 6.2 displays the latent PDFs for the final model.

One interpretation of this result is that under the min operator, the membership function of the seven-syndrome set can be seen as a genuine bandwidth filter rather than as a fuzzy set dilator. For another example that is more like a dilator, consider the set of children who are both withdrawn and anxious/depressed (syndromes A_1 and A_3). The fuzzy intersection using the min operator can be fitted with a homosecedastic two-limit tobit model with a distribution for the nonreferred sample whose mean is –0.408 and one for the referred sample whose mean is 0.278, both of which have a standard deviation of 0.403. However, this intersection is heteroscedastic ($\chi^2(1) = 11.126$, $p = .0009$). For the nonreferred sample, the mean is –0.493 and the standard deviation is 0.473, whereas for the referred sample the mean is 0.279 and the standard deviation is 0.392. This finding suggests that the membership function for $A_1 \cap A_3$ behaves more like a dilator than a simple bandwidth filter.

6.3 Multiple and Partial Intersection and Inclusion

We have already encountered the concept of "partialing out" membership in a set when we computed single-syndrome rates in Example 6.4, that is, the degree to which a person has one syndrome but not any of the others. One additional component leads to the concept of partial intersection. Let $\{A_i\}$ be a collection of sets for $j = 1, \ldots, J$ and for a sample of N observations, denote the membership of the ith case in the jth set by m_{ij}. The *partial intersection* of Sets A_g and A_h with the other sets removed is

$$m_{igh.J} = \min(\min(m_{ig}, m_{ih}), 1 - \max_{iJ}), \quad [6.8]$$

where $max_{ij} = \max_{j \neq g,h}(m_{ij})$.

In English, this is the intersection between Sets A_g and A_h and the complements of the other sets. When $g = h$, then we return to our single-syndrome example.

Part and partial inclusion coefficients are based on those discussed in pages 162–163 of Smithson (1987) and Smithson (1994, pp. 17–18) and use inclusion measures rather than fuzzy logical implication. *Partial inclusion* measures the proportion of a focal set that intersects a second set but not the union of a collection of other sets:

$$I_{hg/J} = \sum_i m_{igh.J} \Big/ \sum_i m_{ig}. \quad [6.9]$$

Part inclusion, on the other hand, measures the proportion of the smallest part of the focal set with the union of other sets represented by max_{iJ} removed, which intersects with part of the second set but not the union of a collection of other sets:

$$I_{hg.J} = \sum_i m_{igh.J} \Big/ \sum_i \min(m_{ig}, 1 - \max_{iJ}). \quad [6.10]$$

Another useful ratio is the partial intersection divided by the raw intersection:

$$I_{hg/J} = \sum_i m_{igh.J} \Big/ \sum_i \min(m_{ig}, m_{ih}). \quad [6.11]$$

Multiple intersection, broadly speaking, refers to the intersection of any collection of sets, be they elementary or compound. We have already encountered multiple intersection in Example 6.4, in the form of six- and seven-syndrome comorbidity rates. A useful special case of multiple intersection is the intersection between a focal set and either the union or intersection of a collection of sets. A corresponding *multiple inclusion* coefficient may be defined simply as the proportion of the focal set included in a multiple intersection.

➢ Example 6.5: Part and Partial Inclusions Among Child Pathology Syndromes

To illustrate the application of the concepts outlined above, we return to the child pathology data introduced in Example 6.4 and focus on the intersection between A_3 = anxiety/depression and A_4 = thought disorder. In line with the investigation of single-syndrome rates, suppose we wish to find out how much of the intersection between A_3 and A_4 is left when membership in the other syndromes is removed. The evidence regarding high multiple-syndrome rates in the referred sample suggests that we might expect relatively little of that intersection to remain in the referred sample compared to the nonreferred sample.

This turns out to be the case. For the referred sample, the mean of the $m_{i34.J}$ is .0598, and for the nonreferred sample, it is .0190. The average partial intersection of A_3 and A_4 is indeed greater for the referred sample. However, the mean raw intersection of A_3 and A_4 (using the min operator) is .3958 for the referred sample and .0311 for the nonreferred sample. From Formula 6.11, we have $I_{43|J} = .0598/.3958 = .1511$ for the referred sample and $I_{43|J} = .0190/.0311 = .6109$. Thus, only about 15% of the intersection remains for the referred sample when membership in the other syndromes is removed, whereas about 61% remains for the nonreferred sample. In the nonreferred sample, the intersection between anxiety/depression and thought disorder tends not to involve any other syndromes, whereas in the referred sample, it clearly does.

