特集 オペレーションズリサーチにおけるファジィ応用

ファジィデータからの統計的推論[†]

奥田 徹示*

1. まえがき

ここでは、主観的なあいまいさを伴う観測値を用いる統計的な推測問題について考察する。統計学の分野において、正確な観測ができないかまたはコストの関係で、ある区間に属する事象が何件発生したかというような、いわゆるグルーピングデータに関する分析が古くよりなされてきている。特に、正規分布のような比較的なめらかな分布を念頭においた Sheppard の補正は現在でも有名で、十分実用性を持っていると思われる。この補正については、1898 年 Sheppard[1]によって論文が出され、pairman & Peason[2]や Fisher[3]をはじめ多くの研究者によって 1930 年代まで、主としてモーメント推定に関して多くの分析がなされたが、1943 年の Kendall らの著書[4]にそれまでの研究成果がまとめられている。最近では、最尤推定、回帰分析等いろいろな分野で研究されている。

ここでは、あいまいな観測をファジィ事象として表現して、ファジィ観測から得られるデータに基づいて母集団推測を行なう方法について考察し、一般的なデータ解析の考え方を基本として[5], [6], 主として Sheppard の補正と同様の考え方[4], [7]に基づいてファジィ観測からモーメントとキュムラントを推定するための補正公式について述べる[8]。

2. ファジィ観測の形式と母集団推測

いま、母集団の密度関数を $f(x|\theta)$ としておく。 θ は母数を意味している。特に、必要のない場合は θ を省略する。通常観測値はある 1 点 $x_j \in X$ が観測されたものとして、これをデータの値として用いる。ここでは、点 x_j のまわりのある幅 H に主観的なあいまいさを伴っている場合を考える。このようなファジィ観測を区間 $H_j = [x_j - h/2, x_j + h/2]$ で定義されるファジィ事象としてとらえる。メンバシップ関数 $\chi_j: H_j \rightarrow [0, 1]$ は、 $x \in H_j$ のファジィ事象 χ_j への帰属度を表わしており、 $\chi_j(x_j) = 1$ であるものとする。

さて、このようなファジィ事象として観測値を受け取る場合、ここではこれをファジィ観測値、ないしはファジィデータと呼ぶことにする。このファジィ観測値の出現確率は密度関数 $f(x)$ のもとでは、Zadeh[9]の定義より、

$$P_j = \int_{-h/2}^{h/2} \chi(v) f(x_j + v) dv \quad (1)$$

と表わされる。ここで、 $\chi(v)$ は区間 $[-h/2, h/2]$ で定義されるメンバシップ関数で、 $\chi(x - x_j) = \chi_j(x)$ と変換している。この確率を、区間 H_j の中心値 x_j をこのファジィ観測値 χ_j の代表点とみなし、代表値 x_j の出現確率と考える。これら代表値を用いて統計的処理を施そうとするのがここでの考え方である。なお、 $x_{j+1} = x_j + h/2$ で、 $\sum P_j = 1$ を満たすような左右対称のメンバシップ関数をここでは考えるものとする。

代表的なものとして、次の 3 種類の型を示しておく。

[†] Statistical Inference From Fuzzy Data
Tetsuji OKUDA

* 大阪工業大学 経営工学科
Department of Industrial Engineering, Osaka Institute of Technology

[三角形型]

$$\chi(v) = -2|v|/h + 1,$$

[2次曲線型]

$$\chi(v) = (8/h^2)(|v| - h/2)^2[2(|v| + h/4)/h] \\ + (1 - 8v^2/h^2)\{1 - [2(|v| + h/4)/h]\},$$

[台形型]

$$\chi(v) = \max\{0, 1 \wedge (-|v|/(2e) + h/(8e) + 1/2)\}.$$

ここで、 $[A]$ は A の整数部を意味し、 v および e の範囲は $-h/2 \leq v \leq h/2$, $0 < e \leq h/4$ である。

さて、 d_r をファジィ事象 χ の r 次の“ F -モーメント”とし、

$$d_r = \int_{-h/2}^{h/2} x^r \chi(x) dx \quad (2)$$

と定義する。そこで、ファジィデータに対する r 次のモーメントを計算してみると、

$$m_r = \sum_{j=-\infty}^{\infty} x_j^r \int_{-h/2}^{h/2} \chi(v) f(x_j + v) dv \\ \doteq \frac{2}{h} \int_{-\infty}^{\infty} \int_{-h/2}^{h/2} (u-v)^r \chi(v) f(u) du dv \\ = \frac{2}{h} \sum_{j=0}^r ({}_r C_j) (-1)^j d_j \mu_{r-j} \quad (3)$$

となる。ここで、“ $\doteq$ ”はEuler-Maclaurinの和分公式に基づく近似誤差で、ここでは近似が十分意味をなしている場合を考える。従って、この段階での“ $\doteq$ ”は、以後単に“ $=$ ”と記すことにする。また、 μ_k は母集団の k 次のモーメント(正確なモーメント)を表わしている。

各メンバシップ関数の型に対する d_r は、 r が偶数のとき

[三角形型]

$$d_r = \frac{h^{r+1}}{(r+2)(r+1)2^r},$$

[2次曲線型]

$$d_r = \frac{(2^{r+2}-1)h^{r+1}}{(r+3)(r+2)(r+1)2^{2r}},$$

[台形型]

$$d_r = \frac{\{(1/4 + e/h)^{r+2} - (1/4 - e/h)^{r+2}\} h^{r+1}}{(r+2)(r+1)e/h}$$

