Chapter Two

Literature Review

I. Fuzzy Sets

A. Fuzzy Sets and Crisp Sets

To distinguish between fuzzy sets and classical (nonfuzzy) sets, we refer to the latter as crisp sets, which is widely accepted in the literature (Klir & Yuan, 1995).

1. Crisp Sets

To indicate that an individual object x is a member or element of a set A, we write

$$x \in A.$$

Whenever x is not an element of set A, we write

$$x \notin A.$$

Therefore, a set is defined by a characteristic function that declares whether elements of *X* (the universe of discourse or universal set) are members of the set. Set A is defined by its characteristic function, x_A, as follows:

$$x_A(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \notin A \end{cases}$$

That is, the characteristic function maps elements of X to elements of the set {0, 1}, which is formally expressed by

$$x_A : \mathrm{X} \rightarrow \{0, 1\}.$$

2. Fuzzy Sets

A fuzzy set is a class with a continuum of membership grades (Zadeh, 1965). The membership grades in fuzzy sets are characterized by a membership function that is analogy to the characteristic function of a crisp set. The characteristic function of a crisp set assigns a value of either 0 or 1 to each individual set to discriminate between members and nonmembers, while the membership function of a fuzzy set assigns a value between 0 and 1 to represent membership grades of the element concerned. The membership function of a fuzzy set A is denoted by μ_A; that is,

$$\mu_A : X \rightarrow [0, 1].$$

Membership degrees are fixed only by convention, and the unit interval as range of membership grades, is arbitrary and context-dependent (Dubois et al., 2000.)

B. Fuzzy Numbers

Fuzzy numbers, or fuzzy intervals, are fuzzy sets defined on the set $\Re$ of real numbers. An institutive example of fuzzy number is "numbers that are close to a given number, say 55," illustrated in Figure 2.1. The quantification of fuzzy categories and degrees of membership in fuzzy statistics is represented by the "fuzzy property," which means that information in each category is by nature ambiguous, fuzzy, and tends to "spill over" to the bordering categories. Therefore, we need to put in the categories with "fuzzy numbers" instead of "real numbers." An explanation of fuzzy numbers may be formulated as below (Nguyen & Wu, 2000):

Let U denote a universal set, $\{A_i\}_{i=1}^{n}$ be a subset of discussion factors on U, and $\Lambda(A_i)$ be a level set of A_i for i=1,2,…,n. The fuzzy number of a statement or a term X over U is defined as:

$$\mu_U(X) = \sum_{i=1}^{n} \mu_i(X) I_{A_i}(X).$$

Where $\{\mu_i(X), 0 \le \mu_i(X) \le 1\}_{i=1}^{n}$ are set of membership functions for corresponding factor in $\{A_i\}_{i=1}^{n}$, and . $I_{A_i}(x) = 1$ if $x \in A_i$; $I_{A_i}(x) = 0$ if $x \notin A$. If the domain of the universal set is continuous, then the fuzzy number can be written as: $\mu_U(X) = \int_{A_i \subseteq A} \mu_i(X) I_{A_i}(X)$ 。

Some special fuzzy numbers are proposed to reduce the amount of computational effort. So far, triangular, trapezoidal, and L-R fuzzy numbers are most widely applied to various models. We can interpret the fuzzy number M with a unique peak as a fuzzy quantity "approximately m" and a trapezoidal number may be seen as a fuzzy quantity "approximately in the interval of $[m_1, m_2]$ (Chen & Hwang, 1992). A numerical example was given in Figure 2.2.

1. Triangular and Trapezoidal Fuzzy Numbers

Let x, a, b, c $\in R$. A triangular fuzzy number M (a, b, c) is defined as:

$$\mu_M(x) = \begin{cases} 0, & x \le a \\ (x-a)/(b-a), & a < x \le b \\ (c-x)/(c-b), & b < x \le c \\ 0, & x > c \end{cases} \quad (2.1)$$

