Enhanced Interval Approach for Encoding Words into Interval Type-2 Fuzzy Sets and Convergence of the Word FOUs

Simon Coupland, *Member, IEEE*, and Jerry M. Mendel, *Life Fellow, IEEE* and Dongrui Wu, *Member, IEEE*[1]

***Abstract*— The Interval Approach (IA) [4] is a method for synthesizing an interval type-2 fuzzy set (IT2 FS) model for a word from data that are collected from a group of subjects. A key assumption made by the IA is: each person's data interval is random and uniformly distributed. This means, of course, that the IT2 FS model for the word is random. Consequently, one can question whether or not the IT2 FS model for the word converges in a stochastic sense. This paper focuses on this question. As a part of our study, we have had to modify some steps of the IA, the resulting being an Enhanced IA (EIA). The paper shows by means of some simulations, that the IT2 FS word models that are obtained from the EIA are converging in a mean-square sense. This provides substantial credence for using the EIA to obtain T2 FS word models.**

I. Introduction

Recently, a methodology, called the Interval Approach (IA), was presented by Liu and Mendel [4] for synthesizing an interval type-2 fuzzy set (IT2 FS) model for a word, in which: interval end-point data about a word are collected from a group of subjects (the subjects are asked: On a scale of 0-10, what are the end-points of an interval that you associate with the word _____?); each subject's data interval is mapped into a type-1 (T1) FS; the latter is interpreted as an embedded T1 FS of an IT2 FS; and, an IT2 FS mathematical model is obtained for the word from these T1 FSs.

A key assumption made and justified (in their paper) during the development of the IA is: each person's data interval is random and uniformly distributed (i.e., the end-points of each person's interval establish the parameters of a uniform probability distribution for that interval). This means, of course, that the IT2 FS model for the word is random. Consequently, one can question whether or not the IT2 FS model for the word converges in a stochastic sense. The purpose of this paper is to focus on this question and to demonstrate that the IT2 FS word models that are obtained from the IA converge in a mean-square sense. As with previous study in this area [4] we only focus on adjectives.

The rest of this paper is organized as follows: Section II provides background material about IT2 FSs, the IA and stochastic convergence. Section III outlines the methodology used to collect data for the experiments in this paper. Section IV presents the *enhanced interval approach* method for synthesizing IT2FS models for words. Section V presents an experiment looking at the convergence of these word models. Finally, Section VI draws some conclusions from this work.

[1]The authors' names are ordered alphabetically, each Author made an equal contribution to this work, as such we do not consider there to be a first author.
Simon Coupland is with the Centre for Computational Intelligence, De Montfort University, Leicester, LE1 9BH UK (e-mail: simonc@dmu.ac.uk).
Jerry M. Mendel is with the Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: mendel@sipi.usc.edu).
Dongrui Wu is with the Institute for Creative Technologies and the Signal Analysis and Interpretation Laboratory, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089- 2564 USA (e-mail: dongruiw@usc.edu).

II. Background

In this section some background material is provided that is needed for the rest of the paper.

A. Interval Type-2 Fuzzy Sets

An IT2 FS $\tilde{W}$ [e.g., Mendel [5], [6]] for a primary variable $x(x \in X)$ is characterized by its *footprint of uncertainty*, $FOU(\tilde{W})$, which in turn is completely described by its *lower membership function*, $LMF(\tilde{W})$ [also denoted $\underline{\mu}_{\tilde{W}}(x)$], and *upper membership function* $UMF(\tilde{W})$ [also denoted $\overline{\mu}_{\tilde{W}}(x)$], that are the lower and upper bounding functions of $FOU(\tilde{W})$ respectively. Five generic FOUs are depicted in Fig. 1 – left shoulder (LS), interior and right-shoulder (RS) FOUs; they are the ones that derive from the IA.

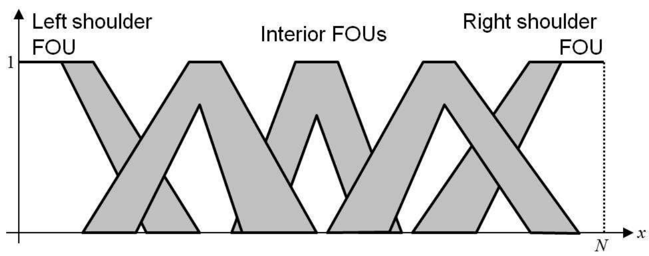


Fig. 1. Left-shoulder, right-shoulder and interior FOUs, all of whose LMFs and UMFs are piecewise linear [4].

It is well known, that $FOU(\tilde{W})$ can be covered by T1

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

FSs[1] (some of which are normal FSs, but most of which are not) called *embedded T1 FSs,* W^i, i.e.:

$$\tilde{W} = \bigcup_{i=1}^{\infty} W^i \quad (1)$$

Equation (1) is often called the Mendel-John wavy-slice representation for a T2 FS [8], specialized to an IT2 FS [7].

