数理解析研究所講究録 第2095巻 2018年 131-137

単調測度を利用した多変量解析

本田あおい (九州工業大学情報工学部)*1
岡崎 悦明 (ファジィシステム研究所)*2

1. はじめに

非加法的な測度による積分は, Choquet 積分, 菅野積分, concave 積分をはじめとして様々な積分が提案されている. 本報告では我々の提案する包除積分の多変量データ解析への応用について考察する. 本報告の構成は, 2節で包除積分の定義をはじめとする数学的準備を行い, 3節で多変量データ解析問題への応用を考察する. 4章では他の代表的な非加法的測度による積分についてデータ解析への応用を中心に考察する. 応用の場合, 非加法的測度空間はもっぱら有限集合となる. 包除積分は非離散的な場合にも拡張可能であるが, 本稿では有限の場合のみ扱うことにする.

2. 包除積分の定義

本稿を通して Ω は有限集合とし, Ω のべき集合を $\mathcal{P}(\Omega)$ とかく. 集合関数 $\mu : \mathcal{P}(\Omega) \to [0, +\infty]$ が非加法的測度であるとは次の (MM1), (MM2) を満たすことである [1, 2].

(MM1) $\mu(\emptyset) = 0, \mu(X) < +\infty$.
(MM2) $A \subset B, A, B \in \mathcal{P}(\Omega)$ ならば $\mu(A) \leq \mu(B)$.

Ω 上の非加法的測度全体を $\mathcal{M}(\Omega)$ で表し, $[0, K]$ に値をとる Ω 上の関数全体を $\mathcal{F}_{[0,K]}(\Omega)$ とする. つまりこの場合 $\mathcal{F}_{[0,K]}(\Omega) = [0, K]^n$ である.

我々は非加法的測度による積分として次の包除積分を提案した.

定義 1 (非負有限包除積分 [4])
I を $([0, K], \mathcal{P}(\Omega))$ 上の interaction operator とする. 任意の $f \in \mathcal{F}_{[0,K]}(\Omega)$ に対して f の μ と I に関する包除積分 $(I) \int_\Omega f d\mu$ は次で定義される:

$$(I) \int_\Omega f \ d\mu \ := \sum_{A \in \mathcal{P}(\Omega)} M^I(f \mid A) \ \mu(A) := \sum_{A \in \mathcal{P}(\Omega)} \left(\sum_{B \supset A} (-1)^{|B \setminus A|} I(f \mid B) \right) \mu(A).$$

ここで interaction operator とは次で定義される多項演算である.

定義 2 (interaction operator) $I(\boldsymbol{x} \mid A) : [0, K]^n \times \mathcal{P}(\Omega) \to [0, K], K \in (0, +\infty]$ が次の (I1)–(I4) を満たすとき $([0, K], \mathcal{P}(\Omega))$ 上の interaction operator とよぶ:

(I1) 任意の $\boldsymbol{x}$ に対して, $I(\boldsymbol{x} \mid \emptyset) = K$.
(I2) 任意の $\boldsymbol{x}$ と任意の $i \in \Omega$ に対して, $I(\boldsymbol{x} \mid \{i\}) = x_i$.
(I3) 任意の $A \in \mathcal{P}(\Omega)$ と $\boldsymbol{x}, \boldsymbol{y} \in [0, K]^n$ に対して, $\boldsymbol{x} < \boldsymbol{y}$ ならば $I(\boldsymbol{x} \mid A) \leq I(\boldsymbol{y} \mid A)$.
(I4) 任意の $\boldsymbol{x} \in [0, K]^n$ と $A, B \in \mathcal{P}(\Omega)$ に対して, $A \subset B$ ならば $I(\boldsymbol{x} \mid A) \geq I(\boldsymbol{x} \mid B)$.

$(\Omega, \mathcal{P}(\Omega), \mu, I, K)$ を finite interactive monotone measure space とよぶことにする.

