統計的方法によるメンバーシップ関数の推定[†]

渡辺 則生*

メンバーシップ関数の推定という問題を、ファジィ確率変数の期待値を推定することと定式化し、推定法に関する客観的な議論が可能であることを示した。そして、メンバーシップ関数に関する情報が反復して得られる状況のもとでは、ファジィネスを推定することが可能であることを明らかにし、具体的な推定法を提案した。さらに、実際例に応用し、提案した方法の有効性を示した。また、従来の方法のひとつである、反復データの頻度に基づく方法についての理論的妥当性は得られないことを示した。

キーワード：メンバーシップ関数、ファジィ確率変数、推定、期待値

1. はじめに

ファジィ集合を用いたモデルを応用する場合、メンバーシップ関数を同定する必要がある。その方法としては、大きくふたつにわければ、主観的に適当な形に設定する方法と、何らかのデータをもとにして客観的に推定する方法がある。たとえば、ある言葉が世代によってどのような意味で使われるかを、メンバーシップ関数の差異によって議論しようとする場合には、主観的に設定する方法は適当ではなく、対象とする集団から得られたデータをもとにしてメンバーシップ関数を客観的に推定する必要がある。ここでは、集団における言葉あるいは概念の意味といった種類のあいまいさの問題を考え、意味を表すメンバーシップ関数を、どのようなデータを用い、どのような方法によって推定すればよいかという点について考察する。本論文のひとつの目的は、客観的な推定がどのような条件のもとで可能であることを明らかにすることである。

一般に、メンバーシップ関数を推定する方法には

ふたつの代表的な方法がある。すなわち、(I)集団における頻度データをもとにして推定する方法と(II)評定尺度法などにより直接推定する方法である(山下[5]、藪本他[4]、Leung[11]、Norwich and Turksen[12]、Thole 他[14]参照)。頻度データによる方法では、ある要素がある集合に属しているかどうかの問いに対して、“yes”と答えた人数の集団における割合をグレードとみなすという考え方が基本となる。評定尺度法は、1から5までの数値や、0と100の間の数値といった数値尺度によって評定者がある事柄について判断を行う方法である。このような推定法に関してはいろいろと議論があり、特に(I)については、確率との関係という点で明らかでない部分がある。本論文では、ファジィ確率変数を用いたモデルを導入することによってメンバーシップ関数の推定問題を定式化し、一般には、(I)と(II)では本質的に異なることを明らかにし、メンバーシップ関数の推定法としては原則として(II)によるべきであることを示す。さらに、(II)に対応した推定法を提案する。また、実際例に応用した結果を紹介する。本論文の結果は、(I)の方法に対する理論的根拠が一般には得られないことを示している。また同時に、(I)の方法が経験的に妥当な結果を与えるかどうかの実証分析においては、本論文のモデルが検証のために使用できることもわかる。

[†] A Statistical Method for Estimation of a Membership Function

Norio WATANABE

* 中央大学 理工学部 管理工学科

Department of Industrial and Systems Engineering, Chuo University

言葉あるいは概念の意味のあいまいさには主観的要素が強く含まれ、集団のなかの個人ごとのばらつきが大きいと考えられる。このような状況のもとで、「言葉の意味」が何を意味するのか、言い替えると、推定しようとしているメンバシップ関数とは何であるのかを明確にする必要がある。本論文では、ある意味で平均的な言葉の意味、あるいはそのメンバシップ関数が推定したい目的のものであると考える。これは、主体として集団における平均的人物を想定することによってよいであろう。集団のなかでばらつくという現象は、確率で表現するのが一般的である。また、ある個人が言葉に与える意味というのはファジィ集合で表すことができよう。このことは、ファジィネスとランダムネスという2種類のあいまいさを考慮にいれるのが自然であることを示している。ファジィネスとランダムネスを同時に扱うことのできる概念として、Kwakernaak [9, 10], Puri and Ralescu [13]などによりファジィ確率変数が提案されている(本論文では Puri and Ralescu による定義に従う。付録参照)。ここでは、集団における個人の言葉の意味がファジィ確率変数によって表現されるものと仮定する。従って、推定したいメンバシップ関数とはファジィ確率変数の平均的な値すなわち期待値であるファジィ集合のメンバシップ関数のことである、と定式化できる。本論文では、このような立場からメンバシップ関数の推定の問題について考察する。ファジィネス(意味のあいまいさ)とランダムネス(個人差による意味の変化)を同時に考慮に入れた理論的な議論はこれまで十分にはなされていない。本論文のモデルは、このような議論を可能にする一般的なモデルである。

意味が集団内で固定的なときはファジィ確率変数を導入する必要はない。この場合には評定尺度法などによりメンバシップ関数を直接推定することが可能であろう。本論文では固定的でない場合を扱う。