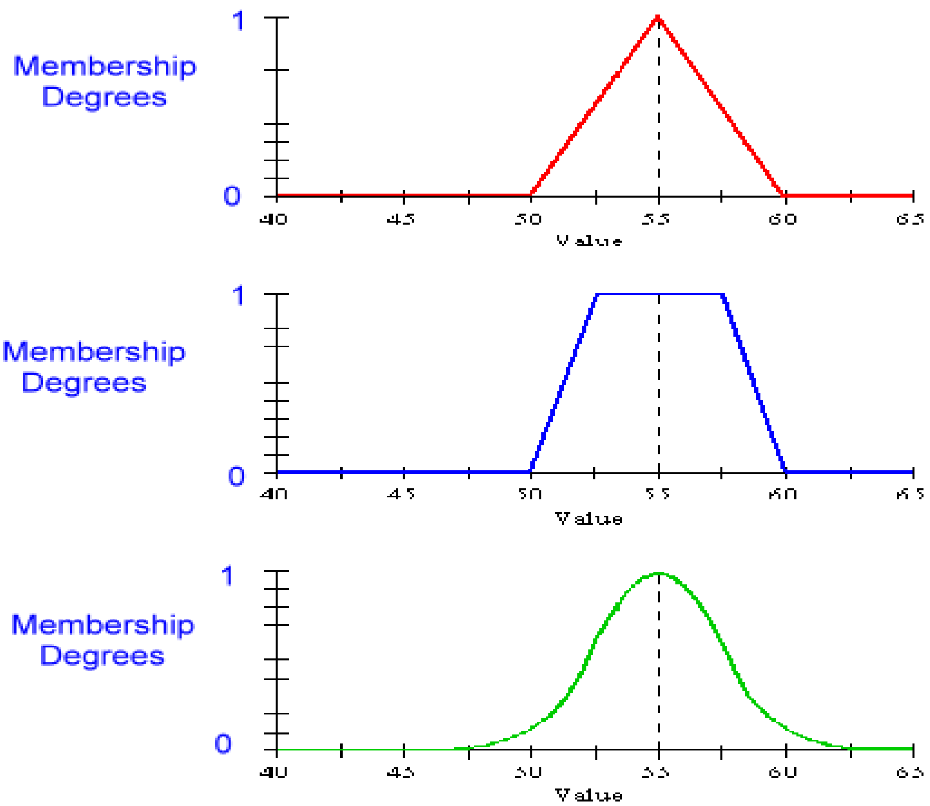

Figure 2.1 Examples of Fuzzy Numbers for Characterizing Real Number Close to 55 in Various Contexts.

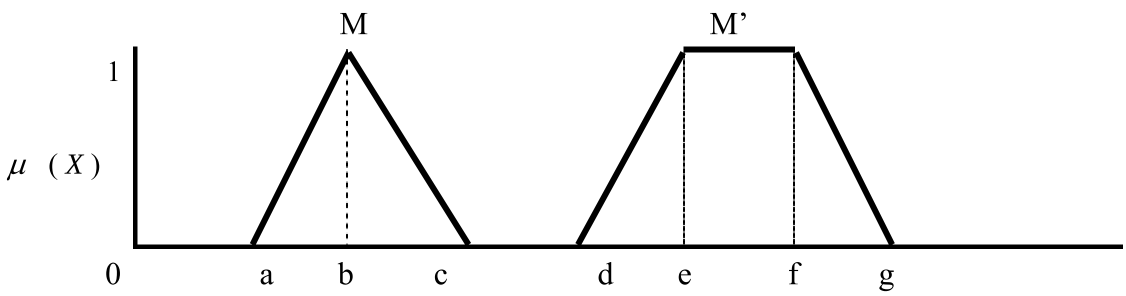

Figure2.2 Triangular Fuzzy Number (M) and Trapezoidal Fuzzy Number (M').

In Figure 2.2, M = (a, b, c) with a and c being the lower and upper bounds, respectively. When there are multiple peaks, fuzzy number M' is denoted as

M'= (d, e, f, g)

with the [e, f] interval being the most likely values for M' and any value above g or below d being totally impossible. The membership value decreases gradually (or linearly) from d to e and f to g.