Similarity of word IT2 FSs plays a very important role in this paper. A measure of similarity between words must simultaneously provide a measure of the similarity between two word's FOU-shapes and the proximity of the word's FOUs, because word FOUs are ordered on the scale of a primary variable. Wu and Mendel [11] explain why the Jaccard similarity measure, $s_J(\tilde{A}, \tilde{B})$, is most useful for computing the similarity between two word IT2 FSs, $\tilde{A}$ and $\tilde{B}$. Their formula for $s_J(\tilde{A}, \tilde{B})$, used by us to study the convergence of a word's FOU, is:

$$s_J(\tilde{A}, \tilde{B}) = \frac{\sum_{i=1}^{N}[\min(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) + \min(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))]}{\sum_{i=1}^{N}[\max(\overline{\mu}_{\tilde{A}}(x_i), \overline{\mu}_{\tilde{B}}(x_i)) + \max(\underline{\mu}_{\tilde{A}}(x_i), \underline{\mu}_{\tilde{B}}(x_i))]} \quad (2)$$

B. Interval Approach (IA) for Encoding a Word into an IT2 FS

The IA consists of two parts, the *Data part* (Fig. 2) and the *Fuzzy Set (FS) part* (Fig. 3). In the Data part, data intervals $[a^{(i)}, b^{(i)}]$ that have been collected from a group of n subjects ($i = 1, \ldots, n$) are pre-processed, after which data statistics are computed for the m surviving intervals. In the FS part, FS uncertainty measures are established for a pre-specified T1 membership function (MF) [always beginning with the assumption that the FOU is an interior FOU (Fig. 1), and, if needed, later switching to a shoulder FOU (Fig. 1)]. Then the parameters of the T1 MF are determined using the data statistics, and the derived T1 MFs are aggregated using union leading to an FOU for a word, and finally to a mathematical model for the FOU.

Referring to the Data part of the IA, observe in Fig. 2 that preprocessing the n interval end-point data $[a^{(i)}, b^{(i)}]$ ($i = 1, \ldots, n$) consists of *four stages*: (1) bad data processing, (2) outlier processing, (3) tolerance-limit processing, and (4) reasonable-interval processing[2]. For the details of each of these steps, see [4]. As a result of data preprocessing, some of the n interval data are discarded and the remaining m intervals are re-numbered, $1, 2, \ldots, m$. A uniform probability distribution is assigned to *each* of the m surviving data intervals after which the mean and standard deviation are computed for each of them.

The FS part of the IA consists of nine steps (Fig. 3):

[1]On the continuous case: for any point within the FOU, we can find a continuous T1 FS passing through it. As an FOU is the union of all such points, the FOU is completely covered by the continuous T1 FSs.

[2]A data interval is said to be [4] *reasonable* if it overlaps with another data interval – the same word must mean something in common for different people,

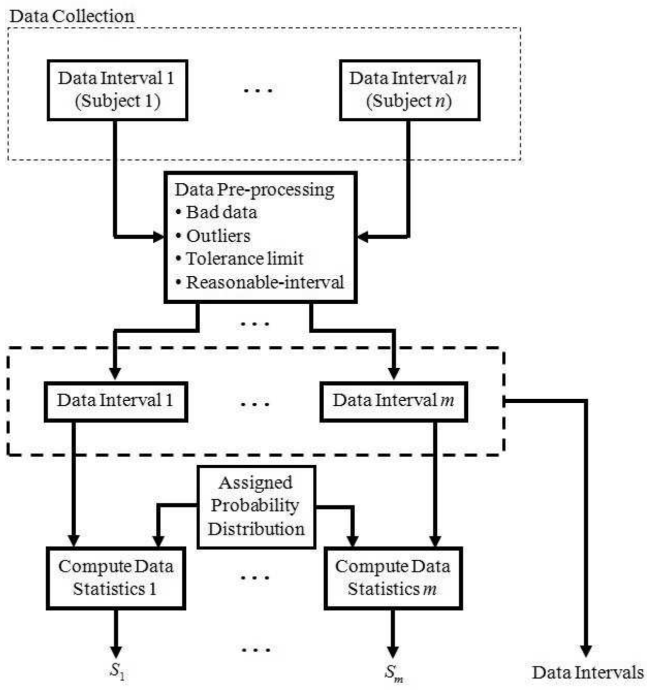


Fig. 2. Data part of the IA approach. Note that the output statistics feed into the Fuzzy Set part of the IA in Fig. 3 [4].

1) Because the mapping from an interval of data to a T1 MF only uses the mean and standard deviation of the assigned uniform probability distribution, only T1 MFs with two degrees of freedom can be used, namely a symmetrical triangle interior T1 MF, a left-shoulder T1 MF, or a right-shoulder T1 MF.
2) The mean and standard deviation are chosen as the uncertainty measures for these deterministic T1 MFs.
3) The mean and standard deviation are computed for these three T1 FSs (see Table II in [4]).
4) The parameters of each T1 FS (triangle, left- or right-shoulder) are computed by equating the mean and standard deviation of that T1 FS to the mean and standard deviation, respectively, of a data interval. Because these results are also used in this paper, they are summarized in Table I.
5) The set of m data intervals, are classified as an interior, left-shoulder or right-shoulder FOU, using the classification diagram[3] depicted in Fig. 4. On that diagram, m_l and m_r are the mean values of the left and right end-points of the surviving m intervals.
6) Once a classification has been made as to the kind of FOU for a specific word, each of the word's remaining

[3]This diagram was obtained by using three simple obvious inequalities: (a) the right-end of a data-interval for a word has to be larger than the left-end of that interval, $b_{MF}^{(i)} > a_{MF}^{(i)}$; (b) the left-end of a symmetric triangle T1 MF has to be greater than or equal to zero, $a_{MF}^{(i)} \geq 0$; and, (c) the right-end of that triangle has to less than or equal to 10, $b_{MF}^{(i)} \leq 10$, and by then using the formulas that are given in the first row of Table I for $a_{MF}^{(i)}$ and $b_{MF}^{(i)}$ in these inequalities.