となり、 r が奇数のときは $d_r=0$ となる。

したがって、たとえば三角形型の場合には、

$$\left. \begin{aligned} \mu_1 &= m_1, \\ \mu_2 &= m_2 - (1/24)h^2, \\ \mu_3 &= m_3 - (1/8)m_1 h^2, \\ \mu_4 &= m_4 - (1/4)m_2 h^2 + (1/160)h^4, \\ \mu_5 &= m_5 - (5/12)m_3 h^2 + (1/32)m_1 h^4, \\ \mu_6 &= m_6 - (5/8)m_4 h^2 + (3/32)m_2 h^4 - (5/2688)h^6, \\ &\dots\dots\dots \end{aligned} \right\} \quad (4)$$

というような補正公式を得る。これは、正確なモーメント μ_r がファジィデータを用いたモーメント m_r から補正公式によって導き出されることを示している。しかも、いまの場合あいまいさの幅 h によって誤差の大きさが評価できるようになっている。すなわち、幅 h が比較的小さければ、補正は h^2 の項までで十分実用的であると考えられる。したがって、 h は1以下となることが望ましいが、シミュレーション等の経験からは必ずしもこのことが絶対的な条件ではないことが分かる。

なお、正確な補正公式は次のようになる。

[三角形型]

$$m_r = \sum_{j=0}^{[r/2]} (2/\{(2j+2)(2j+1)\}) \\ \times (1/2)^{2j} ({}_r C_{2j}) h^{2j} \mu_{r-2j},$$

[2次曲線型]

$$m_r = \sum_{j=0}^{[r/2]} (2/\{(2j+3)(2j+2)(2j+1)\}) \\ \times \{(1/2)^{2j-2} - (1/4)^{2j}\} ({}_r C_{2j}) h^{2j} \mu_{r-2j},$$

[台形型]

$$m_r = \sum_{j=0}^{[r/2]} (2/\{(2j+2)(2j+1)e/h\}) \\ \times \{(1/4 + e/h)^{j+2} - (1/4 - e/h)^{2j+2}\} \\ \times ({}_r C_{2j}) h^{2j} \mu_{r-2j}.$$

なお、正規母集団の場合にはエルミート多項式を利用して、同様の補正公式を導くこともできる。また、2次曲線型のメンバシップ関数を次のように n 次の関数

[n 次曲線型]

$$\chi(v) = \begin{cases} (2^{2n-1}/h^n) \{v+h/2\}^n; & -h/2 \leq v \leq -h/4 \\ -(2^{2n-1}/h^n) v^n + 1; & -h/4 \leq v \leq h/4 \\ (2^{2n-1}/h^n) \{v-h/2\}^n; & h/4 \leq v \leq h/2 \end{cases}$$

に一般化したときは(ただし、 n は偶数),

$$d_r = \frac{2^{n-1}}{2^{2r+2}} \left\{ \frac{2}{2^n(r+1)} \frac{n+r+1}{(n+r+1)} + \sum_{j=0}^n ({}_nC_j) \frac{(-1)^{n-j} (2^{r+1+j}-1)}{2^j(r+1+j)} \right\} h^{r+1}$$

となる。なお、この n 次曲線型で、 $n=2$ とすると、 d_r の形は2次曲線型と完全に一致する。また、台形型のメンバシップ関数で $e=h/4$ とすると三角形型に一致し、 $\lim d_r (e \rightarrow 0)$ は区間幅 $[-h/4, h/4]$ で単に級分けするような通常の Sheppard の補正の場合と一致する。

次に、キュムラントの補正について考察する。代表値 x_j から計算される積率母関数は

$$\phi(s) = \sum_{j=-\infty}^{\infty} \exp(sx_j) \int_{-h/2}^{h/2} \chi(v) f(x_j+v) dv \quad (5)$$

であり、Euler-Maclaurin の和分公式が妥当性を失わないとして、

$$\phi(s) = (2/h) \int_{-\infty}^{\infty} \int_{-h/2}^{h/2} \exp\{s(u-v)\} \times \chi(v) f(u) du dv \quad (6)$$

が導ける。そこで、

$$\psi(s) = (2/h) \int_{-h/2}^{h/2} \exp(-sv) \chi(v) dv \quad (7)$$

として、 $\phi(s)$ の対数をとって、

$$\begin{aligned} \log \phi(s) &= \log \psi(s) + \log \int_{-\infty}^{\infty} \exp(su) f(u) du \\ &\equiv \sum_{k=1}^{\infty} c_k s^k / (k!) \end{aligned} \quad (8)$$

と表わせば、 c_k はファジィ観測値の代表値から計算される生のキュムラントということになる。つまり、 β_k を密度関数 $f(x)$ に対する正確なキュムラントとすると、

$$\sum c_k s^k / (k!) = \log \psi(s) + \sum \beta_k s^k / (k!) \quad (9)$$

なる関係が成立することになる。したがって、 $\log \psi(s)$ のベキ展開を求めれば、キュムラントの補正が導けることになる。

ところで、Bernoulli 多項式

$$z / (\exp(z) - 1) = \sum_{j=0}^{\infty} B_j z^j / (j!) \quad (10)$$

を

$$\frac{2}{\exp(z)-1} - \frac{2}{z+1} = 2 \sum_{j=2}^{\infty} \frac{B_j z^{j-1}}{j!} \quad (11)$$