2. L-R Fuzzy Number

A fuzzy number M(α, m, β) is L-R type if and only if

$$\mu_M(x) = \begin{cases} L((m - x/\alpha), & x \le m, \quad \alpha > 0 \\ R((x - m/\beta), & x \ge m, \quad \beta > 0 \end{cases} \qquad (2.2)$$

where α, β are the left and right "spreads", respectively, and m is the mean of fuzzy number M. When α and β are set to be zero, a fuzzy number M becomes a crisp number m. Moreover, we may say that the triangular and trapezoidal fuzzy numbers are more restricted than L-R ones by constraining all "legs" linear.

C. Fuzzy Variables and Linguistic Variables

A fuzzy variable is characterized by a triple(*X, U, R(X)*) in which *X* is the name of the variable, *U* is the universe of discourse, and *R(X*) is the fuzzy subset *U* that represents a fuzzy restriction imposed by *X* (Lin & Lee, 1999). For instance, *X*="tall" with *U*= {150, 160, 170, 180, 190}, and *R(X*) = 0.1/150+0.4/160+0.7/170+.9/180+1/190 is a fuzzy restriction of "tall".

A linguistic variable is a variable of a higher order than a fuzzy variable, and it takes fuzzy variable as its value. A linguistic variable is characterized by a quintuple (x, T(x), U, G, M) in which x is the name of the variable; T(x) is the term set of x, that is, the set of names of linguistic variables in x with each value being a fuzzy variable defined on U; G is a syntactic rule for generating the names of x; and M is a semantic rule for associating each value of x with its meaning. The following example was given to illustrate fuzzy variables.

If speed is interpreted as a linguistic variable with *U* = [0, 120], that is, *x*="*speed*", then its term set *T(speed)* could be

T(speed) = { Very Sow, Slow, Medium, Fast…}

The semantic rule M could be defined as:

M (Slow) = the fuzzy set for "a speed below about 40 kilometers per hour (kph)" with membership function μ_{Slow}.

M (Medium) = the fuzzy set for "a speed between 60 to 80 kph" with membership function μ_{Medium}.

M (Fast) = the fuzzy set for "a speed above 80 kph" with membership function μ_{Fast}.

These terms can be depicted by fuzzy sets shown in Figure 2.3.

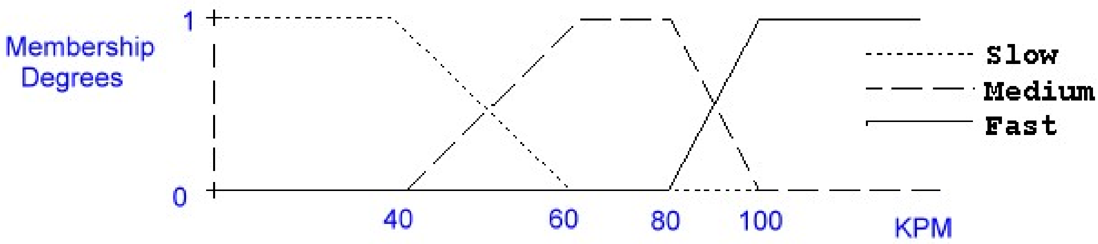

Figure 2.3 Membership Function for a Fuzzy Set of Speed Descriptors.

II. Membership Grades

Since Zadeh (1965) introduced fuzzy sets, one of the main difficulties has been the meaning and eliciting of membership grades. A set is a rather abstract concept and set membership does not mean the same thing in every context; therefore, what fuzzy sets mean is left to the sagacity of users. Fuzzy set opponents criticize that no uniform meaning and measurement of membership grades exist. Our claim is that, far from being weakness, the existence of manifold meanings of membership grades provides the potential richness of the concept of fuzzy set, and explains its pervasiveness in many related fields of investigations (Dubios, Ostasiewicz, & Prade, 2000).