なお、ここでは、言葉の意味を表すファジィ集合の台集合が線形空間の部分集合である場合に限

定して議論する。台集合が線形空間ではなく離散的な抽象空間である場合は、たとえば Thole 他 [14]によるメンバシップ関数の推定法が適用できる。従来の他の方法との関連については2.4節で論ずる。

2. モデルと推定法

2.1 モデル

n 人の集団から、ファジィ集合で表される n 個のデータ $X_1, X_2, \dots, X_n$ が観測されたとする。ファジィ集合 X_i の台集合は、線形空間の部分集合であると仮定する。ここでは簡単のために実数空間上の区間であるとする。本論文では、前節で述べた理由から、 X_i について次のモデルを採用する。モデル. $\{X_1, X_2, \dots, X_n\}$ はあるファジィ確率変数 X からの無作為標本である。

ここでは確率変数 X が直接観測されるという条件のもとで議論を進める(X のメンバシップ関数を規定するパラメータが観測される場合も含まれる)。 X が直接観測されるということは、 X の取り得る値であるファジィ集合のメンバシップ関数がどのような形をしているかがわかることを意味する。これは、あらかじめ予備調査などを行い、メンバシップ関数を同定しておくことによって可能になる。この点については4節において触れる。

このとき、メンバシップ関数の推定という目的は次のように表現できる。

目的. $X_1, X_2, \dots, X_n$ によって EX を推定すること。ここで、 EX はファジィ確率変数の期待値を表す。ファジィ確率変数およびその期待値については Kwakernaak [9, 10] や Puri and Ralescu [13] などにより議論されている。本論文では、Puri and Ralescu による定義に従う(付録参照)。また、 X の取り得る値であるファジィ集合については次の条件をおく。

仮定 1. X の任意の実現値を A で表したとき、ファジィ集合 A は正規であり、 A の α レベル集合は閉区間で与えられる。ただし $0 < \alpha \leq 1$ 。

また、 X の期待値は存在するものと仮定する。

このような定式化のもとで、ファジィ集合 EX あるいはそのメンバシップ関数の推定という統計的問題を考える。ここでは、典型的なふたつの場合を考え、それぞれについて具体的な推定法を与える。第1は離散型のファジィ確率変数の場合で、簡単のために取り得る値のファジィ集合が有限個とする。第2は連続型のファジィ確率変数の場合で、メンバシップ関数の関数型が既知であり、関数のパラメータが連続的であるときを考える。第2の場合、パラメータ自身を実数値確率変数とみなすことができ、またその値が直接観測できるものとする。それぞれの場合について EX の推定量を構成する。なお、Kwakernaak[10]や Puri and Ralescu[13]は EX の計算法も与えている(おもに離散型の場合)。しかし、推定といった統計的問題に関する一般的議論はまだ十分にはなされていない。

2.2 推定法1—離散型

取り得るファジィ集合を $A_1, A_2, \dots, A_K$ とおく。ここで、 A_i の α レベル集合を $[a_i, b_i]$ 、 $X=A_i$ となる確率を p_i とおくと、仮定1のもとで EX の α レベル集合は次式で与えられる。

$$(EX)_\alpha = \left[\sum_{i=1}^K p_i a_i, \sum_{i=1}^K p_i b_i \right] \quad (1)$$

従って、 p_i の推定値を(1)の右辺に代入すれば、 EX の推定量が得られることになる。今、 n 個のデータの中で、 A_i であるものの個数を n で割った値を $\hat{p}_i$ とおけば、 $\hat{p}_i$ は p_i の一致推定量となる。従って、 α レベル集合

$$(\hat{EX})_\alpha = \left[\sum_{i=1}^K \hat{p}_i a_i, \sum_{i=1}^K \hat{p}_i b_i \right] \quad (2)$$

によって定義されるファジィ集合を $\hat{EX}$ とすれば、 $\hat{EX}$ の任意の α レベル集合が EX の α レベル集合へ確率収束するという意味で推定量 $\hat{EX}$ は EX に収束することがわかる。また、 $\hat{p}_i$ は p_i の不偏推定量であり、 $E\left(\sum_{i=1}^K \hat{p}_i a_i\right) = \sum_{i=1}^K p_i a_i$ 、 $E\left(\sum_{i=1}^K \hat{p}_i b_i\right) = \sum_{i=1}^K p_i b_i$ となり、このような意味で $\hat{EX}$ は EX の不偏性・一致性を持つ推定量になっている。

これらのことは(2)式によってメンバシップ関数を推定する方法の妥当性を示している。ここで、(2)式の右辺は、ファジィ数の算術平均「平均ファジィ数」を与える α レベル集合と一致している(Kaufmann and Gupta[8]§2.3 参照。なお、[8]の THEOREM 2.3 で与えられている「期待ファジィ数」は、(1)式に相当する)。ファジィ確率変数を用いることによって、算術平均あるいは標本平均と期待値の対応関係が明確になり、標本平均を用いることの根拠が得られる。

