日本オペレーションズ・リサーチ学会
2023 年春季研究発表会

一次の最大・最小の期待値に対する 誤差とブレイクポイント数の保証付き区分線形近似

05000653 青山学院大学 *高澤 陽太朗 TAKAZAWA Yotaro

1. はじめに

確率計画においては、確率変数を X としたときに、以下のような期待値計算を伴う関数 $c_X : \mathbb{R} \to \mathbb{R}$ がしばしば登場する。

$$c_X(z) = \mathbb{E}[\max(a_1 z + b_1 X, a_2 z + b_2 X)] \quad (1)$$

ただし、$a_i, b_i \in \{-1, 0, 1\}$ $(i \in \{1, 2\})$ である。例えば、ある商品の仕入れや在庫管理の文脈で、その商品の不確実な需要量を X、その仕入れ量を s とすると

- $\mathbb{E}[\max(X - z, 0)]$ = (期待機会損失)
- $\mathbb{E}[\max(z - X, 0)]$ = (期待在庫コスト)
- $\mathbb{E}[\min(X, z)]$ = (期待売上数)

といった形式で利用されることが多い。初めの二つは、一次の損失関数（first-order loss function）と呼ばれている [4]。

一般に関数 (1) は非線形であるため、MILP（Mixed Integer Linear Programming）といったモデルに組み込む際には、区分線形近似といった近似手法が用いられる。ある関数 f の区分線形近似とは、関数の定義域からブレイクポイントと呼ばれる点をいくつか選び、ブレイクポイント z とその関数値 $f(z)$ からなる f のグラフ上の点 $(z, f(z))$ を順に結んで得られる区分線形関数によって元の関数を近似する方法である。

区分線形近似を行う際に、適切なブレイクポイント数の決定は重要な課題である。一般にブレイクポイント数を増やすことで近似精度は向上するが、MILP において最適解を求めるための計算量が増えるというトレードオフがあるためである。このトレードオフを事前に定量的に把握することは難しく、一般的には予備実験を通して適切なブレイクポイント数を決めていることが多い。

一次の損失関数に対する区分線形近似の既存研究では、ブレイクポイント数を固定したときの誤差が最小となる条件を導出し、最適なブレイクポイントの位置を求めるというアプローチが取られている [3]。これは、凸関数の区分線形近似の研究 [1] が土台となっている。これらの研究では、前もってブレイクポイントと誤差の関係が定量的にわからないことに加え、

1. 対象となる関数が微分可能かつ狭義の凸関数である必要がある
2. 最適なブレイクポイント配置のための非凸最適化問題を解く必要がある

という制限がある。近年、上記の制限はあるものの、凸関数の区分線形近似を対象にブレイクポイントの数と誤差の関係に対する解析が与えられた [2]。

本研究では、関数 (1) に対して、精度とブレイクポイント数の両方に保証のある区分線形近似関数を得るシンプルな方法を提案する。この手法では、関数 (1) に限っていえば、事前に精度とブレイクポイントの関係が定量的に把握できるため、予備実験無しに、適切なブレイクポイント数をもつ区分線形近似を得ることが可能となる。

2. 一次の損失関数に対する区分線形近似

本研究は Rossi [3] らの一次の損失関数に対する区分線形近似の分析が土台となる。まず、いくつかの前提を確認する。X は期待値をもつ確率変数とし、その台を $S \subseteq (a, b] \subseteq \mathbb{R}$ と仮定する。関数 (1) を関数 $\tilde{c} : \mathbb{R} \to \mathbb{R}$ で近似する場合、その誤差は

$$\max_{z \in \mathbb{R}} |c_X(z) - \tilde{c}(z)| \quad (2)$$

として評価する。

Rossi らの研究では、一次の損失関数の区分線形近似を行う際に、定義域からブレイクポイントを選ぶ代わりに確率変数のとりうる値の領域の分割を考え、新たな確率変数を考えるような方針をとる。

定義 1. 確率変数 X が与えられたときに、X がとり得る領域 $(a, b]$ に対して次のような分割 $I = (I_1, \ldots, I_n)$ を考える:

- $I_i = (a_i, b_i]$ とすると、 $i \in \{1, \ldots, n-1\}$ に対して $b_i = a_{i+1}$ が成立,
- $a_1 = a$ かつ $b_n = b$,
- $P(X \in I_i) > 0 \ \forall i \in \{1, \ldots, n\}$.