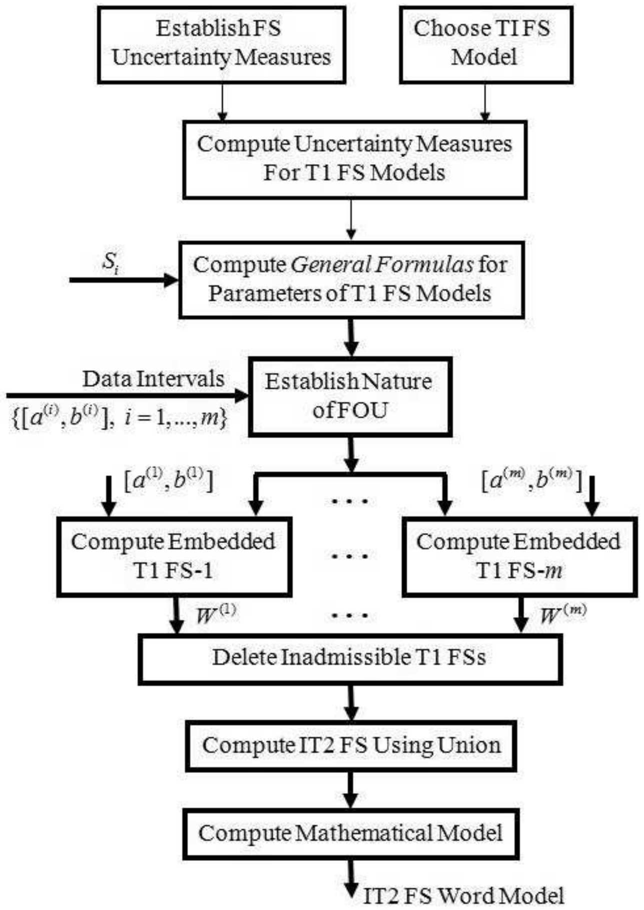


Fig. 3. Fuzzy set part of the IA [4].

m data intervals are mapped into their respective (embedded) T1 FSs using the equations that are given in Table I.

7) It is possible that some of the m embedded T1 FSs are inadmissible, i.e. they violate $a_{MF}^{(i)} \geq 0$ and $b_{MF}^{(i)} \leq 10$, because the FOU classification procedure was based on statistics and not on each realization. Those T1 FSs are deleted, so that there will be m^* remaining embedded T1 FSs, where $m^* \leq m$. The number m^* is the starting point for our convergence study. Using the Wavy Slice Representation Theorem for an IT2FS [8], [7] a word's IT2 FS $\tilde{W}$ is then computed as in (1), where W^i is the just-computed i^{th} embedded T1 FS ($i = 1, 2, \ldots, m^*$).
8) A mathematical model is obtained for $\tilde{W}$ by upper bounding and lower bounding it using piece-wise linear bounds.

A word (see Fig. 1) that is modeled by an interior FOU has an UMF that is a trapezoid and a LMF that is a triangle, but in general neither the trapezoid nor the triangle are symmetrical. A word that is modeled as a left- or right-shoulder FOU has trapezoidal upper and lower MFs; however, the legs of the respective two trapezoids are not necessarily parallel.

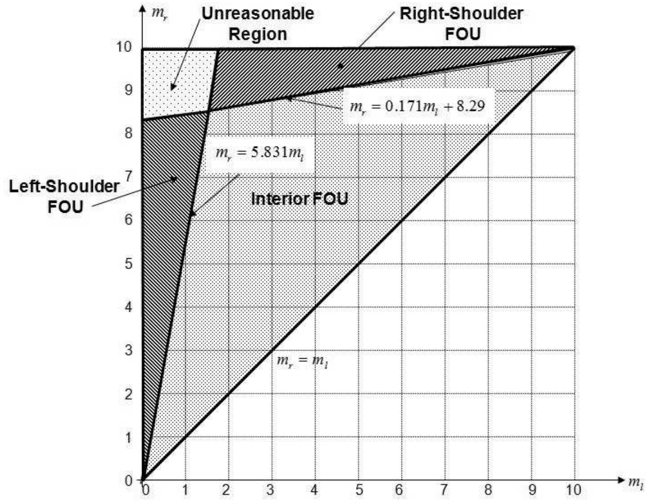
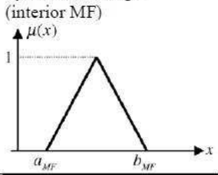
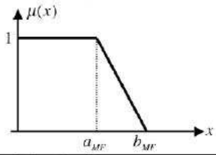
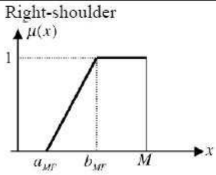


Fig. 4. Classification diagram for the IA [4].

TABLE I
TRANSFORMATIONS OF THE UNIFORMLY DISTRIBUTED DATA INTERVAL $[a^{(i)}, b^{(i)}]$ INTO THE PARAMETERS $a_{MF}^{(i)}$ AND $b_{MF}^{(i)}$ OF A T1 FS [4].

MF	Transformations
Symmetric triangle (interior MF) 	$a_{MF}^{(i)} = \frac{1}{2}[(a^{(i)} + b^{(i)}) - \sqrt{2}(b^{(i)} - a^{(i)})]$ $b_{MF}^{(i)} = \frac{1}{2}[(a^{(i)} + b^{(i)}) + \sqrt{2}(b^{(i)} - a^{(i)})]$
Left-shoulder 	$a_{MF}^{(i)} = \frac{(a^{(i)} + b^{(i)})}{2} - \frac{(b^{(i)} - a^{(i)})}{\sqrt{6}}$ $b_{MF}^{(i)} = \frac{(a^{(i)} + b^{(i)})}{2} + \frac{\sqrt{6}(b^{(i)} - a^{(i)})}{3}$
Right-shoulder 	$a_{MF}^{(i)} = M - \frac{(a'^{(i)} + b'^{(i)})}{2} - \frac{\sqrt{6}(b'^{(i)} - a'^{(i)})}{3}$ $b_{MF}^{(i)} = M - \frac{(a'^{(i)} + b'^{(i)})}{2} + \frac{(b'^{(i)} - a'^{(i)})}{\sqrt{6}}$ $a'^{(i)} = M - b^{(i)}$ $b'^{(i)} = M - a^{(i)}$

C. Stochastic Convergence and its Application in this Paper

Four popular forms of stochastic convergence are [9]: convergence in distribution, convergence in probability, convergence with probability 1 and convergence in mean square. It is well known (e.g., [9]) that convergence in mean square implies convergence in probability (the converse is not true), and convergence in probability implies convergence in distribution (the converse is not true). In this paper, our focus is on convergence in mean square of the FOU word models. We do this by testing for convergence of the similarity of the FOUs, more on this in Section V.