数値例を以下に示す。

例1. $K=3$ として、図1上の A_1, A_2, A_3 が X の取り得る値であるとする。データから $\hat{p}_1=0.1$ 、 $\hat{p}_2=0.5$ 、 $\hat{p}_3=0.4$ が得られたとすると、 $\hat{EX}$ は図1下のメンバシップ関数で与えられる。

図1 EX の推定例

ところで、 $\hat{p}_i$ の平均、分散はそれぞれ $\hat{p}_i$ 、 $\hat{p}_i(1-\hat{p}_i)/n$ であり、さらに $\hat{p}_i$ は漸近的に有効な推定量でありかつ漸近分布は正規分布である。このことから、 EX の変動を考慮にいたした EX の区間推定が、以下のようにして可能となる。

$(\hat{p}_1 \ \hat{p}_2 \ \dots \ \hat{p}_K)'$ の平均値ベクトル、分散共分散行列はそれぞれ

$$\begin{bmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \vdots \\ \hat{p}_K \end{bmatrix}, \frac{1}{n} \begin{bmatrix} \hat{p}_1(1-\hat{p}_1) & -\hat{p}_1\hat{p}_2 & \cdots & -\hat{p}_1\hat{p}_K \\ -\hat{p}_2\hat{p}_1 & \hat{p}_2(1-\hat{p}_2) & \cdots & -\hat{p}_2\hat{p}_K \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{p}_K\hat{p}_1 & -\hat{p}_K\hat{p}_2 & \cdots & \hat{p}_K(1-\hat{p}_K) \end{bmatrix}$$

となる(たとえば、竹内・藤野[1]p.26)。従って

$\sum_{i=1}^K \hat{p}_i a_i, \sum_{i=1}^K \hat{p}_i b_i$ の分散はそれぞれ

$$\sigma_a^2 \equiv \left\{ \sum_{i=1}^K p_i a_i^2 - \left(\sum_{i=1}^K p_i a_i \right)^2 \right\} / n$$

$$\sigma_b^2 \equiv \left\{ \sum_{i=1}^K p_i b_i^2 - \left(\sum_{i=1}^K p_i b_i \right)^2 \right\} / n$$

であることが示される。このことから、 n が十分大きい場合、正規近似によって α レベル集合の端点の信頼区間が得られるので、 EX の区間推定が可能となる。たとえば α レベル集合の左端の近似的な信頼区間は、(2)式の左端の値に幅を持たせた次の区間で与えられる。

$$\left[\sum_{i=1}^K \hat{p}_i a_i - z \hat{\sigma}_a, \sum_{i=1}^K \hat{p}_i a_i + z \hat{\sigma}_a \right]$$

ただし z は標準正規分布のパーセント点であり、 $\hat{\sigma}_a$ は σ_a に含まれる p_i に $\hat{p}_i$ を代入して得られる値である。 z を上側 $\beta/2$ 点とすれば、この区間は近似的な $(1-\beta)$ 信頼区間となる。

実データへの適用例は3.1節で示す。

2.3 推定法2—連続型

ファジィ集合 X のメンバシップ関数 $\mu_X(x)$ が、 $\mu_X(x) = f(x; \theta)$ と表されるものとする。ここで f は $[0, 1]$ 区間の値を取るある関数で、 θ はパラメータ(ベクトル)である。標本を ω と表せば、 $X(\omega)$ という実現値は $\theta(\omega)$ という実現値によって次式によって定まるとする。

$$\mu_{X(\omega)}(x) = f(x; \theta(\omega)) \quad (3)$$

このことは θ が確率変数であることを意味する。今、確率変数であるパラメータ θ が直接観測されるものとする。また、 θ の2次モーメントの存在を仮定する。ここで X の α レベル集合の下限、上限を η, ν とおく。 η, ν は確率変数であり次のように表される。

$$\eta(\alpha) = \inf_{\{f(x; \theta) \geq \alpha\}} x \quad (4a)$$

$$\nu(\alpha) = \sup_{\{f(x; \theta) \geq \alpha\}} x \quad (4b)$$

仮定1のもとでは、 EX の α レベル集合は次式で与えられる。

$$(EX)_\alpha = [E(\eta(\alpha)), E(\nu(\alpha))] \quad (5)$$

(5)の右辺における $E(\cdot)$ は、通常の実数値確率変数の期待値を意味する。従って、 $E(\eta), E(\nu)$ の推定値を(5)に代入することによって EX が推定される。 n 個のデータが与えられた場合、 $X_1, X_2, \dots, X_n$ に対応して $\theta_1, \theta_2, \dots, \theta_n$ が観測され、さらに(4)式によって $\eta_1, \eta_2, \dots, \eta_n$ および $\nu_1, \nu_2, \dots, \nu_n$ が得られる。そして η_i, ν_i の標本平均 $\bar{\eta}, \bar{\nu}$ は、一般に $E(\eta), E(\nu)$ の一致性・不偏性を有する推定量となっている。従って、 α レベル集合

