λ ファジィ測度とショケ積分を利用した中央値，平均値の中間の評価[*1]

高萩栄一郎

専修大学商学部

Eiichiro Takahagi

Senshu University

要旨

この論文では，ショケ積分を用いて，中央値，算術平均値，および（最大値＋最小値－中央値）の中間の評価方法を示す．この評価モデルは，2 つのサブモデル – 中央値より高い部分を評価するサブモデル (PU) と中央値より低い部分を評価するサブモデル (PL) – に分けられる．この 2 つのサブモデルは，λ ファジィ測度によるショケ積分モデルを使った 最大値，平均値，最小値の中間の評価を用いて評価される．全体のモデルは，(PU + PL − 中央値) で計算される．

1　はじめに

1 組のデータがあり，そのデータの代表値を求める問題を考える．もっとも一般的なものは，平均値であろう．他にも，中央値 (median)，最頻値，区間中央値（(最大値＋最小値)/2) などが考えられる．

平均値は，もっともよく使われるが，その 1 組のデータに異常値や外れ値が含まれていると，平均値の値は，その値に引っ張られと言われている．それに対して，中央値は，異常値が含まれていても，それに対して引っ張られることは少ないと言われている．しかし，中央値は，平均値に比べて，中央値以外のサンプルの変化の影響をその代表値へ反映

[*1] 本稿は，「第 11 回ファジィシステムシンポジウム講演論文集」(1995) での同名の原稿を再構成したものである．

しにくい．あるサンプルの値が増大しても，そのサンプルが中央値より低いサンプルで中央値より大きくならない限り，（またはその逆でない限り）代表値に影響を与えない．[1]では，このような異常値や外れ値に対して強い性質を抵抗性が高いと呼んでいる．すなわち，「データを集約して得られたある統計量が，一部の不適切なデータの影響を受けにくいことを，抵抗性 (resistance) が高いという」[1] と述べられており，また，「統計的仮定からの逸脱の影響を受けにくい統計量は，頑健 (robust) であるといわれている」[1] と述べられている．そこで，本稿では抵抗性という言葉を用いる．

測定の失敗などによる異常値の場合，もしその値が異常値であることが認識できれば，そのサンプルを除外して，分析を行えばよい．しかし，異常値であると認識することが難しい場合が多い．

外れ値の場合，そのデータの代表値への影響を小さくとどめたいのならば，中央値を用いればよい．しかしながら，中央値を用いると先程あげたような中央値以外あまり影響を与えないという弊害が起こる．

人は，ふだんは平均に近い評価をするが，外れ値や異常値がある場合，中央値に近い評価をしているのではないだろうか？ 例えば，ある地域の年間の降雪回数が，90 年は 2 回,91 年が 40 回,92 年が 3 回であったとする．このときの代表値を求めることを考える．平均をとれば 9 回, 中央値をとれば，3 回である．「ここ数年，一冬にだいたた何回位の雪が降るか？」という問いに対して，人は，「4-5 回」と答えるのではないだろうか？ これは，91 年が以上に多かったと判断して，平均の 9 回より小さい値を，でもたまには，大雪が降ることもあると考え，中央値の 3 より大きな数字をこと得たのではないだろうか？

本手法は，代表値を中央値を用いる場合と平均値を用いる場合の中間のものを求めようとするもである．指標 ξ を用いて，0 で中央値，0.5 で平均値，その間を連続的に変化する評価値を求めようとするものである．この指標 ξ は，代表値や評価値を求める目的などによって，主観的に定めることになる．

2 最大値・平均値・最小値の拡張としてのファジィ測度とショケ積分

2.1 定義

一般的な定義にしたがって，ファジィ測度，ショケ積分を定義する．評価項目の集合 X をとし，評価項目の数は，有限個 n とする．すなわち，$X = \{x_1, x_2, \ldots, x_n\}$ とし，$F = 2^X$ とする．

ファジィ測度 μ は，次のように定義する (正規性を仮定).