本研究は科研費 (課題番号:15K05003) の助成を受けたものである。
2010 Mathematics Subject Classification: 28B15, 05A19
キーワード：monotone measure, inclusion-exclusion integral, Möbius transformation, interaction operator
*1 e-mail: aoi@ces.kyutech.ac.jp
*2 e-mail: okazaki@flsi.or.jp

interaction operator の具体例として t-norm に基づく interaction operator を導入する. まず, t-norm の定義は次の通りである.

定義 3 2項演算 $\otimes : [0,K]^2 \to [0,K]$ が次の (GT1)–(GT4) を満たすとき, $[0,K]$ 上の t-ノルムと呼ぶ. 任意の $x,y,z \in [0,K]$ に対して,

(GT1) $0 \otimes 0 = 0, x \otimes K = x$ for $x \in [0,K]$.
(GT2) $x \le y$ ならば $x \otimes z \le y \otimes z$.
(GT3) $x \otimes y = y \otimes x$.
(GT4) $(x \otimes y) \otimes z = x \otimes (y \otimes z)$.

(GT4) より, t-norm は次のように多項演算に拡張できる:

$$\bigotimes_{i\in A} x_i := \begin{cases} K, & A=\emptyset, \\ x_j, & A=\{j\}, \\ \bigotimes_{i\in A} x_i, & otherwise. \end{cases}$$

一般に t-norm は $[0,1]$ 上の二項演算として定義される. $[0,1]$ 上の t-norm $\otimes$ は

$$x \otimes_G y := K\left(\frac{x}{K} \otimes \frac{y}{K}\right), x,y \in [0,K]$$

とすることで簡単に $[0,K]$ 上の t-norm に拡張できる. そして $\boldsymbol{X} = \{x_1,\ldots,x_n\}, A \in \mathcal{P}(\Omega)$ に対して

$$I(\boldsymbol{x} \mid A) := \bigotimes_{i\in A} x_i$$

とすると t-norm $\otimes$ に基づく interaction operator となり, これを用いることで t-norm に基づく包除積分が定義できる.

3. 包除積分の多変量データ解析への応用

多変量データ解析問題は, 目的変数 y を説明変数 $(x_1,\ldots,x_n)$ から推定するものである. 一般には $(y,x_1,\ldots,x_n)$ のデータの組がいくつかわかっている. 既知のデータをもとに精度のよいモデル式 $\hat{y} = F(x_1,\ldots,x_n)$ を構築する. t-ノルム $\otimes$ に基づく包除積分モデルは次のように表される.

$$\hat{y} = F(x_1,\ldots,x_n) = \sum_{A\in\mathcal{P}(\Omega)\setminus\{\emptyset\}} a_A \left(\sum_{B\supset A} (-1)^{|B\setminus A|} \bigotimes_{i\in B} x_i\right) + a_\emptyset.$$

$n=3$, つまり $\Omega = \{1,2,3\}$ の場合を書き下すと,

$$\begin{aligned} F(x_1,x_2,x_3) = {} & a_{\{1\}}(x_1 - x_1\otimes x_2 - x_1\otimes x_3 + x_1\otimes x_2\otimes x_3) \\ & + a_{\{2\}}(x_2 - x_1\otimes x_2 - x_2\otimes x_3 + x_1\otimes x_2\otimes x_3) \\ & + a_{\{3\}}(x_3 - x_1\otimes x_3 - x_2\otimes x_3 + x_1\otimes x_2\otimes x_3) \\ & + a_{\{1,2\}}(x_1\otimes x_2 - x_1\otimes x_2\otimes x_3) + a_{\{1,3\}}(x_1\otimes x_3 - x_1\otimes x_2\otimes x_3) \\ & + a_{\{2,3\}}(x_2\otimes x_3 - x_1\otimes x_2\otimes x_3) + a_{\{1,2,3\}}(x_1\otimes x_2\otimes x_3) + a_\emptyset. \end{aligned}$$

である. 既知のデータを教師データとし当てはまりのよい回帰係数 $a_A, A \in \mathcal{P}(\Omega)$ を決定する. この a_A が非加法的測度 $\mu(A)$ に相当する. あるいは包除積分の別表現を用い