The IA maps the assumed random interval end-points into an FOU. Even though this mapping is linear (see Table I), by the time the set of m^* T1 FSs is upper and lower bounded the resulting upper and lower MFs for the FOU are very non-linear functions of the surviving m^* data intervals. This means that it is not possible to compute the mathematical probability distributions for the parameters of the FOU (and their associated population means and variances) or for the FOU (it depends jointly on all of its parameters). Instead, the FOU is viewed herein as a generic non-linear function of the m^* data intervals, i.e.,

$$FOU(W) = h(\{[a^{(i)}, b^{(i)}], i = 1, 2, \ldots, m^*\}) \quad (3)$$

Another well-known fact from probability theory is [10]: AIf the parameters of a continuous function converge in probability to their true values, then that function also converges to its true value.. Unfortunately, we do not know what the “true” values are for the parameters of the FOU or for the function h; hence, this result is not used by us at this time. Instead, our approach is to study the mean-square convergence of the *entire FOU* by using similarity numbers, as explained next.

Given m^* surviving data intervals, we choose 25, 50, 75, ..., etc. subsets of these m^* intervals. For the purposes of the present discussion, let m_1^* and m_2^* denote successive subsets of intervals where the first value of m_1^* must be 25 (our smallest subset). By using random sampling (explained more in Section V), 100 sets of m_1^* and m_2^* intervals are created. Using the IA (actually an enhanced IA–EIA, as explained in Section IV) the following FOUs are computed: $FOU(W(i|m_1^*))(i = 1, 2, \ldots, 100)$ and $FOU(W(j|m_2^*))(j = 1, 2, \ldots, 100)$. Then the following 10,000 Jaccard similarity measures are computed:

$$\mathrm{s_J}[FOU(W(i|m_1^*)), FOU(W(j|m_2^*))] \; i, j = 1, 2, \ldots, 100 \quad (4)$$

This collection of 10,000 random numbers is denoted

$$\mathrm{s_J}(l(m_1^*, m_2^*)) \quad (5)$$

Because all of these 10,000 random numbers are used, as explained next, their ordering is unimportant, so the exact mapping from i, j to l is not important.

Our target number for the similarities in (4) is 1. If it can be shown that the random sequence of 10,000 numbers in (5) converges in mean-square to this target number, then it can be said, that: *using the IA, convergence in mean square occurs for $FOU(W)$.*

Both the sample mean and standard deviation of the 10,000 $\mathrm{s_J}(l(m_1^*, m_2^*))$ are computed. The difference between the sample mean of $\mathrm{s_J}(l(m_1^*, m_2^*))$ and the target number 1 is called the *bias*.

In order to prove convergence in mean-square, yet another well-known result from probability theory [9], is used, and it is stated next for our particular problem: *If the bias and the standard deviation for the 10,000 $\mathrm{s_J}(l(m_1^*, m_2^*))$ both approach zero as m_1^* and m_2^* both approach ∞, then the FOU obtained from the IA converges in mean square.* Because the population-mean and standard deviation for the similarity random variable are unknown, the sample mean and variance are used as their approximations. Convergence properties for these two well-known statistics are very well known (e.g., [9]).

By repeating the above calculations for $(m_1^* = 25, m_2^* = 50)$, $(m_1^* = 50, m_2^* = 75)$, $(m_1^* = 75, m_2^* = 100)$, etc., it is possible to compute a sequence of sample mean and standard deviations, and observe if the sample mean converges to 1 and the standard deviation converges to 0. Simulations in Section V demonstrate that both do occur.

A difficulty with this approach is it needs lots of data intervals, and unfortunately they were not available to us, i.e., Liu and Mendel [4] only had access to data from 28 subjects, a number that was felt to be too small to perform a convergence study. To remedy this situation, our first task was to obtain more data intervals. This is explained next.

III. Data Collection Procedure

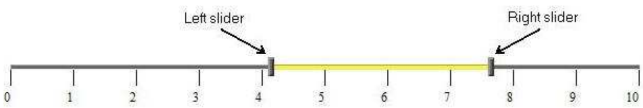


Fig. 5. Double ended slider used to collect intervals.

An online survey was conducted where participants were invited to give the interval which best describes a word on the interval scale of 0 to 10 using a pair of sliders as depicted in Figure 5. This interval data was gathered for a set of 32 words (taken from [4]). Each user was presented with the words in a random order and the majority of users did not give data for every word. Although the words were presented to the user in a randomised order, words which had less data entries were presented first. This meant that an equal number of samples could be captured about each word. The users were free to enter any value between 0 and 10 for each endpoint, with condition that the left endpoint must be less than or equal to the right endpoint. This methodology has some advantages and some disadvantages when compared to the previous work [4]. The online format of this survey meant that the level of participation could be much higher (175 participants as opposed to 28) than a paper based survey and the results could be more easily collated and without the risk of transcription errors. The lack of contact between the people conducting the survey and the participants meant there was less risk of influencing the data being collected, however it also meant there was no opportunity to explain the survey face-to-face or to answer questions about the survey which may have led to some participants not understanding the survey. The data collection method, a two tailed slider, meant that participants could enter data simply and intuitively, however, it may not have been clear to all participants that both sliders could be changed. We observed a small number of users who entered entire sets of words where the left endpoint was at 0 for every single word. We did not remove this data as we believed it would be captured by the preprocessing stage of the enhanced interval approach method presented next.