$$\mu : F \to [0,1] \tag{1}$$

$$A, B \in F, A \subset B \Rightarrow \mu(A) \leq \mu(B) \tag{2}$$

$$\mu(\emptyset) = 0,\ \mu(X) = 1 \tag{3}$$

i 番目の評価値を $h(x_i)$ $(i = 1, \ldots, n)$ とする．すなわち，

$$h : X \to [0, +\infty) \tag{4}$$

とする．

各評価項目の評価値の総合評価値は，ファジィ測度を用いてショケ積分を行うことにより求める．ショケ積分は，

$$z = (C) \int h d\mu \equiv \int_0^{+\infty} \mu(\{x; h(x) > r\}) dr \tag{5}$$

と定義される．また，λ ファジィ測度は，

$$\begin{aligned} &\forall A, B \subseteq X, A \cap B = \emptyset \Rightarrow \\ &\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\mu(A)\mu(B) \\ &\text{where } \lambda \in (-1, +\infty) \end{aligned} \tag{6}$$

と定義する．

2.2 λ ファジィ測度による拡張

λ を与え，すべてのシングルトンのファジィ測度が同じになるように，λ ファジィ測度の定義により各ファジィ測度を同定し，ショケ積分を行えば，最大値，平均値，最小値の拡張となる．図1は，$X = \{x_1, x_2, x_3\}$ の場合，$\mu(\{x_1\}) = \mu(\{x_2\}) = \mu(\{x_3\})$ という関係を保つように，与えられた λ で各ファジィ測度を求めたときの図である．

横軸は，左に $\lambda = +\infty$，右に $\lambda = -1$ となっている．実際は，式 (12) の ξ の値を横軸としている．$\lambda = 0$ と $\lambda = +\infty$ の場合，それぞれの極限を求める．すなわち，$\lambda \to -1$

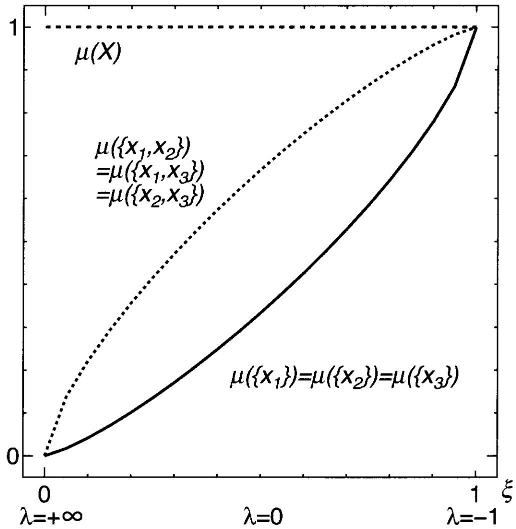

図1 λ とファジィ測度の関係

と $\lambda \to +\infty$ の値を $\lambda = -1$ と $\lambda = +\infty$ の時の値とする．

$$\lambda = -1 \Rightarrow \mu(A) = 1, \forall A \in F \setminus \emptyset \tag{7}$$

$$\lambda = +\infty \Rightarrow \mu(A) = 0, \forall A \in F \setminus X \tag{8}$$

この場合明らかに，最大値 ($\lambda = +\infty$)，平均値 ($\lambda = 0$)，最小値 ($\lambda = -1$) の拡張になっており，λ をそれぞれの中間の値にすることにより，最小値と平均値の間の評価や平均値と最大値の中間の評価を行うことができる．