上記を満たす I が与えられた際に、確率 $P(X \in I_i)$ で値 $\mathbb{E}[X \mid X \in I_i]$ をとる離散確率変数 $\tilde{X}$ を考える。この $\tilde{X}$ を I によって X から誘導された確率変数と呼ぶ。

このとき、(1) において X を $\tilde{X}$ に置き換えた関数 $c_{\tilde{X}}$ はブレイクポイントとして $\mathbb{E}[X \mid X \in I_i]$ を持つ c_X の区分線形近似関数となる [3]。本研究においてもこの結果に基づいて、ブレイクポイントを選ぶ代わりに領域の分割を考える方針をとる。

3. 結果 1: 誤差とブレイクポイント数の関係

本研究の一つ目の結果は、最適なブレイクポイントの仮定を必要としない区分線形近似の誤差評価をした点である。これによって（非凸最適化問題の解となる）最適なブレイクポイントを求めなくとも誤差の評価が可能となった。具体的には、定義 1 の条件を満たす任意の分割 I によって X から誘導された $\tilde{X}$ を利用した際の誤差評価を行った。

定理 1. $\tilde{X}$ を定義 *1* の条件を満たす分割 I によって X から誘導された離散確率変数とする。このとき、以下の不等式が成立する。

$$\max_{z \in \mathbb{R}} |c_X(z) - c_{\tilde{X}}(z)| \leq \max_{i \in \{1, \ldots, n\}} P(X \in I_i) \frac{b_i - a_i}{4} \quad (3)$$

さらに定理 1 より、誤差 $\epsilon > 0$ を達成するために必要なブレイクポイント数の下界を得た。

結果 1. 任意の区間 $(a, b]$, 誤差 $\epsilon > 0$ に対して、次を満たすような確率変数 X が存在する。定義 *1* の条件を満たす任意の分割 I によって X から誘導された確率変数 $\tilde{X}$ が

$$\max_{z \in \mathbb{R}} |c_X(z) - c_{\tilde{X}}(z)| \leq \epsilon \quad (4)$$

を満たすならば、I の要素数（=ブレイクポイント数）は次の値以上である*:*

- $\left\lfloor \frac{\sqrt{2}}{4} \sqrt{\frac{b-a}{\epsilon}} \right\rfloor$ *(*X が連続の時*)*
- $\left\lfloor \frac{\sqrt{2}}{2} \sqrt{\frac{b-a}{\epsilon}} \right\rfloor$ *(*X が離散の時*).*

4. 結果 2: 区分線形近似関数の構成方法

定理 1 の主張を踏まえて、誤差が ϵ を超えないように定義 1 の条件を満たす分割 I を構成する以下のようなシンプルな方法を提案する。

提案手法（概略）

入力: 区間 $(a, b]$, 分布関数 F_X, 許容誤差 $\epsilon > 0$.

出力: 分割 $I = ((a_1, b_1], \ldots, (a_n, b_n])$.

Step 1: $i = 1$、$a_i = a$ とする。

Step 2: b_i を $(F_X(b_i) - F_X(a_i)) \frac{b_i - a_i}{4} \leq \epsilon$ を満たす最大の $b_i \in (a, b]$ とする。$b_i = b$ ならば終了。そうでないならば、$a_{i+1} = b_i$、$i = i + 1$ として Step2 の最初に戻る。

本研究では、上記手法が誤差とブレイクポイント数の両方に保証がある方法であることを示した。

結果 2. 確率変数 X に提案手法を適用して得られた分割を I とし、$\tilde{X}$ を I によって X から誘導された確率変数とする。このとき、以下を満たす。

$$\max_{z \in \mathbb{R}} |c_X(z) - c_{\tilde{X}}(z)| \leq \epsilon \quad (5)$$

さらに、分割 I の要素数は次の値で抑えられる。

- $\frac{1}{2} \sqrt{\frac{b-a}{\epsilon}} + 1$ *(*X が連続の時*)*
- $\frac{3}{4} \sqrt{\frac{b-a}{\epsilon}} + 1$ *(*X が離散の時*)*

この結果を結果 1 で得られた下界と照らし合わせると、かなり近い値となっていることがわかる。

さらに、正規分布などを対象に数値実験を行ったところ、結果 2 で得られた分割数の上界と実際に提案手法で生成される分割数はそれほどギャップがないことがわかった。数値実験の結果は当日発表にて報告する。

参考文献

[1] Alyson Imamoto and Benjamim Tang. A recursive descent algorithm for finding the optimal minimax piecewise linear approximation of convex functions. In *Advances in Electrical and Electronics Engineering-IAENG Special Edition of the World Congress on Engineering and Computer Science 2008*, pages 287–293. IEEE, 2008.

[2] Bo Liu and Yi Liang. Optimal function approximation with relu neural networks. *Neurocomputing*, 435:216–227, 2021.

[3] Roberto Rossi, S Armagan Tarim, Steven Prestwich, and Brahim Hnich. Piecewise linear lower and upper bounds for the standard normal first order loss function. *Applied Mathematics and Computation*, 231:489–502, 2014.

[4] Lawrence V Snyder and Zuo-Jun Max Shen. *Fundamentals of supply chain theory*. John Wiley & Sons, 2019.