Another way to make this comparison is via partial inclusion coefficients. The mean membership in anxiety/depression is .5185 for the referred sample and .0757 for the nonreferred sample. For the referred sample, the proportion of anxiety/depression included in its intersection with thought disorder is .3958/.5185 = .7634, but with the other syndromes removed, this reduces considerably to $I_{43/J} = .0598/.5185 = .1153$. For the nonreferred sample, the proportion of anxiety/depression included in its intersection with thought disorder is .0311/.0757 = .4108, and with the other syndromes removed, this reduces only moderately to $I_{43/J} = .0190/.0757 = .2510$.

Finally, this example also illustrates how part inclusion coefficients assess an aspect of inclusion that is distinct from partial inclusion. For the referred sample, $\Sigma_i \min(m_{i3}, 1 - max_{iJ}) = .0894$, and for the nonreferred sample, $\Sigma_i \min(m_{i3}, 1 - max_{iJ}) = .0536$. From Formula 6.10, we can see that the referred sample's part inclusion coefficient is $I_{43.J} = .0598/.0894 = .6689$, whereas for the nonreferred sample, $I_{43.J} = .0190/.0536 = .3545$. This pattern closely parallels that in the raw inclusion coefficients, whose values were produced in the preceding paragraph. That is, the inclusion relationship between anxiety/depression and thought disorder remains quite stable when the intersection of both sets with other syndromes is removed from the focal set as well as from the intersection.

7. CONCLUDING REMARKS

We began this book with the observation that many concepts in the social sciences have both a categorical and a dimensional character. We have presented a case that fuzzy set theory combines ideas from set-theoretic and continuous variable perspectives in a systematic and valuable fashion. We would like to reemphasize the potential for employing fuzzy set concepts and techniques in conjunction with other techniques. For example, several discrete latent trait and latent class models involve "graded-response" models (cf. Heinen, 1996, for an excellent survey), and these would seem to be worthwhile additions to the methods of constructing membership functions reviewed in Chapter 3 (see Manton et al., 1994, for an example).

Likewise, although we have mentioned only the random set approach to fuzzy sets in Chapter 3 (though the results on constant inclusion paths from Chapter 5 can be interpreted in this fashion), an entire edifice of fuzzy probability and statistical methods has been constructed using the random set idea as its cornerstone. Buckley's (2003) recent book on applications of fuzzy probability and statistics provides an example of how the combination of fuzzy sets and statistics can augment our capabilities for handling uncertain data.

Some of the fuzzy set concepts presented here are unfamiliar to social scientists, and only time will tell how useful they turn out to be. For example, fuzziness as explicated in Chapter 4 is related to but distinct from variance. Although assessing fuzziness might seem useful for various purposes, it is only by comparing the benefits thereof with the gains from using variance instead that researchers will be able to decide which is best for their purposes. On the other hand, concepts such as inclusion and the fuzzy set approach to comorbidity immediately offer insights and results that are not readily accessible through more traditional analyses. The examples developed in Chapters 5 and 6 amply demonstrate this case.

As suggested in Chapter 3, there is a long-running debate between fuzzy set proponents and critics who believe that probability theory can handle anything to do with uncertainty. We argue that fuzzy set theory has something to offer that probability theory does not, but we would in no way counsel abandoning the use of probability, and we also recognize that there are plenty of applications in which it may not be clear what roles these frameworks should play. Our recommendations on this matter are threefold.

First, consider which approach makes more conceptual sense and leads to the greatest clarity. If a probabilistic model of category membership clarifies a research problem more than a fuzzy degree-of-membership model, then the probabilistic model should be preferred; if the converse is the case, then choose the fuzzy model. Likewise, as we observed in Chapter 5, if prediction is the primary aim, then a regression model may be the best approach, but if inclusion or other set-wise relations, necessity, and/or sufficiency are the principal focus, then a fuzzy inclusion model may be preferable.

Second, when in doubt, try both approaches and compare their performances. This last injunction may be difficult to carry out because different approaches also may bring with them different measures of performance, as is the case with regression versus inclusion models. Nevertheless, as we illustrated with the extensive comorbidity example in Chapter 6, it is possible to comparatively evaluate two different analyses of the same data on at least an intuitive basis by asking which analysis yields the more plausible or insightful results.