てモデル式としてもよい:

$$\hat{y} = F(x_1, \dots, x_n) = \sum_{A \in \mathcal{P}(\Omega) \setminus \{\emptyset\}} a_A \left(\bigotimes_{i \in A} x_i \right).$$

$n = 3$ の場合を書き下すと,

$$\begin{aligned} F(x_1, x_2, x_3) &= a_{\{1\}} x_1 + a_{\{2\}} x_2 + a_{\{3\}} x_3 + a_{\{1,2\}} (x_1 \otimes x_2) + a_{\{1,3\}} (x_1 \otimes x_3) \\ &\quad + a_{\{2,3\}} (x_2 \otimes x_3) + a_{\{1,2,3\}} (x_1 \otimes x_2 \otimes x_3) + a_{\emptyset}. \end{aligned}$$

この場合の回帰係数 a_A は非加法的測度のメビウス変換 $m^{\mu}(A) := \sum_{B \supset A} (-1)^{|B \setminus A|} \mu(B)$ に相当する. パラメータの決定は通常の重回帰モデル問題に帰着でき, 推定値とデータ値の誤差の二乗和(残差平方和)を最小にするパラメータは, 教師データの個数を M セットとすると

$$R(\{a_A\}_{A \in \mathcal{P}(\Omega)}) = \sum_{j=1}^{M} \left(\hat{y^j} - F(x_1^j, \dots, x_n^j) \right)^2$$

を最小とする $a_A, A \in \mathcal{P}(\Omega)$ を決定すればよい. これは各 a_A についての2次関数であるから

$$\frac{\partial R(\{a_A\}_{A \in \mathcal{P}(\Omega)})}{\partial a} = \mathbf{0}$$

を与える $a_A, A \in \mathcal{P}(\Omega)$ を求めればよい. 実は計算しなくてもデータ解析ソフトの重回帰分析ツールを用いれば簡単に求めることができる. 例えば Microsoft 社の Excel であれば $(y, x_1, \dots, x_n)$ の M セットのデータを用いて, 包除積分モデルの場合は

$$\left(y,\ \sum_{B \supset \{1\}} (-1)^{|B|-1} \bigotimes_{i \in B} x_i,\ \sum_{B \supset \{2\}} (-1)^{|B|-1} \bigotimes_{i \in B} x_i,\ \dots,\ \bigotimes_{i \in \Omega} x_i \right)$$

なるデータの組を M セット, 包除積分別表現モデルの場合は

$$\left(y,\ x_1,\ \dots,\ x_n,\ x_1 \otimes x_2,\ \dots,\ x_{n-1} \otimes x_n,\ x_1 \otimes x_2 \otimes x_3,\ \dots,\ \bigotimes_{i \in \Omega} x_i \right)$$

なるデータの組を M セット作成する. M 行 $N+1$ 列のデータが作成できたらデータメニューから分析ツールの重回帰分析を選び, 目的変数として y のカラムを, 説明変数として残りのカラムを選択するだけで自動的に回帰係数をはじめ必要な統計量が自動計算される. ただし Excel の場合は説明変数は 16 個までしか選べないことに注意する.

以下に, 包除積分モデルを用いる場合の注意点を述べる.

包除積分モデルと包除積分別表現モデルの違い

別表現モデルの方が若干使いやすいモデルである. 包除積分モデルはデータを準備する際に, 多くの積和(差)を用いた計算が必要であるため n が大きい場合, 丸め誤差が蓄積する可能性がある. モデル式が重回帰モデルの拡張になっている. t-ノルムを掛け算としたものが, 非線形重回帰モデルと一致している. 重回帰モデルの場合にはこの項は「交互作用項」と呼ばれ, データ間のなんらかの相互作用を表しているとされている. し

