区間推定と可能性計画法を用いた 最適コストの評価手法*

山川 栄樹[†]・山下 恒平[‡]

An Application of Interval Estimation and Possibilistic Programming to Estimate the Operation Cost*

Eiki YAMAKAWA[†] and Kohei YAMASHITA[‡]

1. はじめに

不確実な状況下で意思決定を行う方法に、確率計画法 [7] やファジィ数理計画法 [1] がある。ファジィ数理計画法は、メンバシップ関数により帰属度を連続的に定められるファジィ集合 [2] を用いて、あいまいな状況を定式化する。とくに、係数にファジィ数を含む最適化問題を機会制約の考え方をを用いて確定的問題に直して解く方法は、可能性計画法 [4,9] とよばれている。

三角ファジィ数を係数とする回帰式を求める方法に、ファジィ回帰分析 [11] や区間回帰分析 [10] がある。これらの方法では、全観測データが回帰式の定める区間に含まれるという条件のもとで、あいまいさが最小となるようにファジィ数の中心値と幅を定める。そのため、幅の決定に有効な一部のデータしか利用されず、異常データがあると幅が広がりすぎてしまう。そこで、あいまいさの最小化と回帰式の定める区間に含まれるデータ数の最大化を図る多目的最適化問題を解く方法 [5] や、最小二乗法により中心値を定める方法 [8] が提案されている。

本速報では、最適化問題がモデル誤差を伴う回帰式を含むとき、予測値に対する区間推定 [6] を用いて、従属変数を独立変数の可能性関数として表現する。文献 [12] では、区間推定を用いて定式化した 2 レベル数理計画問題を解いて、最悪の状況における最適コストを評価している。本速報では、可能性測度と必然性測度 [3] を用いて、ある水準以上の可能性または確実性で実現できる最適コストを評価する。また、数値例を用いて、提案した方法の有効性と異常データに対する頑健性を示す。

2. 最適コストの評価モデル

対称三角ファジィ数 $\tilde{a} = \langle a^c, a^s \rangle$ のメンバシップ関数は、

$$\mu_{\tilde{a}}(z) = \max \left\{ 0, \frac{|z - a^c|}{a^s} \right\} \quad (1)$$

と定義される。ファジィ数 $\tilde{a}$ が集合 S に含まれる可能性の度合いを可能性測度、確実性の度合いを必然性測度といい、それぞれ $\Pi(\tilde{a} \in S)$ および $N(\tilde{a} \in S)$ と書く [3]。(1) 式より、 $0 \leq h \leq 1$ に対して次式が成り立つ。

$$\begin{aligned} \Pi(\tilde{a} \leq u) &\geq h \Leftrightarrow a^c - (1-h)a^s \leq u \\ \Pi(\tilde{a} \geq l) &\geq h \Leftrightarrow a^c + (1-h)a^s \geq l \\ N(\tilde{a} \leq u) &\geq h \Leftrightarrow a^c + ha^s \leq u \\ N(\tilde{a} \geq l) &\geq h \Leftrightarrow a^c - ha^s \geq l \end{aligned} \quad (2)$$

以下では、変数 $x = (x_1, \dots, x_n)^T$ を与えたとき、その値のとり得る範囲が対称三角ファジィ数 $\langle y_i^c(x), y_i^s(x) \rangle$ に制限される関数を可能性関数とよび、 $\tilde{y}_i(x)$ と書く。

本速報では、 y が x の可能性関数 $\tilde{y}(x) = (\tilde{y}_1(x), \dots, \tilde{y}_m(x))^T$ となるつぎのような数理計画問題を考える。

$$\begin{aligned} \text{目的関数: } &c^T x \rightarrow \text{最小} \\ \text{制約条件: } &p_i^T x + q_i^T y = r_i \quad (i \in \mathcal{I}_E) \\ &p_i^T x + q_i^T y \geq r_i \quad (i \in \mathcal{I}_I) \\ &y = \tilde{y}(x) \end{aligned} \quad (3)$$

ただし、 $c, p_i \in \mathbb{R}^n, q_i \in \mathbb{R}^m, r_i \in \mathbb{R}$ は確定的パラメータである。(2) 式より、問題 (3) の制約条件が、式

$$\begin{aligned} p_i^T x + q_i^T \tilde{y}(x) &= r_i \quad (i \in \mathcal{I}_E) \\ p_i^T x + q_i^T \tilde{y}(x) &\geq r_i \quad (i \in \mathcal{I}_I) \end{aligned} \quad (4)$$