Third, it is important to keep in mind the fact that any given mathematical model will often have two distinct components: systematic variation and noise. Fuzzy sets offer new possibilities for translating theory into models that can be tested by augmenting them with modern statistical estimation and model comparison techniques. That is to say, it allows for a novel collection of systematic relationships that can be tested, such as fuzzy set inclusion. However, fuzzy set theory is relatively silent on how to handle error-laden measurement. Treating a membership like a random variable and then adapting existing technology such as maximum likelihood to these new models provides a means to handle noise in a rigorous way.

We intend the material presented in this book to stimulate researchers and methodologists in further developments and applications using fuzzy sets in conjunction with other quantitative and qualitative techniques. We believe that the potential for such developments is enormous, with the scope ranging from traditional hypothesis testing to data mining.

REFERENCES

ACHENBACH, T. M. (1991). *Manual for the Child Behavior Checklist 4–18 and 1991 profile.* Burlington: University of Vermont Department of Psychiatry.

ADCOCK, R., & COLLIER, D. (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review, 95,* 529–546.

AGRESTI, A., & COULL, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician, 52,* 119–126.

ALLISON, P. D. (1978). Measures of inequality. *American Sociological Review, 43,* 865–880.

AMAREL, S. (1984). On the goals and methodologies of work in fuzzy sets theories. *Human Systems Management, 4,* 309.

BÁRDOSSY, A., & DUCKSTEIN, L. (1995). *Fuzzy rule-based modeling with applications to geophysical, biological, and engineering systems.* New York: CRC Press.

BECK, A. T., & STEER, R. A. (1996). *Beck Depression Inventory manual.* San Antonio, TX: Psychological Corporation.

BILGIÇ, T., & TÜRKŠEN, I. B. (2000). Measurement of membership functions: Theoretical and empirical work. In D. Dubois & H. Prade (Eds.), *International handbook of fuzzy sets and possibility theory, vol. 1: Fundamentals of fuzzy sets* (pp. 195–202). Boston: Kluwer Academic.

BINZEL, R. P. (1999). *The Torino Scale.* Available: http//:impact.arc.nasa.gov/torino

BLACK, M. (1937). Vagueness: An exercise in logical analysis. *Philosophy of Science, 4,* 427–455.

BOLLMAN-SDORRA, P., WONG, S. K. M., & YAO, Y. Y. (1993). A measurement-theoretic axiomatization of fuzzy sets. *Fuzzy Sets and Systems, 60,* 295–307.

BOPPING, D. (2003). *Secrecy and service-loyalty in the Australian Defence Force: Toward a psychology of disclosure behaviour.* Unpublished doctoral dissertation, The Australian National University, Canberra.

BROUGHTON, R. (1990). The prototype concept in personality assessment. *Canadian Psychology, 31,* 26–37.

BUCKLEY, J. J. (2003). *Fuzzy probabilities: New approach and applications.* Heidelberg: Physica-Verlag.

BUDESCU, D. V., KARELITZ, T. M., & WALLSTEN, T. S. (2003). Predicting the directionality of probability words from their membership functions. *Journal of Behavioral Decision Making, 16*(3), 159–180.

BURISCH, M. (1993). In search of theory: Some ruminations on the nature and etiology of burnout. In W. B. Schaufeli, C. Maslach, & T. Marek (Eds.), *Professional burnout: Recent developments in theory and research* (pp. 75–93). Washington, DC: Taylor & Francis.

CERIOLI, A. & ZANI, S. (1990). A fuzzy approach to the measurement of poverty. In C. Dagum & M. Zenga (Eds.), *Income and wealth distribution, inequality and poverty* (pp. 272–284). Berlin: Springer-Verlag.

CHELI, B., & LEMMI, A. (1995). A "totally" fuzzy and relative approach to the multidimensional analysis of poverty. *Economic Notes, 24,* 115–134.

CLEVELAND, W. S. (1993). *Visualizing data.* Murray Hill, NJ: AT&T Bell Laboratories.

CONOVER, W. J. (1980). *Practical nonparametric statistics* (2nd ed.). New York: Wiley.

COOMBS, C. H. (1951). Mathematical models in psychological scaling. *Journal of the American Statistical Association, 46,* 480–489.