Membership grades play important roles in three types of information-driven task: classification and data analysis, decision-making, and approximate reasoning. From the three basic tasks investigated by many researchers, we may correspond and exploit four semantics of membership grades, respectively in terms of similarity, likelihood, random sets, utility, preference, and measurement view (Bilgic & Turksen, 2000; Dubios et al., 2000; Zimmermann, 1996). For illustration, we first begin with a mathematical definition of a membership function. A fizzy set A has a membership μ_A defined as a function from the well-defined universe, the referential set X, into the unit interval as μ_A: X → [0, 1]. In this section, we adapted a predicate "Achilles(x) is tall (T)" represented by a number in the unit interval, $\mu_T(x)$. There are several possible ways to interpret the question: "What does it mean to say $\mu_T(x)$, the membership degree of Achilles is tall, is 0.9)?" (Bilgic & Turksen, 2000):

1. Similarity view: Achilles' height is away from the prototypical object, which is truly "tall" to the degree 0.1.
2. Likelihood view: 90% of a given population declared that Achilles is tall.
3. Random set view: 90% of a given population described "tall" as an interval containing Achilles' height.
4. Measurement view: When compared with others, Achilles is taller than some and this fact can be encoded as 0.9 on some scales.

The above-mentioned interpretations of membership degrees can be classified into two trends: objective and subjective view. The former include both in likelihood and random set views while the latter includes similarity view. Measurement view, connecting the objective and subjective view, defines problem in both ways depending on the observers who are making the comparisons. Bilgic & Turksen (2000) proposed a comprehensive framework for measurement of membership grades. Adopting their framework, we discussed the topic utilizing the example of "the membership degree of Achilles is tall is 0.9" noted previously.

A. The similarity view

Regarding membership degree $\mu_T(x)$ as the degree of proximity or similarity of x from the prototype element of T has been the oldest interpretation since Zadeh advocated the concept of fuzzy set to pattern recognition in 1966. This view is particularly functional in clustering analysis, regression analysis, and fuzzy control technique. Since notation of similarity, arising from prototype theory, takes membership as a notion of being similar to the representation of a category, it assumes that a perfect prototype exists and requires a metric space on which distance can be defined (Bilgic & Turksen, 2000; Dubios et al., 2000; Tversky, 1977).

Regarding the measurement of membership, Zimmermann (1996) proposed a Type-A model, direct scaling method, and a Type-B model, indirect method. In Type-A model, we ask some subjects directly for membership values. However, this approach may be distorted by a number of response bias and achieve only the level of ordinal scale. Since ordinal judgment provides little information and mathematical operation requires at least interval-quality scale, Thole, Zimmermann, & Zysno suggested, according to Thurstone's "Law of Categorical Judgment", asking subjects to grade membership on a percentage scale (Zimmermann, 1996).

The judgment of membership grades can be regarded as the comparison of object x with an ideal (standard), then resulting in a psychological distance. According to this perspective, a single human being is treated as the measurement device and assumed that fuzziness is assumed to arises from the insufficient cognitive abilities of this person when facing with the task of "comparing the object with a certain prototype or imaginable ideal." This approach naturally leads to a notion of *distance* which Zysno relates to the membership function as(Bilgic & Turksen, 2000):

$$\mu(x) = \frac{1}{1+d_x} \qquad (2.3)$$

where d_x is the distance of object x from the ideal. The distance function has to be specific that a specific monotonic function of the similarity with the ideal could, as first approximation, be

$$d'(x) = 1/x. \quad (2.4)$$

The relation between variation in physical dimension of stimuli (Φ) and their associated responses is generally exponential (Nunnally & Berstain, 1994). Therefore, the distance function may be expressed as:

$$d(x) = \frac{1}{e^{a(x-b)}} . \qquad (2.5)$$

Where *a* is the evaluation unit and *b* is the ideal. Substituting equation (2.5) into (2.4) yields the logistic function:

$$\mu(x) = \frac{1}{1+e^{-a(x-b)}} \qquad (2.6)$$

It is S-shaped. Moreover, *b* is the inflexion point and *a* is the slope of the function.

From the perspective of linear programming, equation (2.6) can be linearized by the transformation:

$$-\ln\frac{1-\mu(x)}{\mu(x)} = \ln\frac{\mu(x)}{1-\mu(x)} = a(x-b) \qquad (2.7)$$

B. **The likelihood view**

According to TEE (threshold, error, assumption of equivalence) model (Hisdal, 1988), fuzziness comes from errors in measurement, incomplete information, and interpersonal contradictions. In view of this, when someone is asked to assess whether a person is tall, he/she, due to imperfect information, may construct an imprecise measurement of the person's height, say 190±10 cm, and a error curve to quantify the possible measurement errors. Therefore, μ_{tall}(190cm)= P(tall │ x= 190cm) = 0.9

C. **The random set view**

In this view, we adapted the probability theory and the crisp set approach, to interpret membership degree. For membership degrees μ_T(x) is 0.9, the random set view is: 90% of the population defined an interval on the referential set, X, as an interval containing x(Achilles' height) on the basis of an evaluation T(e.g. tallness). The remaining 10% defined intervals that excluded x to be T.