と変形して、この両辺を0から z まで積分すると、

$$\log \frac{2 \cosh(z) - 2}{z^2} = 2 \sum_{j=2}^{\infty} \frac{B_j z^j}{(j!) j} \quad (12)$$

となる。 B_j は Bernoulli 数と呼ばれ、 $B_0=1$, $B_1=-1/2$, $B_2=1/6$, $B_4=-1/30, \dots, B_3=B_5=\dots=0$ の値を持っている。実は、三角形型の場合、

$$\psi(s) = \frac{2 \cosh(sh/2) - 2}{(sh/2)^2} \quad (13)$$

となるので、式(12)の関係をを用いると、

$$\log \psi(s) = \sum \{ B_j h^j / (j 2^{j-1}) \} \{ s^j / (j!) \} \quad (14)$$

となり、

$$c_r = B_r h^r / (r 2^{r-1}) + \beta_r (r > 1) \quad (15)$$

なる関係が得られる。したがって、キュムラントの補正公式として、

$$\left. \begin{aligned} \beta_2 &= c_2 - h^2/24, \\ \beta_4 &= c_4 + h^4/960, \\ \beta_6 &= c_6 - h^6/8064, \\ &\dots \end{aligned} \right\} \quad (16)$$

が導かれる。

もちろん、この補正公式は、一般的によく知られているモーメントとキュムラントの関係式から、すでに求めたモーメントの補正公式を用いて導くこともできるが、公式の形に表わすにはここでの計算法が便利であろう。

同様な計算法を用いて、2次曲線型の場合、

$$\psi(s) = \frac{\sinh(sh/4) (2 \cosh(sh/4) - 2)}{(sh/4)^3}, \quad (17)$$

$$\beta_r = c_r - B_r(2^r + 2) h^r / (r 2^r) \quad (r > 1), \quad (18)$$

台形型の場合、

$$\psi(s) = \frac{\sinh(se)}{se} \cdot \frac{\sinh(sh/4)}{sh/4}, \quad (19)$$

$$\beta_r = c_r - (B_r h^r / r) \{1/2^r + (2/q)^r\} \quad (r > 1) \quad (20)$$

が導かれる。ただし、 $e=qh$ と変換している。

三角形型、2次曲線型、台形型の何れの場合も、 r 次の補正項のオーダーは h^r になっている。したがって、 h がそれほど小さくなく、 h^4 以上の項が無視できるとすれば、キュムラントの補正は2次の項

のみについて考慮すればよいことになる。たとえば、 $N(\mu, \sigma^2)$ からのファジィ観測値の分布は、 $0(h^r)$ ($r \geq 4$) の項が無視できる場合、 $c_1 \doteq \mu$, $c_2 \doteq \sigma^2 + h^2/24$, $c_3 \doteq c_4 \doteq \dots \doteq 0$ となるので、ファジィ観測値の分布も、ほぼ平均が μ で、分散が $\sigma^2 + h^2/24$ の正規分布と考えても大きな障害にはならないと考えられる場合があるであろう。このような意味で、キュムラントの補正公式は理論上の見通しを分かりやすくするといった利点があるのではないと思われる。

3. 数値例

ここでは、数値例として次のベータ分布

$$dF = \{x^9(1-x)/B(10,2)\} dx, \quad 0 \leq x \leq 1 \quad (21)$$

について考える[8]。 $h=0.4$ として、三角形型のファジィ観測の場合について、表1の頻度を用いてモーメントとキュムラントの補正について調べるとそれぞれ表2、および表3のようになる。

表1. ベータ分布に対する頻度

k	x_k	P_k	k	x_k	P_k
0	0.0	0.0000001	3	0.6	0.1251655
1	0.2	0.0001396	4	0.8	0.5560721
2	0.4	0.0087903	5	1.0	0.3098322

(計)0.9999998

表2. モーメントの補正

k (次数)	μ_k (真値)	m_k (非補正)	(誤差)	補正值	h^2	(誤差)
1	0.8333333	0.8333332	(-0.0000001)			
2	0.7051282	0.7121900	(0.0070618)	0.7055233	(0.0003951)	
3	0.6043956	0.6221406	(0.0177450)	0.6054739	(0.0010783)	
4	0.5238695	0.5540460	(0.0301765)	0.5255584	(0.0016889)	

表3. キュムラントの補正

k (次数)	β_k (真値)	c_k (非補正)	(誤差)	補正值	(誤差)
1	0.8333333	0.8333332	(-0.0000001)	—	—
2	0.0106838	0.0177458	(0.0070620)	0.0110791	(0.0003953)
3	-0.0010176	-0.0009273	(0.0000903)	—	—
4	-0.0000901	-0.0000014	(0.0000887)	0.0000253	(0.0001154)