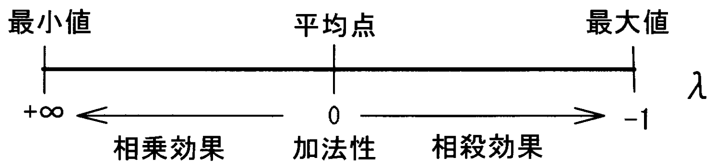

図2　λファジィ測度の意味

ファジィ測度の計算は，塚本 [3] の ϕ_s 変換を利用することができる．ϕ_s 変換は，

$$\phi_s : [0,1] \to [0,1],\ s \in [0,+\infty] \tag{9}$$

$$\phi_s(u) = \begin{cases} [u] & \text{if } s = 0 \\ u & \text{if } s = 1 \\ 1-[1-u] & \text{if } s = +\infty \\ \dfrac{s^u - 1}{s-1} & \text{otherwise} \end{cases} \tag{10}$$

$$\text{where } [u] = \begin{cases} 1 & \text{if } 0 < u \leq 1 \\ 0 & \text{if } u = 0 \end{cases}$$

である．λファジィ測度との関係は，$s = \lambda + 1$ である．

ϕ_s 変換は，任意の $0 \leq a, b, c \leq 1, c = a + b, s \in [0, +\infty]$ に対して，

$$\phi_s(c) = \phi_s(a) + \phi_s(b) + (s-1)\phi_s(a)\phi_s(b) \tag{11}$$

という性質を持っており，λファジィ測度に対応する．したがって，$s = \lambda + 1$ とし，$A \subset S$ のファジィ測度 $\mu(A)$ は，$\phi_s(|\ A\ |/n)$($|\ A\ |$ は，A の要素の数) として求められる．

相乗効果，相殺効果の呼び方は，評価の対象によって異なる．例えば，複数の科目の試験の総合評価の場合，相乗効果は，バランス重視の評価，相殺効果は，個性重視の評価となる．また，骨董品を複数の評価者で評価し，その総合評価値を求める場合，相乗効果は，「悲観的な評価」，相殺効果は「楽観的な評価」となる [2]．実際に λ を与え，ファジィ測度を同定する場合，λ は人間の直観に合いにくい．そこで，次のような指標 ξ を使用する．

$$\xi(s) = \frac{1}{1+\sqrt{s}} = \frac{1}{1+\sqrt{\lambda+1}} \tag{12}$$

表1 相互作用の呼び方

λ	$\lambda > 0$	$-1 < \lambda < 0$
ξ	0.5より小	0.5より大
ファジィ測度	優加法的	劣加法的
一般的な呼び方	補完的 相乗効果	代替的 相殺効果
その他の呼び方(相互作用が外部にある場合)	バランス重視 消極的な 慎重な 悲観的な 保守的な 確実性重視 悪い点がないことを評価	個性重視 積極的な 大胆な 楽観的な 冒険的な 可能性重視 よい点があることを評価
極端な場合	$\xi = 0$	$\xi = 1$
(入力の数基準の場合のみ)	AND MIN(最小値)	OR MAX(最大値)

3 中央値・平均値の拡張としてのファジィ測度とショケ積分

3.1 中央値・平均値の中間の評価への拡張方法

まず，問題を2つ分ける．問題(P)を中央値を基準にして，中央値より大きなデータを評価する問題(PU)と中央値より小さなデータを評価する問題(PL)に分ける．

表2　データ例 (中央値 6)

4	3	3	5	7	20	8	9	2	6	8

PU

$$h^U(x_i) = \begin{cases} h(x_i) - \text{Median} & \text{if } h(x_i) \geq \text{Median} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

$$z^U = (C)\int h^U d\mu \tag{14}$$

PL

$$h^L(x_i) = \begin{cases} 0 & \text{if } h(x_i) \geq \text{Median} \\ \text{Median} - h(x_i) & \text{otheerwise} \end{cases} \tag{15}$$

$$z^L = (C)\int h^L d\mu \tag{16}$$

例題として，表2の11個の数値をあげる．中央値は6，算術平均値は6.81である．問題は図3のようになり，20が外れ値である．図4のように中央値より上のデータを評価する問題PUと，図5のような中央値より下のデータを評価する問題PLに分ける．