IV. Enhanced Interval Approach

When the IA was applied to the newly collected data we observed that the resulting FOUs were much broader than the ones in [4], something that had not been observed before (Fig. 6(a)). This caused us to examine the data and to re-examine the IA to try and understand why this had happened. One finding from our examination of the collected data was that many of the intervals were much broader than the ones that had been collected from the 28 JPL subjects. The JPL subjects had just completed a class about fuzzy sets and had also received instructions about the survey, whereas the people who took the web-survey had neither of those benefits. One could say that the JPL respondents were knowledgeable about fuzzy sets, whereas the web-based respondents were either less knowledgeable or not knowledgeable at all. A reviewer of the paper felt that the broader FOUs were due to the questions being asked out of a specific context. This is also correct; however, the difficulty with asking the questions in a specific context when one is modifying a general method is that there are an infinite number of contexts. Regardless, the EIA now guards against broad FOUs, something that would be true in or not in a specific context.

We also observed that some of the words seem to have confounded the web-based respondents, especially the term *quite a bit*. Consequently, for the purposes of this study we focused on five words that did not seem to confound those respondents, namely: very small, small, some, large and very large.

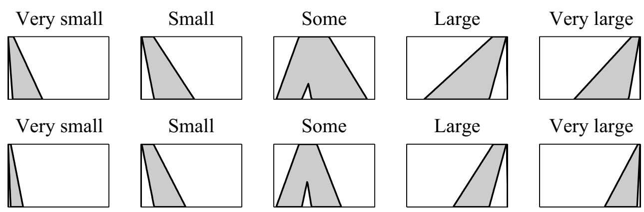


Fig. 6. FOUs for the five selected words (a) produced using IA (b) produced using EIA.

A close examination of the IA led us to make some modifications to it, all of which are in the spirit of the original IA but are felt to enhance it, which is why we are referring to our modified IA as the *Enhanced IA (EIA)*.

One of the key pre-processing steps in the IA is called (Stage–4) "Reasonable Interval Processing." Its purpose is commensurate with the adage "Words must mean similar things to different people," that was translated by Liu and Mendel [4] to mean that only overlapping intervals (obtained from the subjects) should be kept. They developed a *Reasonable interval test* that derives from probability theory. Although it did a very good job for the small 28-subject JPL data set, it did a poor job on our much larger data set, because of the very long surviving intervals in the latter. Reflecting further upon the preceding adage, we realized that translating this just into keeping overlapping intervals was an incomplete translation, because if the overlap is small but the lengths of the surviving intervals are long, then there will not be much similarity between the overlapping intervals.

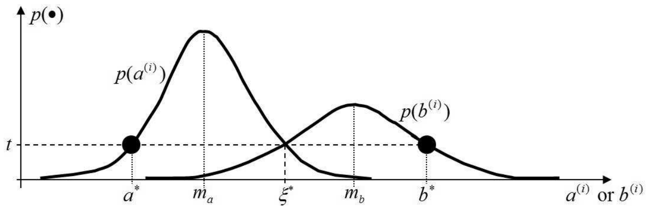


Fig. 7. Diagram for new *Reasonable interval tests.*

A close study of the derivation in [4] revealed that more results could be obtained from it, results that non only ensure overlapping intervals, but also ensure that those intervals are not overly long. Fig. 7 (adapted from Fig. 19a of their paper) depicts the situation. In their derivation, a threshold ξ^* is determined from probability theory, and they retain only those intervals for which $a^{(i)} < \xi^*$ and $b^{(i)} > \xi^*$. A close examination of the derivation of ξ^* reveals that their Eq. (A5), whose solution is ξ^*, can be interpreted geometrically as "ξ^* occurs at the intersection of the two normal distributions $p(a^{(i)})$ and $p(b^{(i)})$." Observe that this intersection occurs when $p(a^{(i)}) = t$ and $p(b^{(i)}) = t$. Observe, also, that this simple equation has three solutions, and not just the one at ξ^*. The two other solutions occur at

$$\begin{cases} a^{(i)} = a^* = m_a - (\xi^* - m_a) = 2m_a - \xi^* \approx 2m_l - \xi^* \\ b^{(i)} = b^* = m_b + (m_b - \xi^*) = 2m_b - \xi^* \approx 2m_r - \xi^* \end{cases} \tag{6}$$

Where m_l and m_r are the mean values of the left and right end-points of the surviving intervals. Consequently, our new *Reasonable interval test* is: Keep only the intervals for which

$$2m_l - \xi^* \le a^{(i)} < \xi^* < b^{(i)} \le 2m_r - \xi^* \tag{7}$$

This modification of the *Reasonable interval test* in [4] has added constraints on the lower limit of $a^{(i)}$ and the upper limit of $b^{(i)}$, both of which help to control the breadth of the surviving intervals, as desired.