$$(\widehat{EX})_\alpha = [\bar{\eta}(\alpha), \bar{\nu}(\alpha)] \quad (6)$$

によって定義されるファジィ集合 $\widehat{EX}$ は、 EX の妥当な推定量になっていることがわかる。

また、 $E(\eta), E(\nu)$ の信頼区間が α レベル集合の端点の信頼区間となることから、 EX の区間推定を考えることも可能である。 $E(\eta), E(\nu)$ の信頼区間を求めるためには、次のような $(1-\beta)$ 信頼区間を、(4)式を評価して求めればよい。

$$P(\eta_1 \leq \eta \leq \eta_2) = 1 - \beta$$

$$P(\nu_1 \leq \nu \leq \nu_2) = 1 - \beta$$

$\eta_i, \nu_i (i=1, 2)$ は $\theta_1, \dots, \theta_n$ から計算される統計量であり、 $P(\cdot)$ は $\theta_1, \dots, \theta_n$ についての確率を意味する。具体的な η_i, ν_i の導出は $f(x; \theta)$ に依存する。3.2節において具体例を示す。なお、 α レベル集合の端点 η, ν が θ の非線形関数で表される場合、正確な信頼区間を求めることは簡単ではないが、 n が大きいときには η, ν の標本分散を用いて正規近似することが可能となる。

Remark 1. θ が観測されることから、 θ の標本平均を(4), (5)に代入して EX を推定する方法を考えることもできる。メンバシップ関数が線形の場合、この方法で得られるファジィ集合と(6)式の方法で得られるものとは一致することがわかる。しかし非線形の場合は一致するとは限らない。一致しない場合、 θ の標本平均を用いた場合には(4)

においてバイアスが生じる可能性が大きく、不偏性が失われることになる。従って、(6)式を用いることが望ましいと言えよう。

2.4 従来の方法との関連

2.2 および 2.3 節で示したメンバシップ関数の推定法は、1 節で述べた(II)の方法に含まれるといつてよいであろう。

一方、1 節の(I)によるアプローチは、ある要素がある集合に属しているかどうかの問に対して、属していると判断した人の集団における割合をグレードに対応させるという考え方による(Leung [11]pp. 327-328 参照)。そのような方法のひとつとして、ある種の密度関数を推定して密度関数からメンバシップ関数を構成するという方法がある(たとえば 藪本他[4]など)。この方法は、ふたつのファジィ集合の「境界」のデータをサンプリングして確率密度を推定し、分布関数をメンバシップ関数とみなすという考え方に基づいている(3.3 節参照)。「境界」の値は、われわれのモデルのもとでは、2.3 節の(3)式の θ と考えることができる。従って、このような方法で推定しているのは確率変数 θ の分布関数であって、本来の目的であるあいまいさ、すなわちメンバシップ関数 f とは本質的に異なるものを推定していることになる。

また、頻度データによる方法を拡張したものとして、区間データを利用した Wang の方法がある(Leung [11]pp. 328-333 参照)。Wang の方法も、本質的には頻度をグレードととらえるものであり、頻度データによる方法の一種とみなすことができる。

基本的には、頻度をもとに推定して得られるものはあくまでもある種の密度関数に対応するものであり、本来の目的であるファジィネスとは本質的に異なるものと考えべきである。密度関数から得られるものをファジィネスとみなすことも許されようが、その場合にはあくまでも主観的な判断に基づくものであり、客観的な方法と考えるべきではないであろう。なお、Dubois and Prade[7]は、確率測度から可能性測度が得られることを理

論的に示している。このことは分布をもとにしてメンバシップ関数が得られることを保証する。しかし、得られたメンバシップ関数が求めたいあいまいさを表していることを保証しているわけではないことに注意する必要がある。

われわれの方法では X が観測可能という仮定が含まれており、この仮定は必ずしも一般的とは言えない。しかし、ファジィ確率変数を用いたモデルそのものは十分一般的であると考えられる。このことから、(I)の方法はメンバシップ関数の客観的な推定法としては不適当であると結論づけられる。目的によっては、たとえば θ の分布をメンバシップ関数とみなしてよいこともあろうが、このような場合は客観的な妥当性は失われることになる。(I)の方法が適当かどうかは、実証分析によって経験的に裏付けが与えられるかどうかにより依存する。実証分析は本論文で提案した枠組で行うことができよう(3.3 節参照)。