問題PUとPLそれぞれについて，共通のξまたはλをを与え，$X=\{x_1, x_2, \ldots, x_n\}$として，ファジィ測度を同定する．すなわち$\mu(A)=\phi_s(\mid A \mid /n)$とする．そして，それぞれの問題について，変換後の評価値を使って，ショケ積分を行い，それぞれの問題の総合評価値z^Uとz^Lを求める．

PUは，中央値より大きいデータの評価であり，それらのデータが中央値からどれくらい上に引っ張っているのかの評価である．$\xi=0$ならば，上に引っ張る程度は，常に0であり，$\xi=0.5$ならば，(PUの) データの平均である．また，$\xi=1$ならば，最大値に等しい．同様に，PLは，下に引っ張る程度である．

全体の問題 (P) の総合評価値は，中央値に，上に引っ張られる程度（PUの総合評価値z^U）と下に引っ張られる程度（PLの総合評価値z^L）を勘案したものであり，次式のよう

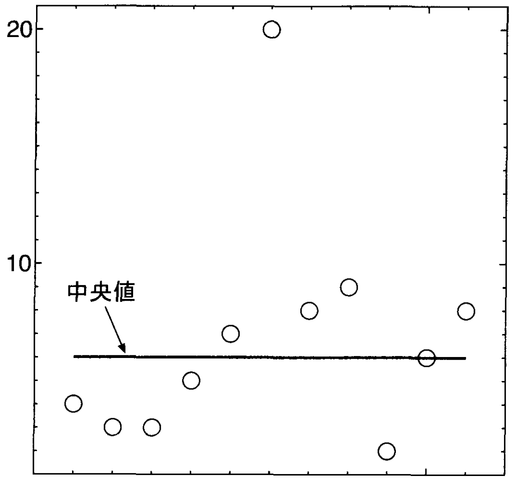

図3 問題P

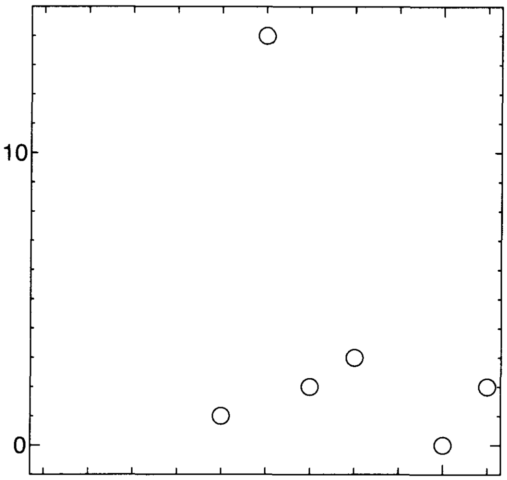

図4 問題PU(中央値より大きいデータを評価)

にする.

$$z = \mathrm{Median} + z^U - z^L \tag{17}$$

$\xi = 0$ ならば，上に引っ張る程度 z^U，下に引っ張る程度 z^L はともに0であり，したがって，総合評価値 z は，中央値に等しくなる．$\xi = 0.5$ ならば，上に引っ張る程度 z^U，下に

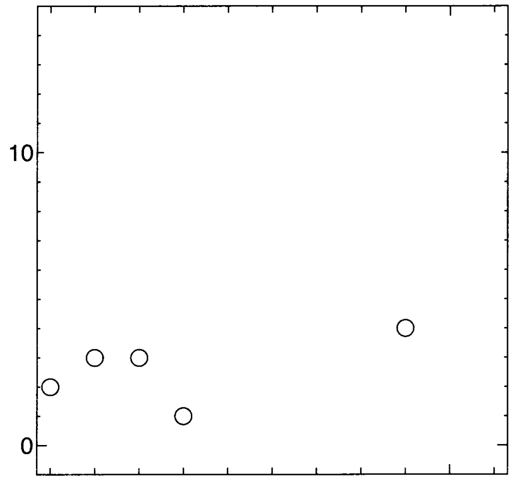

図 5　問題 PL(中央値より小さいデータを評価)