かし特に, 最も有利な点は非加法的測度を k-加法的測度と仮定したときにモデル式の項数を簡単に減らすことができる点にある. k-加法的測度とは $|A| > k$ のとき $m^{\mu}(A) = 0$ となる測度である. つまり $|A| > k$ なる　a_A の値は 0 となる. これにより重回帰モデルに比べ爆発的に増えてしまう項数を適宜調整することが可能である. 上記でも触れたように Excel では説明変数の説明変数は16個までしか選択できない. 包除積分モデルの場合, 説明変数の数が n 個の場合, 項数は $2^n - 1$ 個となり, これが重回帰モデルにおける説明変数の数となる. これでは $n = 4$ の場合までしか扱えない. そこで例えばモデルの非加法的測度を 2-加法的測度と制限することで, 説明変数が 5 個の場合まで分析可能となる. 実際, $|\{\{1\}, \{2\}, \ldots, \{5\}, \{1, 2\}, \ldots, \{4, 5\}\}| = \begin{pmatrix} 5 \\ 1 \end{pmatrix} + \begin{pmatrix} 5 \\ 2 \end{pmatrix} = 15$ である. とはいえ $n = 6$ の場合は 2-加法的と仮定しても項数は31であるため Excel では計算できない. 統計ソフト R や SPSS などは, もっと多くの説明変数を扱うことができる. 因みに他に項数を減らす手法としては, 重回帰分析の変数選択法を採用することもできる.

モデル式の解釈

包除積分モデルを用いたデータ分析が機械学習や他の数理モデルと比較して優れているところは, 得られたモデル式の意味を解釈できる点にある. 回帰係数 a_A が各説明変数や説明変数の合成の重視度を表し, 相互作用の働く範囲を t ノルムで説明できる. しかしながらそのためには事前にデータのスケールや大小関係を調べてデータに適切な前処理を行う必要がある.　説明変数が非積分関数となり目的変数が積分値であるので, 　説明変数の値は大きいほど目的変数が大きくなるよう, 正負の向きを揃える必要がある. また t-ノルムで演算を行うためデータは $[0, K]$ の範囲におさまるように正規化する必要がある. 煩雑な前処理が必要であるが, モデル式を解釈できるメリットは大きい. しかしながら何らかの理由で前処理ができない場合も, 包除積分モデルとしてパラメータを決定することは可能である. この場合はモデル式の意味を解釈することはできないが, この場合にもなんらかのよい推定値を求めるという目的はひとまず達成することはできる. そもそも多くのデータ解析手法においては得られたモデル式はブラックボックスになっている.

測度の単調性をどう考えるか

包除積分は非加法的測度による積分であるので, 包除積分モデルに現れる測度, つまりパラメータは非加法的測度, つまり単調な測度を表現したもの仮定するのが自然であろう. 単に最小二乗法で回帰係数を決定する場合は, 回帰係数が単調測度になっているとは限らない. 単調測度の仮定を入れる場合は単調測度となるパラメータの範囲の中から最も当てはまりのよいものを選ぶ何らかのアルゴリズムが必要である. この場合も単によい推定値を求めるのが目的である場合や, 扱う問題が明らかに測度が単調性を満たさないようなものであれば単調性を考慮する必要はないであろう.

4. 代表的なファジィ積分と多変量データ解析

非加法的な測度による積分は様々なものが提案されている. 最もポピュラーな Choquet 積分である. 他に菅野により提案された菅野積分, 近年では Lehrer による concave 積分が扱いやすさや性質のよさから注目されている. 本節ではこれらの積分について解説する. $\mathcal{F}_{[0,\infty)}(\Omega)$ を Ω 上の非負関数全体とする. 被積分関数 $\mathcal{F}_{[0,\infty)}(\Omega)$ とすると $f = (x_1, x_2, \ldots, x_n) \in [0, \infty)^n$ と表せる.