CROWTHER, C. S., BATCHELDER, W. H., & HU, X. (1995). A measurement-theoretic analysis of the fuzzy logic model of perception. *Psychological Review, 102,* 396–408.

D'AGOSTINO, R. B., & STEPHENS, M. A. (Eds.). (1986). *Goodness-of-fit techniques.* New York: Marcel Dekker.

DECISION SYSTEMS, INC. (1998). DSIGoM V1.01 (Beta): Grade of membership analysis. Raleigh, NC: Author.

DE LUCA, A., & TERMINI, S. (1972). A definition of nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control, 20,* 301–312.

DUBOIS, D., & PRADE, H. (1980). *Fuzzy sets and systems: Theory and applications.* New York: Academic Press.

EFRON, B., & TIBSHIRANI, R. (1994). *An introduction to the bootstrap.* New York: Chapman & Hall/CRC.

FODDY, M., & SMITHSON, M. (1989). Fuzzy sets and double standards: Modeling the process of ability inference. In J. Berger, M. Zelditch, & B. Anderson (Eds.), *Sociological theories in progress: New formulations* (pp. 73–99). Newbury Park, CA: Sage.

GILES, R. (1988). The concept of grade of membership. *Fuzzy Sets and Systems, 25,* 297–323.

GOLDSTEIN, H., RASBASH, J., BROWNE, W., WOODHOUSE, G., & POULAIN, M. (2000). Multilevel models in the study of dynamic household structures. *European Journal of Population: Revue Europeenne de Demographie, 16,* 373–387.

GUPTA, A. K., & NADARAJAH, S. (Eds.). (2004). *Handbook of beta distribution and its applications.* New York: Marcel Dekker.

HEINEN, T. (1996). *Latent class and discrete latent trait models: Similarities and differences.* Thousand Oaks, CA: Sage.

HERSH, H. M., & CARAMAZZA, A. (1976). A fuzzy set approach to modifiers and vagueness in natural language. *Journal of Experimental Psychology: General, 105,* 254–276.

HESKETH, B., PRYOR, R. G., GLEITZMAN, M., & HESKETH, T. (1988). Practical applications and psychometric evaluation of a computerised fuzzy graphic rating scale. In T. Zetenyi (Ed.), *Fuzzy sets in psychology: Advances in psychology* (pp. 425–454). Amsterdam: North-Holland.

HEUBECK, B. G. (2001). *An examination of Achenbach's empirical taxonomy and covariation between syndromes in different sex, age, and clinic status groups.* Unpublished doctoral dissertation, The Australian National University, Canberra.

HISDAL, E. (1988). Are grades of membership probabilities? *Fuzzy Sets and Systems, 25,* 325–348.

HOROWITZ, L. M., & MALLE, B. F. (1993). Fuzzy concepts in psychotherapy research. *Psychotherapy Research, 3,* 131–148.

JACOBY, W. G. (1991). *Data theory and dimensional analysis* (Quantitative Applications in the Social Sciences, Vol. 78). Newbury Park, CA: Sage.

JACOBY, W. G. (1997). *Statistical graphics for univariate and bivariate data* (Quantitative Applications in the Social Sciences, Vol. 117). Thousand Oaks, CA: Sage.

JACOBY, W. G. (1998). *Statistical graphics for visualizing multivariate data* (Quantitative Applications in the Social Sciences, Vol. 120). Thousand Oaks, CA: Sage.

JOHNSON, N. L., KOTZ, S., & BALAKRISHNAN, N. (1995). *Continuous univariate distributions, Vol. 2* (2nd ed.). New York: Wiley.

KARABATSOS, G. (2001). The Rasch model, additive conjoint measurement and new models of probabilistic measurement theory. *Journal of Applied Measurement, 2,* 389–423.

KARABATSOS, G., & ULLRICH, J. R. (2002). Enumerating and testing conjoint measurement models. *Mathematical Social Sciences, 483,* 485–504.

KAUFMANN, A. (1975). *Introduction to the theory of fuzzy subsets, Vol. 1.* New York: Academic Press.

KLIR, G. A., & YUAN, B. (1995). *Fuzzy sets and fuzzy logic.* Englewood Cliffs, NJ: Prentice Hall.

KOCHEN, M., & BADRE, A. N. (1974). On the precision of adjectives which denote fuzzy sets. *Journal of Cybernetics, 4,* 49–59.