引っ張る程度 z^L は，それぞれ，中央値からどれくらい離れているのかの平均値となり，総合評価値 z は，平均値に等しくなる．また，$\xi = 1$ ならば，上に引っ張る程度 z^U，下に引っ張る程度 z^L は，それぞれ，中央値からもっとも離れている値となり，総合評価値 z は，$z =$ 中央値 $+$ (最大値 $-$ 中央値) $+$ (最小値 $-$ 中央値) $=$ 最大値 $+$ 最小値 $-$ 中央値 となる．

3.2　1 回のショケ積分に統合

次に，2 つの問題に分け，2 回のショケ積分を行っていたものを，1 つの問題 (1 回のショケ積分) の問題に変換する．この場合，2 つ合わせた場合のファジィ測度 μ^* をどのように割り当てるのかという問題になる．n が，奇数の場合と偶数の場合で異なるので，それぞれ示す．

n が奇数の場合，$\forall A \subset X$ について，

$$\mu^*(A) = \begin{cases} \phi_s(\dfrac{\mid A \mid}{n}) & \text{if} \mid A \mid < \dfrac{n}{2} \\ 1 - \phi_s(\dfrac{n - \mid A \mid}{n}) & \text{otherwise} \end{cases} \tag{18}$$

とする．

n が偶数の場合，全体集合は，X に中央値を加えた X^E を利用する．すなわち，$\forall \subset X^E = X \cup \{x_{Med}\}$ とし，$h(x_{mid}) =$ 中央値 とする．

$$\mu^*(A) = \begin{cases} \phi_s(\dfrac{|A|}{n}) & \text{if } |A| < \dfrac{n+1}{2} \\ 1 - \phi_s(\dfrac{(n+1) - |A|}{n}) & \text{otherwise} \end{cases} \tag{19}$$

$$h : X^E \rightarrow [0, \infty) \tag{20}$$

偶数，奇数の場合とも，

$$z = (C)\int h d\mu^* \tag{21}$$

とする，奇数の場合，明らかに中央値，平均値の拡張になっている．すなわち，$\xi = 0$ の場合，

$$\mu^*(A) = \begin{cases} 0 & \text{if } |A| < \dfrac{n}{2} \\ 1 & \text{otherwise} \end{cases} \tag{22}$$

となり，中央値となる．また，$\xi = 0.5$ の場合，

$$\mu^*(A) = \begin{cases} \dfrac{|A|}{n} & \text{if } |A| < \dfrac{n}{2} \\ \dfrac{|A|}{n} & \text{otherwise} \end{cases} \tag{23}$$

となり，平均値になる．

偶数の場合，$\xi = 0$ とすると，

$$\mu^*(A) = \begin{cases} 0 & \text{if } |A| < \dfrac{n+1}{2} \\ 1 & \text{otherwise} \end{cases} \tag{24}$$

となる．この場合，要素の数が $n+1$ 個から $(n/2)+1$ 個までのファジィ測度の値が 1 になる．中央値は，大きい方から数えて $(n+1)/2+0.5$ 番目のデータであるので，必ず，中