定義 4 $f \in \mathcal{F}_{[0,\infty)}(\Omega)$ の非加法的測度 $\mu \in \mathcal{M}(\Omega, \mathcal{P}(\Omega))$ による Choquet 積分, 菅野積分, concave 積分の定義は次で定義される.

$$
\begin{aligned}
(C)\int f d\mu &:= \sum_{i=1}^{n} \left(f^*(i) - f^*(i+1)\right) \mu(A_i) \\
(S)\int f d\mu &:= \bigvee_{i=1,2,\ldots,n} \{f^*(i) \wedge \mu(A_i)\} \\
\int^{cav} f d\mu &:= \max\left\{ \sum_{A \in \mathcal{P}(\Omega)} a_A \mu(A) \;\middle|\; \{a_A\}_{A \in \mathcal{P}(\Omega)};\ a_A \geq 0,\ \sum_{A \in \mathcal{P}(\Omega)} a_A \chi_A = f \right\}
\end{aligned}
$$

ただし, f^* は $f^*(1) \geq f^*(2) \geq \cdots \geq f^*(n)$ なる f の rearrangement, $A_i := \{1, 2, \ldots, i\}$, $\vee, \wedge$ は max 演算と min 演算である.

Choquet 積分, 菅野積分, concave 積分は積分型汎函数である. つまり, これらの積分を $F : \mathcal{M}(\Omega) \times \mathcal{F}_{[0,\infty)}(\Omega) \to [0, \infty]$ で表すことにすると

1. 任意の $\mu \in \mathcal{M}(\Omega)$ に対して $F(\mu, 0) = 0$.
2. 任意の $\mu \in \mathcal{M}(\Omega)$ と任意の $f, g \in \mathcal{F}_+(\Omega)$ に対して, $f \leq g$ ならば $F(\mu, f) \leq F(\mu, g)$.

の 2 条件を満たす.

4.1. Choquet 積分

命題 5 包除積分の Interaction operator を $I(f \mid A) := \bigwedge_{i \in A} f(i)$ とすると, Choquet 積分と一致する.

注意 6 包除積分を連続な場合に拡張したものについても, 同じ命題が無条件に成り立つ.

この定義からわかるように, Choquet 積分は包除積分の特別な場合である. min 演算は Interaction operator のうちの最大の演算である. 言い換えると, Choquet 積分は包除積分の計算において interaction の働く部分を最大限に見積もったものであるとみなすことができる.

命題 7 任意の $f \in \mathcal{F}_{[0,\infty)}(\Omega), \mu \in \mathcal{M}(\Omega, \mathcal{P}(\Omega)), a, b \in [0, +\infty)$ に対して

$$
(C)\int (af + b) d\mu \;=\; a\ (C)\int f d\mu + b.
$$

この性質より, Choquet 積分は特に被積分関数の正規化を行わなくてもよいことがわかる. ただし, 全ての $i \in \Omega$ について $f(i)$ のスケールが揃っている必要はある. ちなみに, 命題 7 は Choquet 積分と包除積分の連続版においても無条件で成り立つ.

多変量データ解析において, Choquet 積分モデルを用いる場合, 包除積分モデルの場合と同様にモデルのパラメータ, 積分の測度に相当, を同定することができる.

4.2. 菅野積分

定義式からわかるように, 積分値の計算において被積分関数値と非加法的測度値の比較を行う. このため, この比較が意味を持つ必要がある. 例えば科学者の貢献度を表すとされる h-index は対象となる科学者の全ての論文の被引用件数を非積分関数とし, counting measure で菅野積分したものに相当する [5]. この場合は被引用件数と論文数

を比較するわけでどちらも論文の本数を表している．このような菅野積分で表せる評価や積算値の例は他にも存在しそうである．

3節のような多変量データ解析において菅野積分モデルを用いる場合，パラメータを同定するのは難しく，今のところモデルを同定する手法は見つかっていない．

4.3. concave 積分

concave 積分は，非積分関数を分割して積分値を算出する，いわゆる「分割型」積分の1種である点で包除積分と関係の深い積分である．全ての分割から計算される積算値のうち最大のものを積分値と見なす．concave 積分は次のような性質を持つ．