の成り立つ可能性を h 以上にすることを求めていると考え、問題 (3) はつぎの確定的問題に変換できる。

$$\begin{aligned} \text{目的関数: } &c^T x \rightarrow \text{最小} \\ \text{制約条件: } &p_i^T x + q_i^T y^c(x) \\ &\quad - (1-h)|q_i^T| y^s(x) \leq r_i \quad (i \in \mathcal{I}_E) \\ &p_i^T x + q_i^T y^c(x) \\ &\quad + (1-h)|q_i^T| y^s(x) \geq r_i \quad (i \in \mathcal{I}_E \cup \mathcal{I}_I) \end{aligned} \quad (5)$$

* 原稿受付 2008 年 12 月 8 日

[†] 関西大学 環境都市工学部 Faculty of Environmental and Urban Engineering, Kansai University; 3-3-35 Yamatecho, Suita city, Osaka 564-8680, JAPAN

[‡] 関西大学大学院 工学研究科 Graduate School of Engineering, Kansai University; 3-3-35 Yamatecho, Suita city, Osaka 564-8680, JAPAN

Key Words: interval estimation, multiple regression analysis, fuzzy sets, possibilistic programming, nonlinear programming.

一方, $\mathcal{I}_E = \emptyset$ または $q_i = 0$ ($i \in \mathcal{I}_E$) ならば, 問題 (3) の制約条件が (4) 式の成立する現実性を h 以上にすることを求めていると解釈すると, 問題 (3) を確定的問題

$$\begin{aligned} & \text{目的関数: } c^T x \rightarrow \text{最小} \\ & \text{制約条件: } p_i^T x = r_i \quad (i \in \mathcal{I}_E) \\ & \quad p_i^T x + q_i^T y^c(x) \\ & \quad - h |q_i^T| y^s(x) \geq r_i \quad (i \in \mathcal{I}_I) \end{aligned} \quad (6)$$

に変換できる. 問題 (5) の最適値は, 平均よりよい状況でも h の可能性で必要となるコストを表す. 一方, 問題 (6) の最適値は, 平均より悪い状況でも h の現実性で実現できるコストを表している. なお, 問題 (5) の $h=1$ における最適値と問題 (6) の $h=0$ における最適値は一致し, 平均的状況における最適コストを与える.

3. 可能性関数の定式化

従属変数 y が独立変数 x の 1 次関数になると想定されるとき, 問題 (3) に含まれる可能性関数 $\tilde{y}_i(x)$ は, 対称三角ファジィ数 $\tilde{a}_i = (\tilde{a}_{i1}, \dots, \tilde{a}_{in})^T$, $\tilde{b}_i$ を用いて

$$\tilde{y}_i(x) = \tilde{a}_i^T x + \tilde{b}_i \quad (i = 1, \dots, m)$$

と定式化できる. 観測データ $x^{(k)}$, $y_i^{(k)}$ ($k = 1, \dots, \ell$) が与えられたとき, $i = 1, \dots, m$ のそれぞれに対して問題

$$\begin{aligned} & \text{目的関数: } \sum_{k=1}^p \{(a_i^s)^T |x^{(k)}| + b_i^s\} \rightarrow \text{最小} \\ & \text{制約条件: } |(a_i^c)^T x^{(k)} + b_i^c - y_i^{(k)}| \\ & \quad \leq (a_i^s)^T |x^{(k)}| + b_i^s \quad (k = 1, \dots, p) \\ & \quad a_i^s \geq 0, \quad b_i^s \geq 0 \end{aligned} \quad (7)$$

を解いて $\tilde{a}_i = \langle a_i^c, a_i^s \rangle$, $\tilde{b}_i = \langle b_i^c, b_i^s \rangle$ を定める方法を, 区間回帰分析という [10]. 区間回帰分析では, 線形計画問題を解いて $\tilde{a}_i$, $\tilde{b}_i$ を定めるため, 係数に対する実務的知識を組み込みやすいが, 異常データに敏感で [5], 有効な制約に関与する一部のデータしか利用されない [8].