モーメントの補正については、補正前の誤差が $1/100$ のオーダーであるが、 h^2 の項までの補正後の誤差のオーダーは $1/1000$ のオーダーになっている。キュムラントについては、もともと値が小さいため、2 次のキュムラントでは補正前の誤差が $1/1000$ のオーダーであったものが補正後は $1/10000$ のオーダーとなっている。4 次のキュムラントについて注目してみると、補正をしたからといって必ずしも誤差が小さくはなっていない。これは、補正項そのものが h^4 のオーダーであり、Euler-Maclaurin の和分公式の近似誤差が影響しているものと思われる。すなわち、 h^4 の補正項の値がすでに近似誤差の範囲に入っているものと考えられる。しかしながら、誤差そのものの値は $1/10000$ のオーダーでありそれほど問題にはならないであろう。ここでの数値例からも分かるように、補正は h^2 の項までで十分実用的ではないかと考えられる。

なお、あいまいさの幅 h を標準偏差 σ との比で考慮したい場合は、 $x/\sigma = u$, $h/\sigma = \alpha$ などとスケール変換しておけばよい。このようにすれば、 h の役割は α に置き換えられる。ここでは標準偏差の値 σ は 0.10336 であり、 h の値よりも小さくなっているが、先にも述べたように、実用上は必ずしも α が 1 以下でなければならないということはない。

ところで、ここでの補正が正しく働くかどうかは、Euler-Maclaurin の和分公式の適応の妥当性によるが[10]、一般的には密度関数 $f(x)$ の尾部が

高次の滑らかさを保つことが条件となる。いま、区間 $[a, b]$ を n 等分して、 $h = (b-a)/n$, $x_j = a + (j/2)h$ とすると、

$$z_r(x) = x^r \int_{-h/2}^{h/2} \chi(v) f(x+v) dv$$

として、Euler-Maclaurin の和分公式から、

$$\begin{aligned} \sum_{j=1}^{2n-1} z_r(x_j) &= (2/h) \int_{-h/2}^{h/2} z_r(x) dx \\ &+ \sum_{j=1}^{2m} (B_k/k!) (h/2)^{k-1} \{z_r^{(k-1)}(b) \\ &- z_r^{(k-1)}(a)\} + R^{2m}, \\ R^{2m} &= \frac{n B_{2m+2} (h/2)^{2m+2}}{(2m+2)!} z_r^{(2m+2)}(v), \\ &a \leq v \leq b \end{aligned}$$

となる。したがって、 $|x| \rightarrow \infty$ において、 $z_r(x)$ およびその導関数がことごとく 0 に漸近するのであれば、和分から積分への近似は意味を失わない。つまり、密度関数の尾部が高次の滑らかさを持つことが必要となる。これは、通常の Sheppard の補正に対する条件と同じである[4], [11]。たとえば、Pairman & Peason[2]によって、尾部が打ち切られたような正規分布に対しては、Sheppard の補正が適用困難であることが示されている。また、近似誤差については、Fisher[3]によってフーリエ展開を利用した評価などもなされているが、ここで示した方法は Kendall ら[4]と同じ方法で、文献[11]では Sheppard の補正の成立条件が議論されている。

4. ファジィ観測のデータ処理への影響

まず、ここで導いた補正項をもとに不偏推定量を構成することができるかどうかを調べておく。補正公式を $\mu_r = m_r - c_r$ として、 $-c_r$ は r 次モーメントの補正項を表わすものとしておく。また、ファジィ観測の代表値を実現値としてとる確率変数を X として、確率変数 X に対する観測系列 $\{X_1, \dots, X_n\}$ は独立であるとする。すなわち、

$$P_r\{X_1 = x_1, \dots, X_n = x_n\} = \prod_{j=1}^n P_r\{X_j = x_j\} \quad (22)$$

としておく。

さて、ファジィ観測からの標本モーメントとして、

$$m_r' = \frac{1}{n} \sum_{j=1}^n X_j^r - c_r \quad (23)$$

を定義すると、簡単な計算から $E(m_r') = \mu_r$ となることが分かる。すなわち、 m_r' は μ_r の不偏推定量として用いることができる。参考までに標準偏差を求めておくと、

$$\sqrt{\text{Var}(m_r')} = \sqrt{\mu_{2r} - \mu_r^2 + c_{2r} - c_r^2 - 2c_r\mu_r} / \sqrt{n} \quad (24)$$

となる。なお、

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2 - c_2, \quad \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \quad (25)$$

とおくと、

$$\begin{aligned} E(S_n^2) &= \frac{n}{n-1} \left\{ E(X^2) - \frac{2}{n} E\left\{X_j^2 + \sum_{k=1(k \neq j)}^n X_j X_k\right\} \right. \\ &\quad \left. + \frac{1}{n} E\left\{\sum_j X_j^2 + \sum_{i \neq j} X_i X_k\right\} - c_2 \right\} = \mu_2 - \mu_1^2 \quad (26) \end{aligned}$$

となって、標本モーメントの場合と同じく、 S_n^2 は母集団の分散の不偏推定量になっていることが分かる。このように補正項を用いることによって、ファジィデータを用いた場合でも簡単に不偏推定量をつくることができる。