Since this view applies a random set approach derived from the crisp set, we shall introduce the notation of "α-cut", which bridges the conceptual gap between crisp and fuzzy set. An α-cut (α-level set) of a fuzzy set A is a crisp set A_α that contains all the elements of the universal set U that have a membership grade in A greater than or equal to α. That is,

$$A_\alpha = \{X \in U \mid \mu_A(x) \geq \alpha\},\ \alpha \in (0.1] \qquad (2.8)$$

The set of all levels $\alpha \in (0.1]$ that represents distinct α-cuts of a given set A is called a level set of A. That is,

$$\Lambda_A = \{\alpha \mid \mu_A(x) = \alpha\},\ \alpha \in (0.1] \qquad (2.9)$$

By the notation of α-cut, we may consider a fuzzy set as a nested family of classical subsets. Dubios et al.(2000) call this vertical view of fuzzy set.

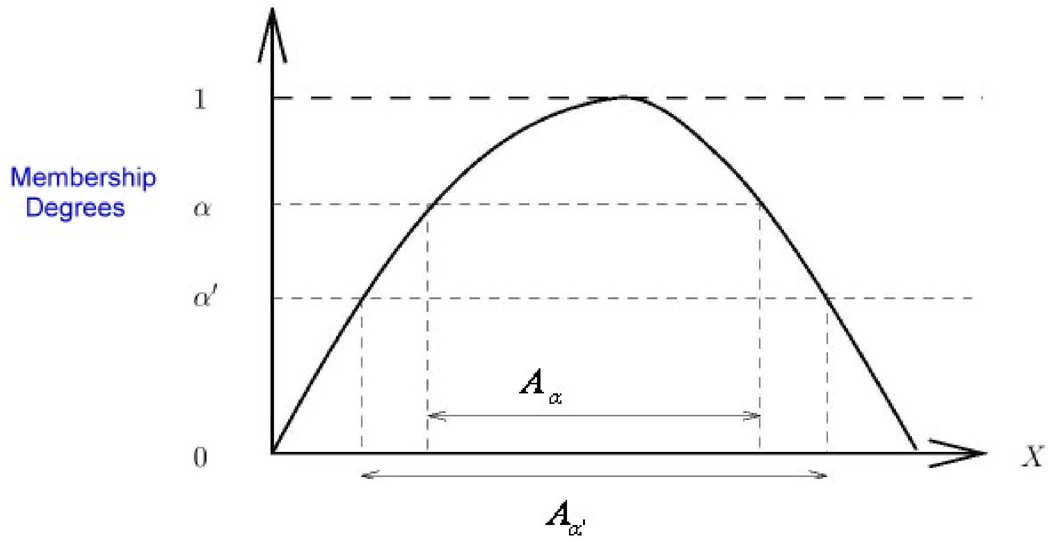

Figure 2.4 α-cut of Fuzzy Set A

The membership function can be stated by an integral form:

$$\mu_A(x) = \int_0^1 \mu\, A_\alpha(x) d\alpha \qquad (2.10)$$

As noted earlier, the random set view, derived from the traditional probability theory, considers membership function as a uniform distributed random set consisting of the Lebesgue measure on [0,1] and the set-valued mapping A_α : (0, 1] →X.

With the random set view, our example could be interpreted as: 90% of the population defined an interval on the referential set, X, as an interval containing x(e.g. x=Achilles' height) on the basis of an evaluation T(e.g. tallness). The remaining 10% defined intervals that excluded x to be T.

D. The measurement view

Measurement theory is about representation and meaningfulness of a particular representation. This theory bridges that gap between abstract mathematical structures and empirical science by taking a phenomenon that can be modeled as an algebraic structure (e.g. fuzziness) and by mapping it into a numerical structure (hence capturing the essential core of the measurement process). Using such theory, one can discuss the representation of a qualitative structure and the meaningfulness of such a representation (Bilgic & Turksen, 2000).