命題 8 任意の $f \in \mathcal{F}_+(\Omega), \mu \in \mathcal{M}(\Omega, \mathcal{P}(\Omega)), a, b \in [0, +\infty)$ に対して μ が劣加法的ならば

$$\int^{cav} f d\mu = \int f d\mu.$$

このことからわかるように，concave 積分は μ の劣加法性を表現することができない．

concave 積分は無数に存在する被積分関数の分割の中から最大の積算値を与える分割を見つけなければならないので，そもそも concave 積分の計算自体が難しそうに見えるが，この計算は可能である．concave 積分の積分値を与える分割は

$$\sum_{A \in i} a_A \leq f(i), i \in \Omega$$

の条件の下で，

$$\sum_{A \in \mathcal{P}(\Omega)} a_A \ \mu(A)$$

を最大化するものである．目的関数と制約条件がすべて線形であるので線形計画法に帰着できる．この条件を満たす分割は無限個存在するが，条件を満たす領域は超多面体になっており解は超多面体の頂点であることが知られている．この頂点を効率的にチェックする手法は既に確立されており，多項式時間で解くアルゴリズムが提案されている．ほとんどのプログラム言語において線形計画法のライブラリが準備されている．小規模な問題であれば Excel のソルバーで簡単に解くことができる．このように測度が与えられている場合の積分の計算は可能であるが，一方3節で述べたような多変量データ解析においては，モデル式に concave 積分を用いた場合，モデルのパラメータを同定するのは難しい．concave 積分と Choquet 積分の関係においては次の命題が成り立つ．

命題 9 次の2つは同値である:

1. concave 積分とショケ積分が一致する．
2. μ が super additive である．すなわち任意の $A, B \in \mathcal{P}(\Omega)$ に対して

$$\mu(A \cup B) + \mu(A \cap B) \geq \mu(A) + \mu(B).$$

super modularity より弱い優加法性では不十分であり，μ が優加法的な場合は concave 積分とショケ積分は一般には一致しない．実際，$\Omega = \{1,2,3\}, \mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = 1, \mu(\{1,2\}) = \mu(\{2,3\}) = 3, \mu(\{1,2,3\}) = 4, f(1) = f(3) = 1, f(2) = 2.$ とすると f の μ による　Choque積分は 5, concave 積分は 6 となり一致しない．

命題 10 $\bar{\mu}(A) := (L) \int \chi_A \ d\mu$ とすると

1. $\bar{\mu}$ は非加法的測度である．

2. $\bar{\mu}(A) \geq \mu(A),\ \forall A \in \mathcal{P}(\Omega)$.
3. $\bar{\bar{\mu}} = \bar{\mu}$.
4. $(L)\displaystyle\int f d\mu = (L)\int f d\bar{\mu}$.
5. $\bar{\mu}$ は優加法的である.

ただし χ_A は定義関数.

これより, $\bar{\mu}$ が super modular である場合, ショケ積分を用いて $\bar{\mu}$ を決定することができる. ただし $\bar{\mu}$ から μ を決定することはできない. $\bar{\mu}$ が supermodular となる条件も今のところわかっていない.

参考文献

[1] M. Sugeno, Fuzzy measures and fuzzy integrals–a survey, In: M. M. Gupta, G. N. Saridis and B. R. Gaines (eds), *Fuzzy automata and decision processes*, pp. 89–102, North Holland, Amsterdam, 1977.

[2] T. Murofushi and M. Sugeno, Fuzzy measures and fuzzy integrals, in: *Fuzzy Measures and Integrals: Theory and Applications*, M. Grabisch, T. Murofushi, and M. Sugeno, eds., pp. 3–41, Physica-Verlag, 2000.

[3] G. Choquet, Theory of capacities, *Ann. Inst. Fourier*, **5**, pp. 131–295, 1953.

[4] A. Honda, Y. Okazaki, Theory of inclusion-exclusion integral, *Information Science*, **376**, pp. 136-147, 2017.

[5] 本田あおい, 岡崎悦明, 包除積分の非離散化, 実解析学シンポジウム 2017 報告集, pp.59-64.