本速報では, 重回帰分析における区間推定を用いて, 可能性関数 $\tilde{y}_i(x) = \langle y_i^c(x), y_i^s(x) \rangle$ を定式化する. 重回帰分析では, 観測データ $x^{(k)}$, $y_i^{(k)}$ ($k = 1, \dots, \ell$) が式

$$y_i^{(k)} = a_i^T x^{(k)} + b_i + \varepsilon_i^{(k)} \quad (i = 1, \dots, m; k = 1, \dots, \ell)$$

を満たすと仮定する. ただし, $\varepsilon_i^{(k)}$ は, 平均 0, 分散 σ_i^2 の正規分布に従う確率変数である. まず, a_i , b_i の推定値 $\hat{a}_i$, $\hat{b}_i$ を求め, $\tilde{y}_i(x)$ の中心を与える関数 $y_i^c(x)$ を

$$y_i^c(x) = \hat{a}_i^T x + \hat{b}_i$$

と定める. このとき, 誤差の分散 σ_i^2 の推定値 v_i は

$$v_i = \frac{1}{\ell - n - 1} \sum_{k=1}^{\ell} \left\{ y_i^{(k)} - y_i^c(x^{(k)}) \right\}^2$$

となる. そこで, $\tilde{y}_i(x)$ の幅 $y_i^s(x)$ を次式で定義する.

$$y_i^s(x) = t_{\tau/2} \sqrt{\left\{ 1 + \frac{1}{\ell} + \varphi(x) \right\} v_i}$$

ただし, $x_j^{(k)}$ ($k = 1, \dots, \ell$) の平均 $\bar{x}_j$ と, 偏差積和

$$s_{jj'} = \sum_{k=1}^{\ell} (x_j^{(k)} - \bar{x}_j)(x_{j'}^{(k)} - \bar{x}_{j'})$$

を (j, j') 成分とする行列の逆行列 $(\bar{s}_{jj'})_{n \times n}$ に対して

$$\varphi(x) = \sum_{j=1}^n \sum_{j'=1}^n (x_j - \bar{x}_j)(x_{j'} - \bar{x}_{j'}) \bar{s}_{jj'}$$

である. また, 自由度 $\ell - n - 1$ の t 分布の確率密度関数を ψ とすると, $t_{\tau/2}$ は次式を満たす t の値である.

$$\int_t^{\infty} \psi(\omega) d\omega = \frac{\tau}{2}$$

独立変数 x の値を一時的に固定すると, 可能性関数 $\tilde{y}_i(x)$ は対称三角ファジィ数 $\langle y_i^c(x), y_i^s(x) \rangle$ となるから, そのメンバシップ関数の値が正となるのは, 従属変数の値が区間 $(y_i^c(x) - y_i^s(x), y_i^c(x) + y_i^s(x))$ に含まれる場合である. ところが, この区間は, 従属変数の理論値 $y_i^c(x)$ に対する予測値 y_i の $100(1 - \tau) \%$ 信頼区間 [6] にほかならない. したがって, このようにして定式化した可能性関数 $\tilde{y}_i(x)$ は, 観測データの分布をよく反映しており, 異常データにも頑健であると期待される. なお, $y_i^s(x)$ は非線形関数であるため, 実際に解く問題 (5) または (6) はいずれも非線形計画問題となる.

4. 計算実験

提案した方法の有効性を確かめるために, 計算実験を行った. 最初に, つぎのようなテスト問題を考える.

$$\begin{aligned} & \text{目的関数: } x_1 + 4x_2 + 6x_3 + 3y_1 + 2y_2 \rightarrow \text{最小} \\ & \text{制約条件: } x_1 + 2x_2 + 2x_3 \geq 10 \\ & \quad 4x_1 + 3x_2 + 2x_3 + y_1 + 2y_2 \geq 41 \\ & \quad 2x_1 + 4x_2 + 12x_3 + 2y_1 + y_2 \geq 58 \\ & \quad y_1 = \tilde{y}_1(x_1, x_2), \quad y_2 = \tilde{y}_2(x_3) \\ & \quad x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0 \end{aligned} \quad (8)$$

問題 (8) は, 目的関数に y を含むが, 不等式制約条件

$$x_0 - x_1 - 4x_2 - 6x_3 - 3y_1 - 2y_2 \geq 0$$

を追加し, 最小化すべき目的関数を x_0 に変更することにより, (3) 式の形式をもつ問題に帰着できる.