2 節で提案した方法は、ファジィ集合の台集合が線形空間の部分集合である場合に適用できる方法である。線形空間であることは、期待値 EX が意味を持つために必要な条件である。台集合は必ずしも連続である必要はなく、離散的であっても、線形空間の部分集合であれば意味を持つこともある(渡辺[2]参照)。一方で、台集合が抽象的な要素からなる離散的な集合であることもある。このような場合には、Thole 他[14]によるメンバシップ関数の推定法が適用可能である。Thole 他[14]は明示していないが、彼らの方法は、台集合の各要素に対するグレードが、集団における個人ごとにランダムに変動するというモデルのもとで、グレードの平均値を推定するという定式化に基づいていると解釈できる。Thole 他[14]の方法は、台集合が線形空間である場合に応用することも可能である。しかし、線形構造が考慮されないため、各個人ごとのファジィ集合が凸集合であるのに、推定されたファジィ集合は凸にならないといった不自然な現象が起こり得ることがわかる(渡辺[2]参照)。これに対して本論文の方法は、線形構造に基づいた方法であり、線形空間の場合についてはよ

り自然な結果を与えるものと考えられる。なお、観測形式はそれぞれの方法で異なっており、結果を直接比較することはできない。Thole 他[14]の方法では、台集合の要素単位でグレードを観測する必要がある。

また、Norwich and Turksen [12]は、個人のメンバシップ関数を推定するための方法として、Thole 他[14]と同様の考え方を用いている。従って、上述の問題点を指摘することができる。一方で、Norwich and Turksen [12]の方法を、たとえば2.2節における $A_1, \dots, A_K$ を同定するための予備調査の手法として利用することが可能である(4節参照)。

3. 応用例

ここでは2節におけるふたつの場合に対応してふたつの実例を示す。いずれの例も、身長や年齢に関する言葉の意味についてのある集団におけるアンケートデータに対する適用結果である。

図2 A のメンバシップ関数

3.1 離散型の例

$n=93$ 人の大学生の集団に対し次のような質問を行った。

15 歳から 35 歳の男性において「年齢が上」という言葉があてはまる度合を 0-1 の数字で表すとする。このとき、図 2 のグラフのなかで一番適当と思われるものの番号はいずれか。ただし、「年齢が上」は、「比較的高い年齢層」という意味を持ち、「比較的低い年齢層」という意味の「年齢が下」と対をなす概念を意味する言葉であるとする。

図 2 のグラフは、 X の取得得る値であるファジィ集合 $A_1, A_2, \dots, A_{10}$ のメンバシップ関数を表している。集計した結果より、表 1 のような p の推定値が得られた。2.2 節の方法を適用して得られた EX の推定値は図 3 の実線で表されるメンバシップ関数で与えられる。得られた結果は、「年齢が上」という言葉の代表的メンバシップ関数としては妥当なものと思われる。また、図 3 の破線で囲まれた領域は、2.2 節で提案した方法による、 α レベル集合の左端の 95%信頼区間よりなる領域を表している。このことにより、得られた推定値の誤差の評価が可能となる。なお、 α レベル集合の右端は、メンバシップ関数が増加型であることから台集合の上限の値(ここでは 35 であるが、一般には $+\infty$)で常に一定となる。

問題としては、図 2 のメンバシップ関数の設定が十分であるかという点があげられる。このアン

表1 確率の推定値

i	1	2	3	4	5	6	7	8	9	10
$\hat{p}_i$	.09	0	.17	.03	.18	.31	.03	.10	.04	.04

図3 EX の推定値のメンバシップ関数

ケートにおいては、グラフの設定に関する満足度の調査も同時に行っている。そして不満のときには自分でグラフを書くように指示を与えている。その結果、若干名は、図2のグラフの番号を選択した上で自らグラフを描いたが、図2との大きな差は見受けられなかった。従って、このような設定は近似的には十分と考えることができ、上記の結果は妥当なものと結論づけられる。

3.2 連続型の例

ここでは、ふたつのファジィ集合を同時に考えた例を示す。 $n=89$ 人の大学生の集団に対し次のようなアンケートを行った。1) 身長が「低い」あるいは「高い」、2) 年齢が「下」あるいは「上」という言葉の意味を、前節と同様に0-1の数字に表すとして、その形状が図4のグラフで与えられると仮定する。このとき、各人の主観によると m , d というパラメータの値はどのような数値が適当と考えるかという質問を行った(年齢が「下」、「上」については3.1節と同様の説明を与えた)。ここで、 m はふたつのファジィ集合の境界と考えることができる。また、 $2d$ は、あいまいな

図4 Aのメンバシップ関数(連続型)