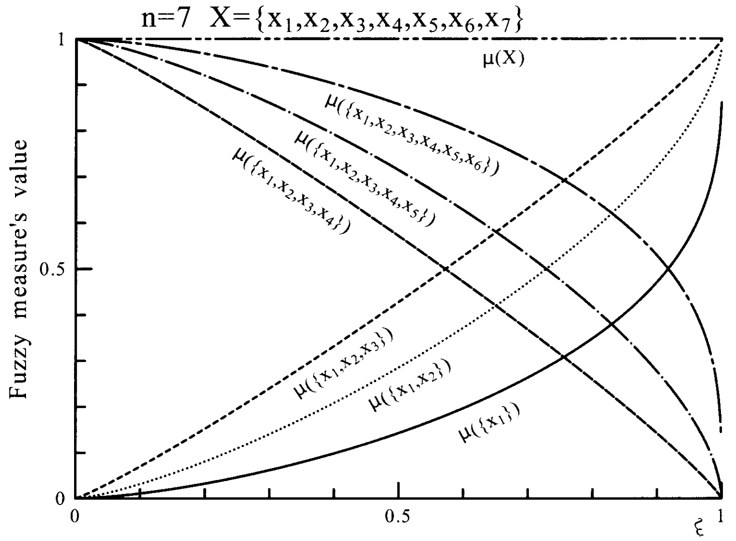

図 6 奇数の場合のファジィ測度の変化

央値のファジィ測度の値から 1 になる．したがって，中央値となる．$\xi = 0.5$ とすると，

$$\mu^*(A) = \begin{cases} \dfrac{|A|}{n} & \text{if } |A| < \dfrac{n+1}{2} \\ \dfrac{|A|-1}{n} & \text{if } |A| = \dfrac{n+1}{2} + 0.5 \\ \dfrac{|A|-1}{n} & \text{if } |A| > \dfrac{n+1}{2} + 0.5 \end{cases} \tag{25}$$

となる．この場合，大きい方から数えて $(n+1)/2+0.5$ 番目（中央値）に対応するファジィ測度と $(n+1)/2+0.5+1$ 番目に対応するファジィ測度は一致する．したがって，中央値は評価されず，平均値となる．図 7 を見てもわかるように，$\xi > 0.5$ で，ファジィ測度が単調性を満たさない部分が出てくる．すなわち，非単調なファジィ測度になっている．

4 他の抵抗性のある代表値との比較

また，[1] では，抵抗性が高い平均値の拡張としていくつか紹介されている．重みづけ平均値は，n 個のデータ $(x_i, i = 1, \ldots, n)$ に，重み $w_i, i = 1, \ldots, n$ をつけて求める．す

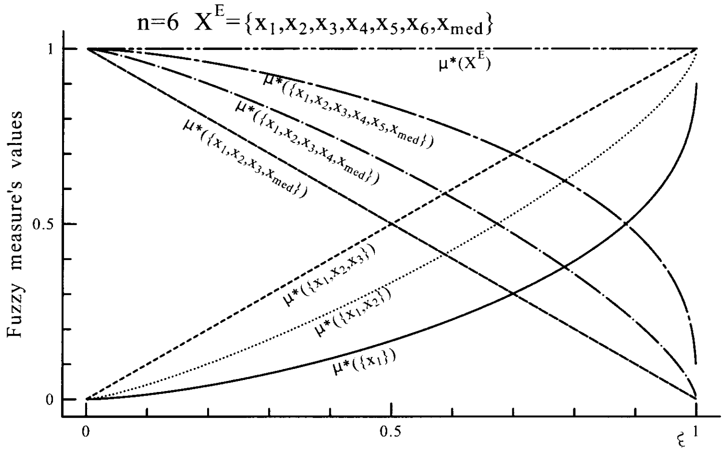

図7 偶数の場合のファジィ測度の変化

なわち，

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \quad w_i = \frac{1}{|\ x_i - \mathrm{Median}\ |} \tag{26}$$
$$\text{where} \quad w_i = 1 \text{ if } x_i = \mathrm{Median}$$

とする．これは，中央値から離れれば，離れるほど，ウエイトが小さくなるようにしている．

また，体操やアイススケートの審査などで用いられている最低点や最高点を除いて，平均を求める方法などがある．最大値と最小値からそれぞれ $\alpha\%$ のサンプルを取り除いたデータの平均値は，$\alpha\%$ 調整平均値 ($\alpha\%$ trimmed mean) と呼ばれている．50% 調整平均値は，中央値に等しくなる．

$\alpha\%$ 調整平均値は，それぞれ $\alpha\%$ のデータを取り除いたが，取り除かずに，残りのデータの最小値と最大値の等しい値に置き換えたものをウインザライズ平均値 (Winsorized mean) と呼ばれている．