さて、ファジィ観測が推定に与える影響をもう少し具体的に調べる。いま、 $y_j \in [x_j - h/4, x_j + h/4]$ なる y_j が現実のデータとして発生した場合、 x_j でデータ処理がなされるとする。この場合、三角形型など一般的なメンバシップ関数の場合、

$$x_j(y_j) = x_{j-1}(y_j) \vee x_j(y_j) \vee x_{j+1}(y_j) \quad (27)$$

となっている。すなわち、現実には y_j が発生しているが、観測がファジィになっているので具体的な処理は y_j の帰属度が一番大きくなるファジィ観測の代表値 x_j でなされるものとして粗い近似を試みる。そこで、 $y_j - x_j$ を観測誤差と考えて、この誤差について分析することにする。

誤差を表わす確率変数を W として、 $W = Y - X$ とする。 $\Delta w \rightarrow 0$ のとき、

$$P_r\{X = x_j, w_0 < W < w_0 + \Delta w\} = f(x_j + w_0) dw \quad (28)$$

となることを利用して、 $E(X^p W^q)$ タイプの期待値が計算できる。以後の計算に必要なものを計算しておく次のようになる。

$$\begin{aligned} E(W) &= 0, \quad E(X) = \mu_1, \quad E(XW) = -\frac{h^2}{48}, \\ E(W^2) &= \frac{h^2}{48}, \quad E(X^2) = \mu_2 + \frac{h^2}{48}, \quad E(W^3) = 0, \\ E(X^3) &= \mu_3 + \frac{h^2}{8} \mu_1, \quad E(XW^2) = \frac{h^2}{48} \mu_1, \\ E(X^2 W) &= -\frac{h^2}{24} \mu_1, \quad E(W^4) = \frac{h^4}{1280}, \\ E(XW^3) &= -\frac{h^4}{1280}, \quad E(X^2 W^2) = \frac{h^2}{48} \mu_2 + \frac{h^4}{1280}. \end{aligned}$$

そこで、現実のデータを用いて推定する平均値とファジィデータを用いて推定する平均値との標準誤差を求めてみる。計算は単純で、

$$D_1 = \left\{ \frac{1}{n} \sum y_j \right\} - \left\{ \frac{1}{n} \sum x_j \right\} \quad (29)$$

として、

$$\sqrt{E(D_1^2)} = \frac{h}{4\sqrt{3}n} \quad (30)$$

となる。

次に、2次モーメントについて調べてみると、

$$D_2 = \left\{ \frac{1}{n} \sum y_j^2 \right\} - \left\{ \frac{1}{n} \sum x_j^2 - c_2 \right\} \quad (31)$$

として、計算は少し複雑になるが(c_2 のオーダーは通常 $O(h^2)$ となる)、

$$E(D_2^2) = \frac{h^2}{n} \left\{ \frac{\mu_2}{12} + \frac{h^2}{2880} \right\} + \left\{ \frac{h^2}{48} - c_2 \right\}^2 \quad (32)$$

となることが分かる。したがって、 h^4 以上の項を無視すれば、

$$\sqrt{E(D_2^2)} \doteq \frac{h\sqrt{\mu_2}}{2\sqrt{3}n} \quad (33)$$

となる。

さらに、分散の推定について調べてみると、

$$D \equiv \left\{ \frac{1}{n-1} \sum (y_j - \bar{y})^2 \right\} - \left\{ \frac{1}{n-1} \sum (x_j - \bar{x})^2 - c_2 \right\} \quad (34)$$

として、計算はかなり複雑になってしまうが、

$$\begin{aligned} E(D_s^2) &= \frac{h^2(\mu_2 - \mu_1^2)}{12(n-1)} \\ &+ \frac{h^4}{n-1} \left\{ \frac{n-1}{n} \frac{1}{1280} - \frac{n+1}{n} \frac{1}{48^2} \right\} + \left\{ \frac{h^4}{48} - c_2 \right\}^2 \end{aligned} \quad (35)$$

が得られ、 h^4 以上の項を無視すれば、

$$\sqrt{E(D_s^2)} \doteq \frac{h\sqrt{\mu_2 - \mu_1^2}}{2\sqrt{3}n-1} \quad (36)$$

となる。

以上より、ファジィデータから平均や分散を推定する場合、補正項を用いることによって正確なデータに基づいて推定する場合と比べてその誤差がデータ数 n と共に減少していくことが式(30)、(33)、(36)から分かる。この種の議論は、Sheppardの補正に関連して、分類に基づく標準誤差としてFisher[12]によって議論されていることと同じ議論である。なお、一般の r 次のモーメント推定に対しては、その計算は複雑になる。しかしながら、重要なことは補正項を利用して一般的に母集団の推測を行う場合、ある程度標本サイズが大

きければ観測に伴うあいまいさの影響が解消できると考えられることである。このような意味で、ここでの補正項をファジィデータに対処するために各種の統計的データ解析法に用いることの実用性を示しているのではないかと考えられる。

明らかに、以上の議論はメンバシップ関数の度合いを厳密には考慮していない粗い議論である。そこで、メンバシップ関数の度合いを厳密に考慮して、モンテカルロ実験に基づくシミュレーションを実施している[8]。その結果、生のデータに基づくモーメントの推定結果とファジィデータ(ファジィ観測値の中心値)に基づくモーメントの推定結果とには一定のバイアスが生じ、補正を施すことによってそのバイアスが解消するというここでの理論的な結果を裏付ける結果を得ている。ただし、実際のシミュレーションでは、現実のデータ解析への適用を考えて、メンバシップ関数の形が多少変動してもかまわないというように、理論的な設定条件を緩和した状況のもとで実行している。また、先にもふれたが、 α が1以上の場合に対してもシミュレーションを実行したが、 α がそれほど大きくなければ良好な結果が得られる。