図5 EXの推定値のメンバシップ関数

領域の幅とみなすことができる。得られたデータから表2のような統計量が得られた。また、2.3の方法により図5の実線で表される EX の推定値が得られた。前節と同様、メンバシップ関数の型についての評価の調査も同時に行った。ここでは、図4に示されるとおり、ふたつの言葉は否定的な関係にあると仮定しているが、このような点を含めて回答を求めた。その結果、身長に関してはそれほど問題はないが、年齢については、対象の側に個人差が大きく、対象によって解が変化するという指摘がいくつかみられた。従って、身長の場合の図5の結果は妥当なものと言える。しかし年齢の場合は設定が不十分である可能性があり、さらに検討する必要があることが示唆される。なお、折れ線型よりもロジスティック曲線のような形状の方がよいという意見がみられた(前節の場合も同様である)。しかし、近似と考えればそれほど問題はないと思われる。

表2 m, d の標本平均

	身長	年齢
$\bar{m}$	171.4	32.8
$\bar{d}$	7.0	6.0

図5の破線は、パラメータ m, d の平均値の95%信頼区間から得られる α レベル集合の端点の信頼区間を表している。この例では、 f というメンバシップ関数の形が図4で表されるものとしており、たとえば「高い」や「上」の単調増加のメンバシップ関数から得られる α レベル集合の左端は、 $m + (2\alpha - 1)d$ となる。従って、左端の平均値の近似的な信頼区間は、 m, d が独立のとき

$$\left[\bar{m} + (2\alpha - 1)\bar{d} - \frac{z}{\sqrt{n}} \hat{\sigma}, \right.$$

$$\left. \bar{m} + (2\alpha - 1)\bar{d} + \frac{z}{\sqrt{n}} \hat{\sigma} \right]$$

ただし

$$\hat{\sigma}^2 \equiv \hat{\sigma}_m^2 + (2\alpha - 1)^2 \hat{\sigma}_d^2$$

で与えられる。ここで、 z は標準正規分布のパーセント点であり、 $\hat{\sigma}_m^2, \hat{\sigma}_d^2$ は m, d の標本分散を表す

ものとして、正規近似を用いた。また α レベル集合の右端の値は台集合の上限の値(ここでは $+\infty$)となる。

3.3 頻度データによる方法との関係

3.2 節の例に対し、2.4 節で言及した、ふたつのファジィ集合の「境界」のデータからメンバシップ関数を推定する方法との関係について考える。

「境界」のデータからメンバシップ関数を推定することは、 m のデータのみから推定を行うことを意味する。 m のヒストグラムを図6に示す。図7の曲線は、 m の経験分布関数をもとにして「身長が高い」および「年齢が上」というメンバシップ関数を推定した結果を示している。また、図7にプロットされた点は、経験分布関数がジャンプした位置を示している。なお、ここではメンバシップ関数を推定するため、簡便法として3次曲線のあてはめを行った(藪本他[4]はスプライン関数による補間を用いている)。しかし、得られた曲線は、あくまでもメンバシップ関数の位置パラメータ m の確率的なばらつきを表現するものであって、言葉の意味を表すファジィネスとは本質的に異なるものであり、図7の曲線をここで問題としている言葉の意味を表すメンバシップ関数とみなすことはできない。従って、結論としては、図7の

折れ線で表される本論文で提案した方法によって得られたメンバシップ関数(図5と同じ)を採用すべきと言える。

ところで、図7の折れ線と m の分布から得られた曲線とを比較すると、あいまいさの幅の程度など、よく似ていることがわかる。すなわち、本質的に全く別個である m の分布関数と、メンバシップ関数とがかなり近いものになっていると言える。この一致は、単なる偶然なのか、あるいはなんらかの経験則の存在の可能性を示唆しているのか、興味を持たれる。このような解析を積み重ねることによって、(I)の方法がどのようなときに妥当な方法といえるのかを、経験的に明らかにすることができるものと期待される。

4. 考察

2.1 で導入したファジィ確率変数によるモデルは、集団における言葉の意味というようなあいまいさを表現するものとして十分現実的なものと考えられる。本論文では、このモデルのもとで、集団から得られるデータによって求めたいメンバシップ関数が客観的に推定可能であるための条件を与えた。ファジィ確率変数の取り得るファジィ集合を離散的にあらかじめ設定し、そのなかから選

図6 ふたつのファジィ集合の境界の分布

図7 境界の分布から得られたメンバシップ関数

択するという形の観測が可能ならば、2.2節の方法1が適用できる。また、あらかじめメンバシップ関数をパラメタライズして、パラメータを観測するような設計が可能であれば、2.3節の方法2が適用できる。これらの条件は、各々の個人に関するメンバシップ関数がなんらかの形で定められるための条件と言える。本論文の結果は、手法の客観性と対象の主観性とは別次元の性質であり、ある程度強い条件が満たされれば、主観的なファジィネスを客観的な方法によって推定することが可能になることを示している。