三項平均値 (trimean) は，中央値 M と上ヒンジ H_U(最大値から中央値のデータの中央

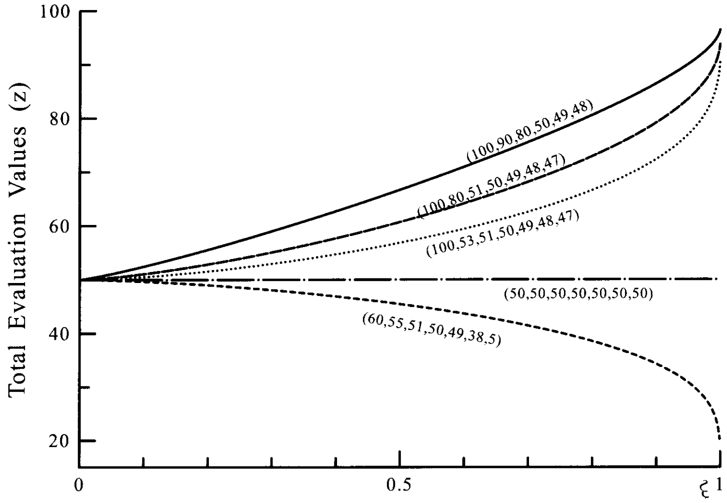

図 8　$n = 7$ の場合の総合評価値の変化 () 内は，評価値の組

値）と下ヒンジ H_L(最小値から中央値のデータの中央値）の重みつき平均値

$$\bar{x}_i = \frac{H_L + 2\mathrm{Median} + H_U}{3} \tag{27}$$

である．

次に，これらの値の特徴を見てみよう．

重みづき平均値では，重みの決め方によって評価が異なる．上の例では，データと中央値の差の絶対値の逆数を取っている．しかし，データと中央値の差が 1 未満のとき，重みが 1 より大きくなるという問題が起こる．他にも，中央値からの近さの順位を無視しているという問題点をもつ．

α% 調整平均値とウインザライズ平均値は，α を調整すれば，平均値と中央値の間の評価が可能である．しかしながら，調整時にデータが切り捨てられたり，残されたデータの最大値や最小値に一致させたりしている．したがって，中央値と同じように，切り捨てたデータの影響がその平均値に影響を与えないという欠点を持つ．

三項平均値も中央値と同様に 3 つの値以外の値がその平均値に与える影響は少ない．本手法は，重み付き平均値に近い．なぜなら，中央値との近さの順位によって，ファジィ測

度が変化するからである．言いければ，重みつき平均法で，重みを中央値との近さの順位によって決定する方法である．ただし，その重みがかかる部分は，自分より1つ中央値に近いデータまでである．ウインザライズ平均値は，ショケ積分の形で表現すれば，次のようななる．本手法のPUとPLのそれぞれの問題のファジィ測度を次のように割り当てる．

$$\mu(\mid A \mid) = \begin{cases} 0 & \text{if } \dfrac{\mid A \mid}{n} \leq \dfrac{\alpha}{100} \\ \dfrac{\mid A \mid}{n} & \text{otherwise} \end{cases} \tag{28}$$

これは，ファジィ測度をある段階から0を割り当てるという方法とも解釈できる．

5 終わりに

最大値，平均値，最小値の中間の評価を利用して，中央値，平均値の中間の評価法を提案した．1つのファジィ測度で表現できるので，その性質を分析することが可能になるだろう．今後，入力データが離散値の場合も検討していきたい．

参考文献

[1] 渡部洋，鈴木規夫，山田文康，大塚雄作「探索的データ解析入門」，朝倉書店,1985.

[2] 高萩栄一郎,"重要度と λ による λ ファジィ測度の同定について", 日本ファジィ学会誌, 12, 5, 665-676,2000, DOI: 10.3156/jfuzzy.12.5_73

[3] 塚本弥八郎，不確実性を伴う意思決定問題のファジィ測度論による分析，「ファジィ理論と人文社会科学」(講座ファジィ 14), 日刊工業新聞社,1994.