このようなメンバシップ関数の設定についてはいろいろと議論のあるところであろうが、本研究では実用性に主眼をおいている。すなわち、ファジィデータの取得時において、観測者が何個かのファジィデータを取るときにいちいちメンバシップ関数の形を報告せねばならないとすれば、それは少し観測者への負担が大きすぎるのではないかと考えられる。元来、実用的なデータ解析の研究では、現実の観測を実行する観測者側とデータの解析者側との負担が適度にバランスがとれていることが必要であろう。たとえば、正規分布に基づく統計理論の構築と誤差が適度に管理されてほぼ正規分布をしていることが、統計理論を有用なものにするとの考え方があるが[5]、これと類似するようなことである。ある特定のメンバシップ関数を最初に設定するという前提のもとで理論を構成し、データ解析時においてメンバシップ関数の形を精密に推定しなければならないような議論展開

は、実用性の観点からは少し不利であろう。そこで、本研究では一応基本的なメンバシップ関数を定めて標準的な理論を構成しておき、理論の適用に際しては理論構成の前提条件が若干崩れてもそれほど影響がないことをシミュレーション等によって検討している。その結果、比較的良好な結果を得ている[13], [14]。このように、観測者側に極度の負担を掛けることなく、大体いくらぐらいのデータが得られたかをどのファジィ区間のデータが発生したかというような形でのみ観測者から報告してもらい、後は観測者の持つおおよそのメンバシップ関数を推定して、それに対して標準的な理論を適用し、少しぐらいのメンバシップ関数の狂いは影響しないというような形での解析を行おうとしているのがここでの立場である。もともとあいまいなデータを取り扱おうとしているのであるから、直接的にメンバシップ関数の形に依存するような議論よりも、大体のメンバシップ関数の形を前提として、後はファジィ理論を適用して解析者側で観測者の持つあいまいさをうまく処理するというような方法を取り、観測者があいまいなデータを厳密に報告しなければならないというような状況は避けたほうがよいのではないかと思われる。

最後に、本モデルの連続モデルへの拡張についてふれておく。本モデルは、もとの母集団は連続空間であるが、空間をいくつかのファジィ区間に区切って議論するため必然的に離散モデルとなる。これは通常の Sheppard の補正を議論する場合と同じことで、元来データをグルーピングして取り扱おうとするのであるから、自然の成りゆきとして離散モデルになるように思われる。ただ、空間をいくつかの区間に区切るのであるが、その区切り方が連続的に任意になるような場合のモデルがある程度の現実感を持って考えられるであろう。すなわち、本モデルに対するいままでの説明は、ファジィ観測の中心値(刻み位置) x_i が固定された場合であったが、刻み位置が変化すれば補正の効果も変化するかも知れない。そこで、刻み位置が区間 $[X_i - h/4, X_i + h/4]$ でランダムに発生した

場合の生のモーメントの期待値について考察しておく[8]。

いま、 $\{x_j\} (j = \dots, -1, 0, 1, \dots)$ を x の刻み位置の集合とする。このとき、 x_i が区間 $[X_i - h/4, X_i + h/4]$ で一様な連続分布にしたがってランダムに発生するとする。ある刻み位置 x_i が発生したとすると、 x_i を含む刻み位置の集合 $\{x_p\}$ に対して、

$$m_r = \sum_{p=-\infty}^{\infty} x_p^r \int_{-h/2}^{h/2} \chi(v) f(x_p + v) dv$$

となる。したがって、 m_r の期待値は

$$\begin{aligned} E(m_r) &= \frac{2}{h} \sum_{j=-\infty}^{\infty} \int_{X_j - h/4}^{X_j + h/4} \left\{ x^m \int_{-h/2}^{h/2} \chi(v) f(x + v) dv \right\} dx \\ &= \frac{2}{h} \int_{-\infty}^{\infty} x^m \left\{ \int_{-h/2}^{h/2} \chi(v) f(x + v) dv \right\} dx \end{aligned}$$

となり、 m_r の期待値が m_r の積分近似と一致することが分かる。すなわち、平均的には補正が正しく働くことが分かる。

さらに一般的に、連続モデルへの拡張が必要となる場合もあるであろうが、数学的な興味からの議論のみではなく、現実のデータ解析に対する実用性の観点をも含めた議論展開が望まれる。

5. 応用

[多変量解析への応用]

ここでは1変量のモーメントの補正について考察したが、2変量のモーメントの補正についても同様の考え方が適用できる[15], [16]。多変量解析においては、多くの変数間の関連の強さを示す統計量として相関係数が最も頻繁に用いられる。この相関係数を、2変量モーメントの補正公式を用いて、1変量の場合と同様にファジィデータによって推定することができる。したがって、相関係数を利用するような最小2乗法などの手法に対しては、2変量に対する補正公式を直接適用することができることになる[17]。最近、古殿ら[18]、