計算実験では, $x_1^{(k)}$, $x_2^{(k)}$, $x_3^{(k)}$ をそれぞれ区間 $[1, 1.5]$, $[1.5, 3.5]$, $[1.5, 2.5]$ の一様乱数として生成したあと, 式

$$\begin{aligned} y_1^{(k)} &= 2x_1^{(k)} + x_2^{(k)} + 1 + \varepsilon_1^{(k)} \\ y_2^{(k)} &= 3x_3^{(k)} + 2 + 0.5\varepsilon_2^{(k)} \end{aligned} \quad (k = 1, \dots, 200)$$

により 200 件の擬似的な観測データ $(x^{(k)}, y^{(k)})$ を生成

Fig. 1 Formulation of possibilistic function $\tilde{y}_2(x_3)$

Fig. 2 Computational results for problem (8)

した．ただし， $\varepsilon_1^{(k)}$ ， $\varepsilon_2^{(k)}$ は標準正規乱数である．Fig. 1 の (a) と (b) に，それぞれ区間回帰分析と提案した方法で求めた $y_2^c(x_3)$ および $y_2^c(x_3) \pm y_2^s(x_3)$ のグラフを示す．ただし，破線はFig. 1中に示すような2件の異常データを加えた場合，実線は異常データを加えない場合である．なお，区間推定において， $\tau = 0.05$ とした．

つぎに， h の値を 0 から 1 まで 0.1 刻みで変化させ，問題 (6)，(7) を汎用ソフト NUOPT で解いた．最適コストの評価結果をFig. 2に示す．ただし，灰色の線は区間回帰分析で，黒線は提案した方法で $\tilde{y}(x)$ 定式化した場合であり，破線は異常データを含む場合，実線は異常データを含まない場合である．これらの結果より，区間回帰分析を用いると，少数の異常データのために $y^s(x)$ の値が過大になり， $y^c(x)$ の値も大きく変化して最適コストの評価を誤る可能性があるのに対して，提案した方法は異常データに対して頑健であることがわかる．

最後に，発電設備の運転最適化問題[12]に提案した手法を適用した．問題の規模は， x の次元が 22， y の次元が 7，回帰式以外の等式制約条件の数が 17，不等式制約の数が 8 である．実験には，実際の運転実績データ 1464 件と人為的に生成した 10 件の異常データを用いた．現実の最適化問題を解く場合には，変数や式のスケールリングが不可欠である．予測値の区間推定はスケール不変性をもつ[12]が，ファジィ回帰分析は一般にスケール不変ではない．また，この問題は y を含む等式制約条件をもつ．そこで，提案した方法で $\tilde{y}(x)$ を求め，問題 (5) を解いて最適コストを推定した結果だけをFig. 3に示す．

Fig. 3 Computational results for operation optimization problem of the electric power plant

Fig. 3 より，提案した方法では異常データによる評価値のズレが 1 割以内に収まっていることがわかる．

5. おわりに

本速報では，制約条件に含まれる線形回帰式を従属変数の区間推定を用いて可能性関数として定式化し，ある水準以上の可能性または確実性で実現可能な最適コストを評価する方法を提案した．また，提案した手法を発電設備の運転最適化問題に適用し，その有効性を確認した．回帰式の従属変数が，高い確実性で満たさなければならない等式制約条件に含まれている場合に最適コストを評価する方法の開発は，今後の課題である．

参考文献

- [1] 浅居，田中：ファジィOR，日刊工業新聞社 (1993)
- [2] D. Dubois and H. Prade: *Fuzzy Sets and Systems: Theory and Applications*, Academic Press (1980)
- [3] D. Dubois and H. Prade: Fuzzy numbers: an overview; *Analysis of Fuzzy Information, Volume I: Mathematics and Logic*, (Ed., J. C. Bezdek), CRC Press, pp. 3–39 (1987)
- [4] 乾口：可能性計画法；オペレーションズ・リサーチ，Vol. 41, pp. 569–574 (1996)
- [5] 石淵，田中：混合 0-1 整数計画問題による区間回帰分析；日本経営工学会誌，Vol. 40, pp. 312–319 (1989)
- [6] 石原，長谷川，川口：Lotus 1-2-3 活用多変量解析，共立出版 (1990)
- [7] I. M. Stancu-Minasian: *Stochastic Programming with Multiple Objective Functions*, D. Reidel Publishing Company (1984)
- [8] 田島，古川：非対称型ファジィ回帰モデル同定法の研究；日本計算工学会論文誌，Vol. 3, pp. 125–130 (2001)
- [9] H. Tanaka and K. Asai: Fuzzy linear programming with fuzzy numbers; *Fuzzy Sets and Systems*, Vol. 13, pp. 1–10 (1984)
- [10] 田中，林，長坂：可能性測度による区間回帰分析；行動計量学，Vol. 16, pp. 1–7 (1988)
- [11] H. Tanaka, S. Uejima and K. Asai: Linear regression analysis with fuzzy model; *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-12, pp. 903–907 (1982)
- [12] 山川，嶋瀬，江本：モデル誤差を伴うシステムにおける最適コストの評価手法；システム制御情報学会論文誌，Vol. 18, pp. 108–117 (2005)