The EIA has the following steps:

1) Data part (Fig. 2):
   a) Bad data processing: Only intervals with $0 \le a^{(i)} < b^{(i)} \le 10$ and $b^{(i)} - a^{(i)} < 10$ are accepted; others are rejected. This step reduces n interval endpoints to n' interval endpoints.
   b) Outlier processing: We perform Box and Whisker tests on $a^{(i)}$ and $b^{(i)}$ first, and then on $L^{(i)} = b^{(i)} - a^{(i)}$, i.e., we first compute $Q_a(.25)$, $Q_a(.75)$, IQR_a, $Q_b(.25)$, $Q_b(.75)$ and IQR_b based on the data from Step 1, and then keep only intervals satisfying

$$a^{(i)} \in [Q_a(.25) - 1.5IQR_a, Q_a(.75) + 1.5IQR_a] \tag{8}$$

$$b^{(i)} \in [Q_b(.25) - 1.5IQR_b, Q_b(.75) + 1.5IQR_b] \quad (9)$$

This step reduces n' interval endpoints to n'' interval endpoints. We then compute $Q_L(.25)$, $Q_L(.75)$, and IQR_L based on the remaining n'' intervals, and keep only intervals satisfying

$$L^{(i)} \in [Q_L(.25) - 1.5IQR_L, Q_L(.75) + 1.5IQR_L] \quad (10)$$

Where $Q_a(.25)$ and $Q_a(.75)$ are the first and third quartile range of the lower limit $a^{(i)}$, similarly $Q_b(.25)$ and $Q_b(.75)$ are the first and third quartile range of $b^{(i)}$ and IQR_a and IQR_b are the interquartile ranges of $a^{(i)}$ and $b^{(i)}$. This step reduces n'' interval endpoints to m' interval endpoints.

Note that in the original IA these three tests are performed simultaneously. Here we separate the test on the length of the intervals from the tests on the endpoints because outliers for $a^{(i)}$ and $b^{(i)}$ can make IQR_L so large that $Q_L(.25) - 1.5IQR_L$ can be negative, in which case the Box and Whisker test on $L^{(i)}$ is not effective for removing short-length intervals that contribute to a small LMF.

c) Tolerance limit processing: We perform tolerance limit processing on $a^{(i)}$ and $b^{(i)}$ first, and then on $L^{(i)} = b^{(i)} - a^{(i)}$. For the former, we keep only intervals satisfying

$$a^{(i)} \in [m_a - k\sigma_a, m_a + k\sigma_a] \quad (11)$$

$$b^{(i)} \in [m_b - k\sigma_b, m_b + k\sigma_b] \quad (12)$$

where k is determined such that one can assert with 95% confidence that the given limits contain at least 95% of the subject data intervals. This step reduces m' interval endpoints to m^* interval endpoints. We then compute m_L and σ_L based on the remaining data and keep only intervals satisfying

$$L^{(i)} \in [m_L - k'\sigma_L, m_L + k'\sigma_L] \quad (13)$$

where

$$k' = \min(k_1, k_2, k_3) \quad (14)$$

in which k_1 is determined such that one can assert with 95% confidence that $[m_L - k'\sigma_L, m_L + k'\sigma_L]$ contains at least 95% of $L^{(i)}$, and

$$k_2 = m_L/\sigma_L \quad (15)$$

$$k_3 = (10 - m_L)/\sigma_L \quad (16)$$

(15) ensures that $m_L - k'\sigma_L \geq 0$, and (16) ensures that $m_L + k'\sigma_L \leq 10$, so that intervals with too small or too large $L^{(i)}$ can be removed. This step reduces m^* interval endpoints to m'' interval endpoints.

d) Reasonable-interval processing: We keep only intervals such that (7) is satisfied where ξ^* is computed by (17)

$$\xi^* = \frac{(m_b\sigma_a^2 - m_a\sigma_b^2) \pm \sigma_a\sigma_b[(m_a - m_b)^2 + 2(\sigma_a^2 - \sigma_b^2)\log(\sigma_a/\sigma_b)]^{1/2}}{\sigma_a^2 - \sigma_b^2} \quad (17)$$

This step reduces m'' interval endpoints to m endpoints.

In summary, data preprocessing starts with all n data intervals and ends with m data intervals, i.e.

$$n \overset{bad\ data}{\rightarrow} n' \overset{outliers}{\rightarrow} m' \overset{tolerance\ limits}{\rightarrow} m'' \overset{reasonable\ interval}{\rightarrow} m \quad (18)$$

2) Fuzzy set part (Fig. 3):

a) FOU classification: The procedure is the same as that in the original IA. This step reduces m interval endpoints to m^* interval endpoints.

b) Compute the embedded T1 FSs: The formulas are the same as those in the original IA.

c) Delete inadmissible T1 FSs: The same as the original IA.

d) Compute the UMF and LMF: The procedures for shoulder FOUs and for the UMF of interior FOUs are correct; however, the procedure for the LMF of interior FOUs needs improvement. Currently it only considers the case that the LMF of an interior FOU is completely determined by the two embedded T1 FSs that also determine the UMF, as shown in Fig. 8(a); however, this is not always true in practice. Three counter-examples are shown in Figs. 8(b), 8(c) and 8(d). The key point is to determine the location and height of the apex, i.e., p and μ_p in Figs. 8(a) – (d). Because in practice we usually have fewer than 100 such embedded T1 FSs and the EIA is used off-line, we use exhaustive search to find this apex, i.e., find all possible intersections of left legs with right legs and then choose the apex as the intersection with the minimum height.

Results from the EIA are depicted in Fig. 6(b). Comparing them with the results in Fig. 6(a), it is clear that we have obtained more reasonable looking FOUs from a group of subjects, whose knowledge about fuzzy sets was unknown to us, and who took the survey on the Internet. By reasonable looking FOUs we mean FOUs capture uncertainty about the words but retain the basic shape of a trapezoid or shoulder FOU.

The following Section describes an experiment to test whether the words constructed using this method show stochastic convergence in the mean square.