2.2節の方法1において、3.1節の図2で示したような形でメンバシップ関数を十分な数だけ提示しその中から選択するような調査法は、多くの場合実用的であると思われる。 $A_1, \dots, A_K$ の設定は、常識的な判断に基づいて可能な場合もあるだろうが、そうでない場合は、予備調査を行えば可能となる。しかし、 $A_1, \dots, A_K$ の提示の仕方によっては中心化傾向が生じやすくなるため、注意が必要となる。これに対し、2.3節の方法のようにパラメータを答える形の調査法は、中心化傾向が生じにくいと思われる。ただしこの場合には、関数型が適当な形で与えられている必要があり、また、パラメータの意味が十分に説明される必要がある。この場合も、予備調査が有効となろう。なお、これらのふたつの方法を、次のように一般化することもできる。あらかじめ有限個のパラメタライズされた関数型を提示し、関数型を選択した上でそのパラメータの値を回答するという形の観測を行うならば、2.2および2.3の方法を組み合わせた推定を行うことが可能となる。

予備調査には、評定尺度法が有効となろう。一般に、評定尺度法によって観測される値は、尺度変換によってメンバシップ関数の値とみなすことができる。従って、台集合の各要素に対して評定尺度法を適用することにより、メンバシップ関数が得られる。台集合が連続的でありかつメンバシップ関数が連続ならば、適当な間隔で選ばれた要素に対して評定尺度法を適用することにより、メンバシップ関数を求めることができる。この際、

各個人ごとに観測を繰り返し、平均を取ってメンバシップ関数を推定する方法がある(Norwich and Turksen[12])。また、予備調査は一般には小集団に対してなされるであろうが、予備調査の対象となる集団は、本調査の対象集団と等質ではあるが重複しない集団であることが望ましいと言える。

予備調査などによって取り得る値であるファジィ集合の族が記述された場合は、本論文で提案した方法によって平均的な意味を表すメンバシップ関数を推定することが可能になる。3節の結果は、このような方法が実際に有効であることを示している。ただし、このようにして得られた結果の信頼性は、前提条件の妥当性に依存する。従って、観測の際には、3節で触れたような方法で妥当性の検証を行う必要がある。また、意味のある結果を得るためには、問題としている言葉や概念をどのような文脈で考えているのかを明らかにして、集団内で回答者が共通の理解を持つ状況をつくる必要がある。

以上のように、主観的なものを客観的な方法で推定することを可能にするためには、先験的な情報を必要とする。逆に、個人の主観がなんらかの形でデータに反映されていない限り、得られた結果を問題としている主観性と関連付けることは原則としてできない。このような条件が満たされていない場合には客観的な推定は困難となる。しかし、推定可能性とは独立に、本論文のモデルは、集団における言葉の意味というようなあいまいさを理論的に考察する上での基本的モデルとして意味を持つ。ひとつの応用として、3.3節で示したように、(I)の方法の妥当性の検証に用いることができる。

ところで、ファジィ確率変数の期待値を推定するという立場にたてば、取り得る値である個々のファジィ集合の同定に対しては、それほどの精密さは要求されないと期待される。その理由は、ここで求めようとしている期待値は平均的人物にかかわるものであり、過剰な精度はあまり意味がないであろうという点と、観測されたファジィ集合

の標本平均を取るという操作によって個々の同定された結果の荒さが相殺されることが期待されるという点にある。後者については一般に保証されているわけではないので、経験的に判断する必要があるだろう。

ここでは、一個人に関するメンバシップ関数に関しては、評定尺度法などでメンバシップ関数を同定することができると仮定して議論してきた。しかし、このような方法についても研究すべき点が残されている([5])。この点については今後の課題と言えよう。

5. まとめ

一般に、データから客観的な方法でメンバシップ関数を推定した場合、推定されたものはどのような意味を持つメンバシップ関数であるのかが明確でない場合がある。このような問題に対してはファジィ確率変数が直接応用でき、またそのことによって推定問題が明確に説明できることを示した。そして、メンバシップ関数が客観的に推定可能であるための条件について考察し、具体的な推定法を提案した。さらにこの推定法を実際例に適用し、このようなメンバシップ関数の推定法が有効であることを示した。また、頻度データから推定する方法には理論的裏付けがないことを指摘すると同時に、この方法が妥当なものであるかどうかを経験的に検証するための理論的枠組を提供できることを示した。

付録. ファジィ確率変数について

ファジィ確率変数は次のように定義される。標本空間 Ω 、 Ω 上のボレル集合体 B 、確率測度 P からなる確率空間 (Ω, B, P) が与えられているとする。あるファジィ集合の族 Λ があり、 Ω から Λ への写像 X が可測性に関する条件を満たすとき、 X をファジィ確率変数と呼ぶ(Kwakernaak[9,10], Puri and Ralescu[13])。基本的には、Kwakernaak[9]による定義と Puri and Ralescu[13]による定義は同等なものと考えてよい。しかし、フ