[19]によって、判別分析や主成分分析への応用がなされ、シミュレーション等によってその実用性が検討されている。この2変量の補正公式では、ファジィデータと通常のデータとの相関を計算することも可能であり、さまざまなデータとファジィデータとが混在するような状況でも適用可能である。したがって、ある種の主観的な意識と人の年齢との相関を求めることなど、社会科学等への応用が考えられる。ここでのモデルの利点は、使用する解析手法が通常の統計理論を基本とするため、導かれた結果が現在の統計理論と同一の理論レベルで解釈でき、データ解析の立場からどのような意味を持つかということが分かりやすくなり、また統計学上の意味を保持できるという点にある。
[標本分布論への応用]

ここでは、モーメントの補正とともにキュムラントの補正についても述べた。キュムラントの補正については既に述べたように、理論上の見通しを分かりやすくするという利点がある。つまり、正規母集団を対象とした場合などは、さきにも述べたように、母集団から抽出されるファジィデータは実際よりは若干大きめの分散を持つ正規母集団にしたがっていると考えるてもよい場合もあることが分かる。このようなことを用いて、ファジィデータに基づく統計量の標本分布を求めることも考えられ、仮説検定などへの応用の可能性がある[20]。また、実際にキュムラントの補正の考え方を品質管理の問題へ応用した例もある[21]。一般にファジィデータからの標本分布を考える場合は、基本的にはスパン $h/2$ の格子点分布を考えることと類似しており、確率分布を近似するための各種の展開公式を利用することなどが考えられるが、シミュレーション等の経験的知識をもとに、どの程度の近似が必要となるかについても考察しておくことが、ここでのファジィデータ解析の実用化にとっては有益ではないかと思われる。

[最尤推定への応用]

ここでは主にモーメント法による推定について説明したが、同様な考え方を最尤推定法に適用することもできる。つまり、ファジィ観測値の代表

値を用いて通常の方法で最尤推定値を求め、ファジィ観測の持つあいまいさの影響によって生じるバイアスを補正項の形で除去しようとする方法である。最尤推定値は一般に漸近的有效推定量となることがよく知られている[6]。この有効推定量を特徴づける量として、フィッシャー情報量があり、ファジィデータを用いて母数を推定した場合、通常のデータを用いて推定する場合に比べてどの程度の情報量が失われるかは、このフィッシャー情報量を計算することによって求めることができる。たとえば、三角形型のファジィ観測のあいまいさの幅 h を標準偏差 σ の半分に取り、正規母集団の平均や分散を推定する場合、フィッシャー情報量の損失比率が約1~2%となることが分かっている[22], [23]。フィッシャー情報量はよく知られているように、推定の漸近効率を与える基本的な量であり、したがって推定効率から考えて、100個のファジィデータは約98~99個の正確に観測されたデータに匹敵することになる。このようにファジィデータに基づく推定によって失われる情報量はそれほど大きなものではないことが分かる。このことは通常のSheppardの補正に対してもよく知られていることである[5]。したがって、補正公式に現れる補正量は標本サイズに無関係なものであり、失われる情報量はそれほど大きくはないのであるから補正を施すことには十分な意味が存在すると考えられる。なお、ファジィデータに基づく最尤推定の考え方を回帰分析へ応用した研究例もある[24]。

[ベイズ論への応用]

さらに、ここでの考え方をベイズ論へ応用することもできる。つまり、ベイズ推定では観測値をもとに事前分布が事後分布に修正されるが、この場合の観測値がファジィデータになった場合である。ここでも、ファジィ観測値の代表値を通常のベイズ法則に適用して事後分布を求め、ファジィ観測に含まれるあいまいさの影響を補正項の形で補正して、あたかも正確な観測値に基づいて修正する場合の事後分布に近似させる方法を考えることができる[25]。このとき、正確な観測値を用い

て修正する事後分布よりも、ファジィデータを用いて修正する事後分布の方がやはり観測にあいまいさの影響のある分だけ緩慢になる。すなわち、不確実性の程度を表わすシャノン型のエントロピーが、ファジィデータに基づく事後分布の方が正確なデータに基づく事後分布よりも大きくなることが分かっている。この研究は未だ初期段階であるが、ファジィデータをベイズ推定に用いることが可能となれば、ベイズ決定分析の分野へも応用可能となる。

[非正規ファジィデータへの対処]

われわれが行ったファジィデータの観測実験は、パーソナルコンピュータのディスプレイを利用した物理的な実験であるが、それらの実験結果の中には、ここで導かれるような補正公式がうまく適応しないものも存在する。その原因の一つは、ファジィ観測値を表現しているメンバシップ関数が非正規になることによると考えられている。このことは、メンバシップ関数の幅 h が少し伸びることを意味しており、現実にも十分起こり得ることであろうと考えられる。確率分布でいえば、分布が少し偏平になってテイルが伸びるというようなことに類似している。そこで、このような非ファジィ正規データに対処するために、メンバシップ関数の形を若干修正して補正公式を修正することを検討している[26], [27]。この方法によって、従来うまく適応できなかったファジィ観測の実験データに対しても、ある程度うまく補正公式が適応できるようになってきている。