V. Stochastic Convergence Experiment

As mentioned above the five words have been chosen for this initial experiment are, in no particular order: *very*

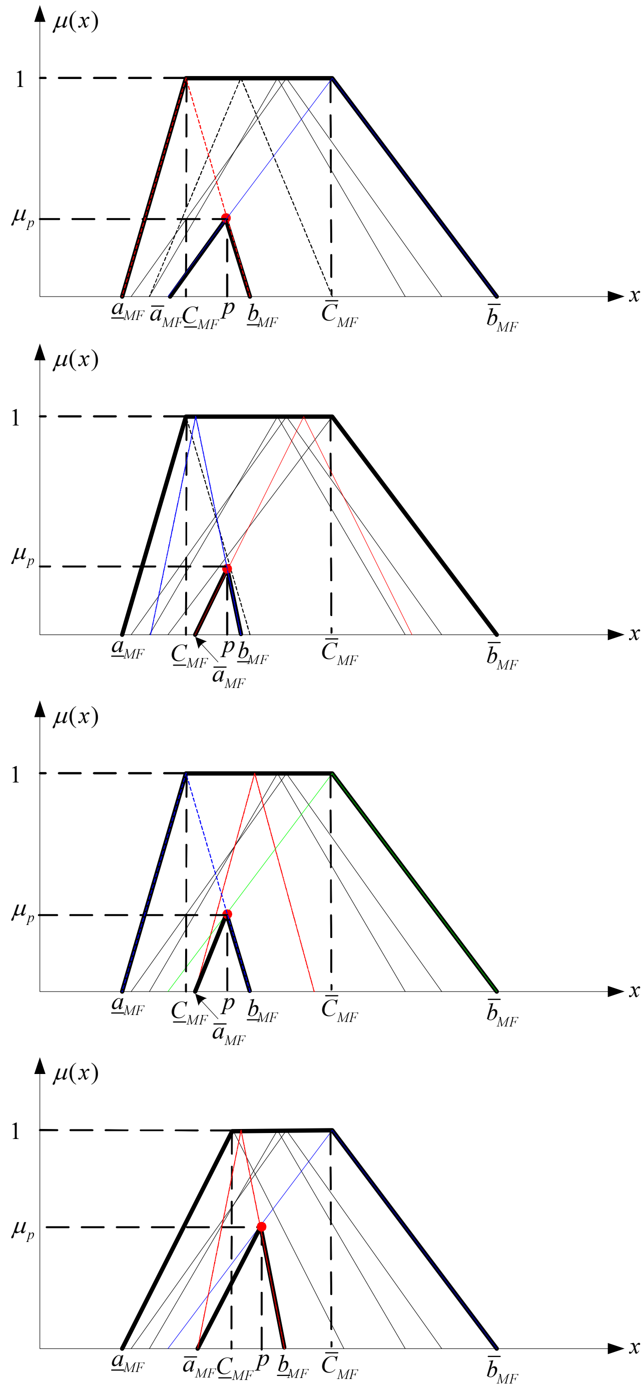


Fig. 8. The four cases for the LMF of an interior FOU.

TABLE II
NUMBER OF INTERVALS REMAINING AFTER PREPROCESSING.

Word	m*
Very Small	75
Small	100
Some	44
Large	94
Very Large	89

small, *very large*, *some*, *large* and *small*. This first experiment investigates whether the consensus FOU for a word converges when data are collected from enough people i.e., will the constructed word model be identical (or in reality very similar) to a word constructed from data from a different group of people? Table II shows the number of interval data points (m^*) which survived the preprocessing steps of the EIA algorithm. The number of surviving data points is dependent on the raw data. In order to conduct a controlled experiment in terms of the number of data points we first ran the EIA preprocessing steps on all the collected data. This gave us a set of intervals which we knew to be good in the sense that they would contribute to the FOU constructed using all available data. We used this preprocessed data as a pool of intervals from which word FOUs be constructed. We sampled a number of intervals from this pool with replacement; this means that one interval may be selected more than once and with these selected intervals we constructed an a FOU using only the fuzzy set part of the EIA. This procedure, which relies on bootstrapping is followed as it ensures that an exact number of intervals are used to constructed an FOU. Bootstrapping [3] is a method to construct a large dataset from a small dataset, or to construct more datasets from a single dataset, by random sampling with replacement. It can be used for statistical inference, e.g., to estimate the variance of an estimator [3], [2]. Usually the original dataset is the empirical distribution of the observed data, and the underlying assumption is that these observations assume an independent and identical distribution. Bootstrapping has been proven to lead to more accurate estimates of sample mean and variance [3]. It is recommended [1] when the theoretical distribution of a statistic is complicated or unknown, and/or when the sample size is insufficient for straightforward statistical inference. Both conditions are satisfied in our convergence experiment; so, bootstrapping was employed.

The procedure for carrying out this experiment was as follows for each word:

1) Perform EIA.
2) Construct FOUs using sample sizes of 25, 50, 75, 100, 125, 150, 175 and 200. Repeat this 100 times (bootstrapping was used).
3) Compute the similarity, using the Jaccard similarity measure in (2) for each pair of consecutive sample sizes (25 and 50, 50 and 75, etc.) for all 100 realizations (see (4)).
4) Calculate the mean and standard deviation of the 10,000 similarity measures for each consecutive similarity pair (see (4) and (5)).

Table III and Table IV respectively give the mean and standard deviation of the similarity values for each consecutive pair for all the words. Figures 9 and 10 show these results graphically.

Observe from these tables and figures that the mean and standard deviation of the similarity functions appear to be converging simultaneously to 1 and 0, which, as explained

TABLE III
MEAN SIMILARITY VALUES OF FIVE SELECTED WORDS.

Word	Sample Size Pairings						
	25-50	50-75	75-100	100-125	125-150	150-175	175-200
Very Small	0.9302	0.9522	0.9663	0.9753	0.9830	0.9866	0.9883
Small	0.9348	0.9608	0.9688	0.9808	0.9862	0.9902	0.9924
Some	0.8845	0.9313	0.9600	0.9793	0.9893	0.9947	0.9966
Large	0.9512	0.9788	0.9893	0.9930	0.9949	0.9965	0.9973
Very Large	0.9144	0.9520	0.9655	0.9798	0.9851	0.9924	0.9957

TABLE IV
STANDARD DEVIATION OF SIMILARITY VALUES OF FIVE SELECTED WORDS.