The measurement view is about the measurement of membership. Assuming a single fuzzy set, F, and a finite number of agents, objects, elements, and so on in A. The question is: "To what degree an agent from A belongs to fuzzy set F?". The first measurement problem takes the relation "an agent is more F than another agent", where F is a fuzzy term (typically an adjective). To capture this graded membership concept, consider a binary relation, $\succsim_F$ on A with the following interpretation:

$$a \succsim_F b \text{ belongs to F at least as much as b belongs to F}$$

The structure is assumed to be bounded (i.e., there exists elements a^- and a^+ in A such that for all other a in A that $a^+ \succsim_F a \succsim_F a^-$). Since the boundary condition is satisfied, we may acclaim that at least ordinal scale representation exists for $\langle A, \succsim_F \rangle$ once the transitivity and connectedness of $\succsim_F$ are accepted. Moreover, the measurement can be measured on an interval scale by asking subjects to compare pairs of agents (Bilgic & Turksen, 2000).

III. Rating Scales

A. Definitions

Rating scales are one of the most common methods for collecting data about human attributes, frequency, quality, or agreement. The purpose of a rating scale is to assign an indicator of more-or-less of an attribute using ordered response categories (O'Neill, 2002). An example of rating scale is shown in Table 2.1.

Likert scales do not differ fundamentally from rating scales. In Likert scales, a statement is followed by category response continuum such as strongly agree, agree, undecided, disagree, and strongly disagree. That is to say, Likert scales use the standard set of response options representing the degree of agreement instead of descriptive terms adopted in rating scales (Hopkins et al., 1990). In this study, we treat a Likert scale as a special form of rating scales. The CES-D, shown in Table 2.2, is an example of a Likert scale.

Table 2.1 An Example of Rating Scales: Extracted from the BDI

Sadness		Changes in sleeping pattern	
0	I do not felt sad.	0	I have not experienced any change in my sleeping pattern
1	I felt much of the time.	1	I sleep somewhat less than usual
2	I felt sad all the time.	2	I sleep a lot less than usual
3	I'm so sad or unhappy that I can't stand it.	3	I wake up 1-2 hours early and can't get beck to sleep

Table 2.2 An Example of Likert Scales: Extracted from the CES-D

During the past week:	Rarely	Some of the time	Occasionally	Most of the time
I felt sad.	0	1	2	3
My sleep was restless.	0	1	2	3

B. Scaling of rating scales

There have been two trends in scaling rating scales: classical test theory (CTT) approach and item response theory (IRT) approach. In CTT, "method of successive categories" is adopted in scaling rating scales (Gilford, 1954). It is assumed that the categories of rating scales are in correct order and their boundary lines are stable except for sampling errors. In this method, alternatives of rating scales have been scaled such that units represent increments of equal size. However, this method only achieves the status of ordinal measurements and the assumption of equal distance between adjacent categories is questionable (Gilford, 1954; Wright, 1999; O'Neill, 2002).

Instead of scaling each category in equal-sizesd increments, item response theories, or latent trait theories, have taken different approaches. An item response model specifies a relationship, often in a logistic form, between the observable examinee item performance and the traits assumed to underlie performance on the test. Some models have been proposed to handle polychotomous data, such as Samejima's(1969) graded response model (GRM), Andrich's(1978) rating scale model, Masters' (1982) partial credit model (PCM), and Muraki's (1992) generalized partial credit model(GPCM). The PCM differs from other models in that it belongs to the Rasch family of models and so share the distinct characteristics of that family: separable person and item parameters, sufficient statistics, and, hence, conjoint additivity. These features enable "specifically objective" comparisons of persons and items and allow each set of model parameters to be conditioned out of the estimation procedure for the others (Masters & Wright, 1984). The PCM is illustrated as follows (Masters, 1982; Masters & Wright, 1997):