KOENKER, R., & HALLOCK, F. K. (2001). Quantile regression: An introduction. *Journal of Economic Perspectives, 15,* 143–156.

KOSKO, B. (1992). *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence.* Englewood Cliffs, NJ: Prentice Hall.

KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. (1971). *Foundations of measurement.* New York: Academic Press.

LAKOFF, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic, 2,* 458–508.

LAZARSFELD, P. F. (1937). Some remarks on typological procedures in social research. *Zeitschrift fur sozialforschung, 6,* 119–139.

LIEBERMAN, E. S. (2000). *Cross national measurement of taxation.* Paper presented at the annual meeting of the American Political Science Association, Washington, DC.

LONG, J. S. (1997). *Regression with categorical and limited dependent variables.* Thousand Oaks, CA: Sage.

MANTON, K. G., WOODBURY, M. A., & TOLLEY, D. H. (1994). *Statistical applications using fuzzy sets.* New York: Wiley.

MARCHANT, T. (2004a). The measurement of membership by comparisons. *Fuzzy Sets and Systems, 148,* 157–177.

MARCHANT, T. (2004b). The measurement of membership by subjective ratio estimation. *Fuzzy Sets and Systems, 148,* 179–199.

MASSARO, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum.

MASSARO, D. W., WELDON, M. S., & KITZIS, S. N. (1991). Integration of orthographic and semantic information in memory retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 277–287.

MICHELL, J. (1990). *An introduction to the logic of psychological measurement.* Hillsdale, NJ: Lawrence Erlbaum.

MICHELL, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88,* 355–383.

MUNCK, G. L., & VERKUILEN, J. (2002). Measuring democracy: Evaluating alternative indices. *Comparative Political Studies, 35,* 5–34.

MUNCK, G. L., & VERKUILEN, J. (2003). *The electoral democracy index*. Unpublished manuscript, University of Southern California, Los Angeles.

NORWICH, A. M., & TÜRKŠEN, I. B. (1982). The fundamental measurement of fuzziness. In R. R. Yager (Ed.), *Fuzzy set and possibility theory: Recent developments* (pp. 49–60). New York: Pergamon.

ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85,* 172–191.

PAOLINO, P. (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis, 9,* 325–346.

PARASURAMAN, R., MASALONIS, A. J., & HANCOCK, P. A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analysing human and machine performance. *Human Factors, 42,* 636–659.

RAGIN, C. C. (2000). *Fuzzy-set social science.* Chicago: University of Chicago Press.

RAGIN, C. C., & PENNINGS, P. (Eds.). (2005). Special issue on fuzzy sets and social research. *Sociological Methods & Research, 33*(4).

RASCH, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: University of Chicago Press.

RAVALLION, M. (2003, April 21). The debate on globalization, poverty, and inequality: Why measurement matters. *The World Bank Group, Working paper 3031.*

ROSSI, P. E., GILULA, Z., & ALLENBY, G. M. (2001). Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association, 96,* 20–31.

SALTELLI, A., TARANTOLA, S., & CAMPOLONGO, F. (2000). Sensitivity analysis as an ingredient of modeling. *Statistical Science, 15,* 377–395.

SANCHEZ, E. (1979). Inverses of fuzzy relations: Applications to possibility distributions and medical diagnosis. *Fuzzy Sets and Systems, 2,* 75–96.

SEITZ, S. T. (1994). Apollo's oracle: Strategizing for peace. *Synthese, 100,* 461–495.

SEITZ, S. T., HULIN, C. L., & HANISCH, K. A. (2001). Simulating withdrawal behaviors in work organizations: An example of a virtual society. *Nonlinear Dynamics, Psychology, and Life Sciences, 4,* 33–66.

SEN, A. (1999). *Development as freedom.* New York: Knopf.

SMITHSON, M. (1982a). Applications of fuzzy set concepts to behavioral sciences. *Journal of Mathematical Social Sciences, 2,* 257–274.

SMITHSON, M. (1982b). On relative dispersion: New solution for some old problems. *Quality and Quantity, 16,* 261–271.

SMITHSON, M. (1987). *Fuzzy set analysis for behavioral and social sciences.* New York: Springer-Verlag.

SMITHSON, M. (1994). *FUZzySTAT v3.1 tutorial manual.* Unpublished manuscript, James Cook University, Townsville, Australia.

SMITHSON, M. (2005). Fuzzy set inclusion: Linking fuzzy set methods with mainstream techniques. *Sociological Methods & Research, 33,* 431–461.