ァジィ集合の族 Λ の条件が異なっている。理論的には、Puri and Ralescu による条件の方がゆるやかであり、また、完備性など理論的に望ましい性質が Puri and Ralescu[13]によって示されている。従って、本論文では Puri and Ralescu によるファジィ確率変数の定義に従う。

ファジィ確率変数に対しては適当な条件のもとで期待値が定義される。ここでは Puri and Ralescu[13]による次の定義に従う。

$$(EX)_\alpha \equiv \int (X)_\alpha dP$$

ここで $(\cdot)_\alpha$ は α レベル集合を表し、右辺の積分は集合値関数に対する Aumann 積分[6]で与えられる。このように定義された期待値は、適当な条件のもとで(1)式や(5)式のように表されることがわかる([13]参照)。Kwakernaak[9]においてもファジィ確率変数の期待値が定義されており、多くの場合 Puri and Ralescu による定義と一致することがわかるが、Kwakernaak[9]では常に凸化して定義するという制約が課されている。従って、期待値の定義についても Puri and Ralescu[13]による定義がより自然であると言える。

なお、ハイブリッド数(Kaufmann and Gupta[8])は、ファジィ確率変数のひとつの例とみなすことができる。3.2節の例において、 d が一定の値をとるときのファジィ確率変数 X は、ハイブリッド数になっている。

ファジィ確率変数およびその期待値の統計的推論への応用は、渡辺・今泉[3]、渡辺[2]、Watanabe and Imaizumi[15]などでなされている。

参考文献

- [1] 竹内啓、藤野和建：「2項分布とポアソン分布」，東京大学出版会，1981.
- [2] 渡辺則生：ファジィ規準による統計的モデル選択。(投稿中)
- [3] 渡辺則生、今泉忠：ファジィデータの統計的モデル，日本ファジィ学会誌，3，4，755-766，1991.
- [4] 藪本剛、宮本定明、村上征文：スプライン関数による確率密度推定法を用いたメンバシップ関数の

- 同定, 第8回「ファジィシステムシンポジウム」講演論文集, 261-264, 1992.
- [5] 山下利之: 心理学におけるファジィ理論的アプローチ, 日本ファジィ学会誌, 2, 4, 542-553, 1990.
- [6] Aumann, R. J.: Integrals of set-valued function, J. Math. Anal. Appl., 12, 1-12, 1965.
- [7] Dubois, D. and Prade, H.: Unfair coins and necessity measures: Towards a possibilistic interpretation of histograms, Fuzzy Sets and Systems, 10, 15-20, 1983.
- [8] Kaufmann, A. and Gupta, M. M.: "Introduction to Fuzzy Arithmetic: Theory and Applications", Van Nostrand Reinhold Company, New York, 1985. (邦訳: 田中監訳「ファジィ数値と応用」オーム社, 1993)
- [9] Kwakernaak, H.: Fuzzy random variables-I. Definitions and theorems, Information Sci., 15, 1-29, 1978.
- [10] Kwakernaak, H.: Fuzzy random variables-II. Algorithms and examples for discrete case, Information Sci., 17, 253-278, 1979.
- [11] Leung, Y.: "Spatial Analysis and Planning under Imprecision", North-Holland, Amsterdam, 1988.
- [12] Norwich, A. M. and Turksen, I. B.: A model for the measurement of membership and the consequences of its empirical implementation, Fuzzy Set and Systems, 12, 1-25, 1984.
- [13] Puri, M. L. and Ralescu, D. A.: Fuzzy random variables, J. Math. Anal. Appl. 114, 409-422, 1986.
- [14] Thole, V., Zimmermann, H.-J. and Zysno, P.: On the suitability of minimum and product operators for the intersection of fuzzy sets, Fuzzy Sets and Systems, 2, 167-180, 1979.
- [15] Watanabe, N. and Imaizumi, T.: A fuzzy statistical test of fuzzy hypotheses, Fuzzy Sets and Systems, 53, 167-178, 1993.

(1992年9月14日 受付)

(1993年2月15日 再受付)

[問い合わせ先]

〒112 東京都文京区春日1-13-27

中央大学 理工学部 管理工学科

渡辺 則生

☎: 03-3817-1949

☎: 03-3817-1943

著者紹介

渡辺 則生 (わたなべ のりお)

中央大学理工学部管理工学科

昭和55年東京工業大学理学部数学科卒業。昭和57年東京工業大学大学院総合理工学研究科システム科学専攻修士課程修了。昭和60年同博士後期課程退学。昭和60年自治医科大学助手。平成元年中央大学理工学部講師。平成4年同助教授、現在に至る。理学博士。時系列解析、統計一般理論の研究に従事。日本統計学会、日本ファジィ学会などの会員。