The PCM is an application of Rasch's model for dichotomies. When an item provides only two scores 0 and 1, the probability of scoring 1 rather than 0 is expected increase with the ability being measured. In Rasch's model for dichotomies, the to probability of person *j* succeeding on item *i* is written as:

$$\frac{P_{ij1}}{P_{ij0}+P_{ij1}}=\frac{\exp(\theta_j-\delta_i)}{1+\exp(\theta_j-\delta_i)} \qquad (2.11)$$

where P_{ij1} is the probability of person *j* scoring 1 on item *i*, P_{ij0} is the probability of person *j* scoring 0, θ_j is the ability of person *j*, and δ_i is the difficulty of item *i* defined as the location on the measurement variable at which a score of 1 on item *i* is as likely as score of 0. The model is written here as a conditional probability to emphasize that it is a model for the probability of person *j* scoring 1 rather than 0.

In PCM, this expectation is modeled using the above-mentioned Rasch model for dichotomies above-mentioned:

$$\frac{P_{ijx}}{P_{ijx-1}+P_{ijx}}=\frac{\exp(\theta_j-\delta_{ix})}{1+\exp(\theta_j-\delta_{ix})}, \quad x=1,2,...,m_i \qquad (2.12)$$

where P_{ijx} is the probability of person *j* scoring *x* on item *i*, P_{ijx-1} is the probability of person *j* scoring *(x-1)*, θ_j is the ability of person *j*, and δ_{jx}, called a "step", is an item parameter governing the probability of scoring *x* rather than *x-1* on item *i*.

The interpretation of "step" is illustrated in Table 2.3. For an item on an attitude questionnaire, "completing the k^{th} step" means choosing the k^{th} response alternative over the $(k\text{-}1)^{th}$ in response to the item. In Table 2.3, a person who chooses to AGREE can be considered to have chosen DISAGREE over STRONGLY DISAGREE (first step taken) and also AGREE over DISAGREE (second step taken), but to have failed to choose STRONGLY AGREE over AGREE (third step rejected).

The third step in item *i* listed in Table 2.2 is from level 2 to level 3. The difficulty of the third step governs how likely it is that a person who has already reached level 2 will complete the third step to level 3. Therefore, the probability of scoring 3 rather than 2 can be expressed as

$$\Phi_{ij3} = \frac{P_{ij3}}{P_{ij2} + P_{ij3}} = \frac{\exp(\theta_j - \delta_{i3})}{1 + \exp(\theta_j - \delta_{i3})}. \qquad (2.13)$$

Likewise, the second step in item i is from level 1 to level 2 since a person cannot make a "0" by failing the second step. Therefore, the probability for a person making a "2" rather than "1" on item i is

$$\Phi_{ij2} = \frac{P_{ij2}}{P_{ij1} + P_{ij2}} = \frac{\exp(\theta_j - \delta_{i2})}{1 + \exp(\theta_j - \delta_{i2})}. \qquad (2.14)$$

Similarly, the first step in item i is to make a "1" rather than a "0":

$$\Phi_{ij1} = \frac{P_{ij1}}{P_{ij0} + P_{ij1}} = \frac{\exp(\theta_j - \delta_{i1})}{1 + \exp(\theta_j - \delta_{i1})}. \qquad (2.15)$$

Finally, as person j must make one of the four possible scores on item i,

$$\sum_{h=0}^{m_i} P_{ijh} = P_{ij0} + P_{ij1} + P_{ij2} + P_{ij3} = 1 \qquad (2.16)$$

Equations (2.13) ~ (2.16) are readily solved to obtain one general expression for the probability of person j scoring x on item i:

$$P_{ijx} = \frac{\exp\sum_{k=0}^{x} (\theta_j - \delta_{ik})}{\sum_{h=0}^{m_j} \exp\sum_{k=0}^{h} (\theta_j - \delta_{ik})} \qquad x = 0, 1, ..., m_i \qquad (2.17)$$

Table 2.3 Pass/fail Scores for A Three-step Item *i*

Person *j*	Performance Levels				Scores X_{ij}
	0 STRONGLY DISAGREE	1 DISAGREE	2 AGREE	3 STRONGLY AGREE	
1	1 → First Step	1 → Second Step	1 → Third Step	1	3
2	1 →	1 →	1		2
3	1 →	1			1
4	1				0