SMITHSON, M., & HESKETH, B. (1998). Using fuzzy sets to extend Holland's theory of occupational interests. In L. Reznik, V. Dimitrov, & J. Kacprzyk (Eds.), *Fuzzy system design: Social and engineering applications: Studies in fuzziness and soft computing, Vol. 17* (pp. 132–152). Berlin: Physica-Verlag.

SMITHSON, M., & ODEN, G. C. (1999). Fuzzy set theory and applications in psychology. In D. Dubois & H. Prade (Eds.), *International handbook of fuzzy sets and possibility theory, Vol. 5: Applications* (pp. 557–585). Amsterdam: Kluwer.

SMITHSON, M., & VERKUILEN, J. (in press). A better lemon-squeezer? Maximum likelihood regression with beta-distributed dependent variables. *Psychological Methods*.

STEENKAMP, J. B. E., & WEDEL, M. (1991). A clusterwise regression method for simultaneous fuzzy market structuring and benefit segmentation. *Journal of Marketing Research, 28,* 385–396.

TABER, C. S. (1992). POLI: An expert system model of U.S. foreign policy belief systems. *American Political Science Review, 86,* 888–904.

THEIL, H. (1967). *Economics and information theory.* Chicago: Rand McNally.

THOMAS, S. F. (1995). *Fuzziness and probability.* Wichita, KS: ACG Press.

TVERSKY, A., & KOEHLER, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567.

UNITED NATIONS DEVELOPMENT PROGRAM. (1999). *Human development report.* CD-ROM 1990–1999. New York: United Nations.

UNITED NATIONS DEVELOPMENT PROGRAM. (2004). *Democracy in Latin America: Towards a citizens' democracy.* New York: United Nations.

VERKUILEN, J. (2005). Assigning membership in a fuzzy set analysis. *Sociological Methods & Research, 33,* 462–496.

WALLSTEN, T. S., BUDESCU, D. V., RAPPOPORT, A., ZWICK, R., & FORSYTH, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348–365.

WATERHOUSE, L., WING, L., & FEIN, D. (1989). Re-evaluating the syndrome of autism in the light of empirical research. In G. Dawson (Ed.), *Autism: Nature, diagnosis, and treatment* (pp. 263–281). New York: Guilford.

WIDIGER, T. A., & CLARK, L. A. (2000). Toward *DSM-V* and the classification of psychopathology. *Psychological Bulletin, 126,* 946–963.

WILSON, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association, 22,* 209–212.

YAGER, R. R. (1979). On the measure of fuzziness and negation, Part I: Membership in the unit interval. *International Journal of General Systems, 5,* 221–229.

ZADEH, L. (1965). Fuzzy sets. *Information and Control, 8,* 338–353.

ZELENY, M. (1984). On the (ir)relevancy of fuzzy sets theories. *Human Systems Management, 4,* 301–306.

ZIMMERMAN, H.-J. (1993). *Fuzzy set theory and its applications* (2nd ed.). Boston: Kluwer Academic.

ZWICK, R., BUDESCU, D. V., & WALLSTEN, T. S. (1988). An empirical study of the integration of linguistic probabilities. In T. Zetenyi (Ed.), *Fuzzy sets in psychology* (pp. 91–126). Amsterdam: North-Holland.

INDEX

ABOUT THE AUTHORS

Michael Smithson is a Reader in the School of Psychology at The Australian National University in Canberra, and received his PhD from the University of Oregon. He is the author of *Confidence Intervals* (Sage, 2003), *Statistics With Confidence* (Sage, 2000), *Ignorance and Uncertainty* (1989), and *Fuzzy Set Analysis for the Behavioral and Social Sciences* (1987); and co-editor of *Resolving Social Dilemmas: Dynamic, Structural, and Intergroup Aspects* (1999). His primary research interests are in judgment and decision making under uncertainty, social dilemmas, applications of fuzzy set theory to the social sciences, and statistical methods for the social sciences.

Jay Verkuilen is a doctoral candidate in the Quantitative Psychology Program at the University of Illinois at Urbana–Champaign, where he also completed a PhD in political science. He has published papers on experimental game theory, measurement, and the development of indexes for measuring democracy. In addition to various collaborations and consulting work, his current research involves applications of the generalized linear mixed model in experimental psychology, particularly in the context of two alternative forced response experiments.