

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 $[0, 1]$:

$$m_A(x) : \Xi \rightarrow [0, 1]. \quad [3.1]$$

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 Türkşen (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 Türkşen (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

<i>Component</i>	<i>Indicator</i>	<i>Lower Goal Post</i>	<i>Upper Goal Post</i>	<i>Membership Between Goal Posts</i>
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:

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

where a is the slope and b is the life expectancy corresponding to $m_H(x) = 1/2$. Setting $a = 0.1$ 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

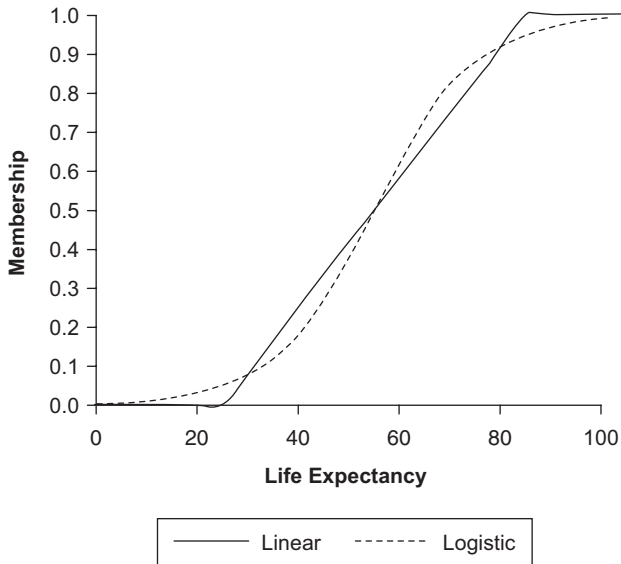


Figure 3.1 Linear Filter and Logistic Membership Functions

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

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

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 $\sim A$ 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 Türkşen (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 Türkşen (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 $\{0 = \text{definite nonmember, possible member, } 1 = \text{definite member}\}$. 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 Türkşen (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 $m_{\neg A}(x) = 1 - m_A(x)$. Moreover, some measurement theorists, such as Bilgiç and Türkşen (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 $m_{\neg A}(2q - x) = m_A(x)$. The minimalist assignment is $\{0 = \text{definite nonmember, possible member, } 1 = \text{definite member}\}$, 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 & Türkşen, 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 $\{x_1, x_2, \dots, x_k\}$ can be ordered in terms of degree of membership in A , so that

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

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

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

For every x_p 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

<i>Transformation</i>	<i>Function</i>	<i>Parameter</i>	<i>Effect on m</i>	<i>Compared to Baseline/Notes</i>
0. Identity	$\text{identity}(m) = m$	—	<ul style="list-style-type: none"> For $k = 1$; all other transforms reduce to the identity 	Exactly the same.
1. Concentration	$\text{conc}(m) = m^k$	$k > 1$, typically 2	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> 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 $.05$	<ul style="list-style-type: none"> 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 $\text{squash}()$	<ul style="list-style-type: none"> Inverts $\text{squash}()$. 	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 $\text{squash}(\cdot)$ 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

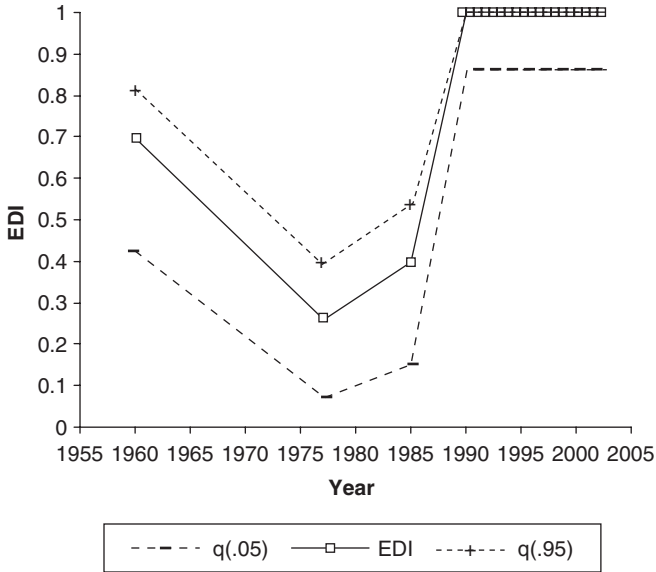


Figure 3.2 Error Bands by Perturbation for EDI Scores for Brazil

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 *vc_score* 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 *vc_score* 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.)

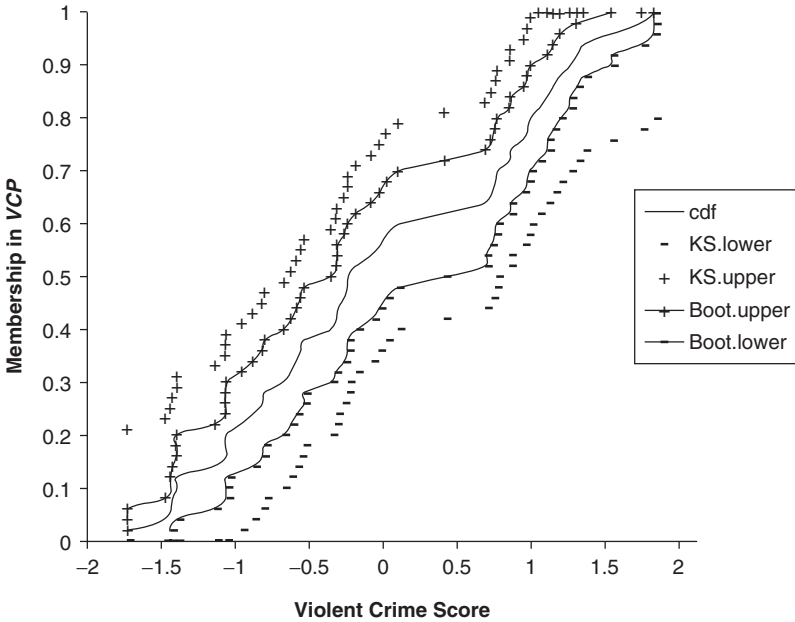


Figure 3.3 Error Bands by Inverting the Kolmogorov-Smirnov Test Versus Bootstrapping

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.