以上、ここでは大阪工業大学浅居喜代治教授のご指導のもと同大学院博士課程古殿幸雄氏をはじめ多くの方々と共同研究の内容について紹介させて頂いた。

参 考 文 献

1. Sheppard, W.H. : On the Calculation of the most Probable Values of Frequency-Constants, for Data arranged according to Equidistant Divisions of a Scale, Proc. London Math. Soc., 29, 353-380, 1898.
2. Pairman, E. & Pearson, K. : On Corrections for the Moment-Coefficients of Limited Range Frequency Distributions when There are Finite or Infinite Ordinates and any Slopes of the Range, Biometrika, 12, 231-259, 1919.
3. Fisher, R. A. : On the Mathematical Foundations of Theoretical Statistics, Phil. Tran. Roy. Soc., A, 222, 309-368, 1921.
4. Kendall, M. G. and Stuart, A. : The Advanced Theory of Statistics, Vol.1, C.Grif., London, 1943.
5. 竹内：数理統計学的方法的基礎：東洋経済，1973.
6. 竹内，大橋：統計的推測－2 標本問題，日本評論社，1981.
7. Langdon, W.H. and O. Ore, Semi-Invariants and Sheppard's Correction, Ann. Math., 31, 230-232, 1930.
8. 奥田，古殿，浅居：ファジィ観測データに基づく統計的推定，計測自動制御学会論文集，Vol.26, No.5, 564-571, 1990.
9. Zadeh, L. A. : Probability Measures of Fuzzy Events, J. Math. Anal. Appl., 28, 421-427, 1968.
10. 赤坂：数値解析，コロナ社，1967.
11. Kendall, M.G. : The Conditions under which Sheppard's Corrections are Valid, J. Roy. Stat. Soc., 101, 592-605, 1938.
12. Fisher, R. A. : Statistical Methods for Research Workers, Fourteenth Edition, Hafner Press, New York, 1970.
13. 広沢，浅野，川下，古殿，奥田，浅居：ファジィデータ解析についてのシミュレーション，JIMA 講演予行集，127-128, 1989.5.
14. 古殿，広沢，浅野，川下，奥田，浅居：ファジィデータ解析についての実験的考察，IFSA 日本支部ファジィシステムシンポジウム講演論文集，319-324, 1989.
15. 奥田，古殿，浅居：2変量ファジィ観測に基づく統計的推定，日本ファジィ学会誌，Vol.2, No.3, 415-428, 1990.
16. Okuda, T., Kodono, Y., and Asai, K. : A

- Statistical Inference by Fuzzy Observation Data in the Bivariate Case, Proc. of 3rd IFSA Congress, 755-758, 1989.
17. 川下, 奥田, 浅居: ファジィ観測データによる最小2乗回帰分析, JIMA 講演予行集, 196-197, 1990.11.
 18. 古殿, 奥田, 浅居, 杉浦: ファジィ観測データからの近似最尤推定に基づく判別関数, 日本ファジィ学会誌, Vol.4, No.1, 1992.
 19. 古殿, 奥田, 浅居, 杉浦: ファジィ観測データを用いる主成分分析, 日本ファジィ学会誌, 掲載予定.
 20. 奥田, 広沢, 浅居: ファジィ観測データに基づくキウムラントの補正と統計的仮説検定, 日本ファジィ学会ファジィシステムシンポジウム講演論文集, 611-614, 1991.
 21. 玉木, 金川, 太田: 言語データ多用管理図の概念と設計法, JIMA 講演予行集, 65-68, 1991.5.
 22. 奥田, 古殿, 前原, 浅居: ファジィ観測からの最尤推定, 日本ファジィ学会ファジィシステムシンポジウム講演論文集, 143-146, 1990.
 23. Okuda, T., Kodono, Y., Maehara, K. and Asai, K.: Maximum Likelihood Estimation From Fuzzy Observation Data, Proc. of 4th IFSA Congress, Vol. of Computer, Management & Systems Science, 185-188, 1991.
 24. Okuda, T., Kodono, Y. and Asai, K.: Approximate Maximum Likelihood Estimates in Regression Models for Fuzzy Observation data, in Fuzzy Regression Analysis, ed. by Kacprzyk, J. and Fedrizzi, M., Omnitech Press, Warsaw, 出版予定.
 25. 奥田, 浅野, 浅居: ファジィ観測データによるベイズ推定, JIMA 講演予行集, 202-204, 1990.11.
 26. 前原, 古殿, 奥田, 浅居: 2変量ファジィデータの観測実験と処理方法に関する研究, JIMA 講演予行集, 119-120, 1990.5.
 27. 前原, 古殿, 奥田, 浅居: 非正規ファジィデータに対するモーメントの補正と相関分析への応用, JIMA 講演予行集, 256-257, 1991.5.

(1991年10月30日 受付)

[問い合わせ先]

〒535 大阪市旭区大宮5-16-1
 大阪工業大学 経営工学科
 奥田 徹示
 ☎: 06-952-3131(内線)3714
 FAX: 06-952-6197

著者紹介

奥田 徹示 (おくだ てつじ)

大阪工業大学 経営工学科
 1972年, 大阪府立大学工学部経営工学科卒業. 1977年, 同大学院博士課程経営工学専攻修了. 同年大阪工業大学経営工学科講師, 現在同教授. ファジィ意思決定やファジィデータ解析の研究に従事. 日本ファジィ学会, 国際ファジィ学会, 計測自動制御学会, 日本OR学会, 日本経営工学会などの会員.