Word	Sample Size Pairings						
	25-50	50-75	75-100	100-125	125-150	150-175	175-200
Very Small	0.0373	0.0285	0.0252	0.0223	0.0197	0.0183	0.0174
Small	0.0397	0.0298	0.0232	0.0151	0.0120	0.0099	0.0094
Some	0.0593	0.0510	0.0455	0.0313	0.0206	0.0134	0.0110
Large	0.0384	0.0182	0.0103	0.0066	0.0062	0.0044	0.0040
Very Large	0.0540	0.0385	0.0322	0.0261	0.0231	0.0152	0.0098

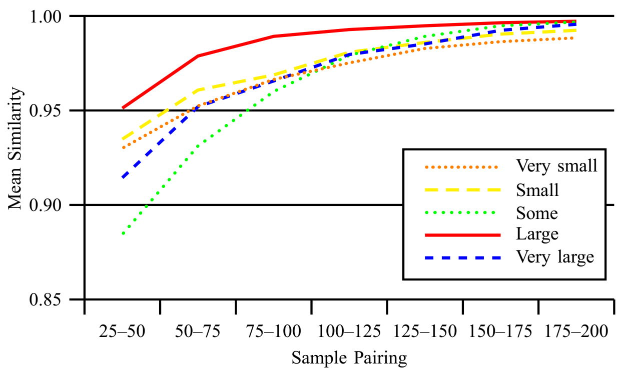


Fig. 9. Mean similarity values of five selected words.

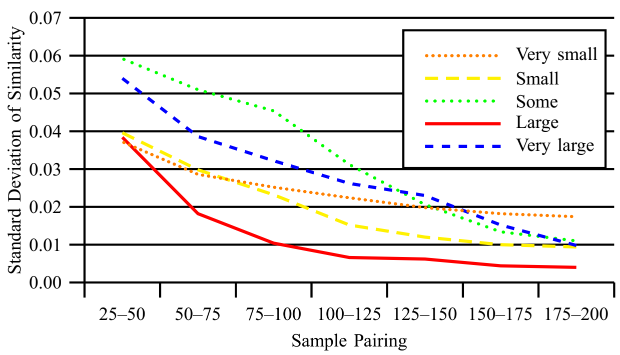


Fig. 10. Standard deviation of similarity values of five selected words.

in Section II.C is indicative of convergence of the similarity in mean square, which in turn is indicative of convergence of each word's FOU in mean square.

VI. CONCLUSIONS

The IA [4] is a method for synthesizing an IT2 FS model for a word from data that are collected from a group of subjects. A key assumption made by the IA is: each person's data interval is random and uniformly distributed. This means, of course, that the IT2 FS model for the word is random. Consequently, one can question whether or not the IT2 FS model for the word converges in a stochastic sense. This paper has focused on this question.

As a part of our study, we have had to modify some steps of the IA; however, each modification was built upon the original steps, the resulting being an *Enhanced IA (EIA)*.

We have shown by means of some simulations, that the IT2 FS word models that are obtained from the EIA are converging in a mean-square sense. This provides substantial credence for using the EIA to obtain T2 FS word models. A reviewer of the paper questioned whether our MS convergence was due to the entire EIA or to mainly the preprocessing that is a part of the EIA. We advocate the preprocessing, because the kind of pre-processing we are doing is done all of the time by statisticians on other data sets, and is considered to be good practice by statisticians. Not to do this would be poor practice.

Additional convergence results for the parameters of the FOUS as well as for the entire FOU will appear in the journal version of this paper. Software that implements the EIA can be downloaded at: http://sipi.usc.edu/~ mendel/software.

ACKNOWLEDGMENT

The authors would like to thank Robert John and Hussam Hamrawi for their help in undertaking this research. We'd also like to thank the reviewers for their very insightful comments. Due to page limitations we have not been able to reply to all of them, but will do so in the journal version of the paper which we are currently preparing.

REFERENCES

[1] H. J. Adér and G. J. Mellenbergh and D. J. Hand "Advising on research methods: A consultant's companion", Huizen, The Netherlands: Johannes van Kessel Publishing, 2008.
[2] L. Breiman, "Bagging predictors," Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.
[3] B. Efron and R. Tibshirani, "An Introduction to the Bootstrap", Boca Raton, FL: Chapman & Hall/CRC, 1993.
[4] F. Liu and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," IEEE Trans. Fuzzy Systems, vol. 16, no. 6, pp. 1503-1521, 2008.
[5] J. M. Mendel, "Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions," Upper-Saddle River, NJ: Prentice-Hall, 2001.
[6] J. M. Mendel, "Type-2 fuzzy sets and systems: an overview," IEEE Computational Intelligence Magazine, vol. 2, pp. 20-29, February 2007.
[7] J. M. Mendel, "Computing with words and its relationship with fuzzistics," Inf. Sci., vol. 177, pp. 988-1006, 2007.
[8] J. M. Mendel and R. I. John, "Type-2 fuzzy sets made simple," IEEE Trans. Fuzzy Systems, vol. 10, pp. 117-127, April 2002.
[9] V. K. Rohatgi, "An Introduction to Probability Theory and Mathematical Statistics", Wiley, New York, 1976.
[10] H. G. Tucker, "A Graduate Course in Probability," Academic Press, New York, 1967.
[11] D. Wu and J. M. Mendel, "A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets," Inf. Sci., vol. 179, no. 8, pp. 1169-1192, 2009.