極値統計に基づく近傍構造解析

01107771　小樽商科大学　*加地太一　KAJI Taichi

1. はじめに

メタヒューリスティクスの解法 [1] は，解の移動を反復することが基調となっている．解の移動操作のベースを作り上げているものが近傍であり，得られる解の良し悪しが近傍により決定される．

著者等 [3] は，組合せ最適化問題の解空間において近傍点のランダムな評価値系列が AR(1) プロセスで支配され，解コストの同時分布にガウス性が伴う仮定を利用して汎用的に解析可能な近傍モデルを構築し，近傍の最小値を推定し，解の移動の性能の解析を試み良好な結果を得た．ただし，多変量正規分布構造において複雑な計算が要求され簡便に利用することが難しい．

そこで，本研究では極値統計学の考えに基づき，容易な方法で近傍の最小値を推定する．極値統計学は，極端な現象，すなわち，母集団分布の端（裾）に対する推測を行うことを対象とし，災害における極端な自然現象や，ファイナンスなどでのリスク評価のために応用される統計手法である．この考えを用い，近傍集合の分布の端（裾）における値を推定し，近傍の最小値に関して論ずる．

2. AR(1) による近傍構造の解析

最小値を推定するために，近傍の確率分布をモデル化しておかねばならない．そこで，組合せ最適化問題の解空間における近傍点のランダムな評価値系列が AR(1) プロセス (first order autoregressive process) で支配されるという仮定を用いてその特徴を導き出す．近傍構造の解析は Kaji[3] により行われており，今回必要な値に関する概略のみ下記に記す．

まず，解空間上を定義した近傍に基づきランダムに移動する評価値 (fitness) の系列 $x_1, x_2, \ldots, x_t, \ldots, x_N$ を考え，F_t をステップ t における解 x_t のコストを示す確率変数とする．また，Δ は平均 0，分散 σ_Δ^2 をもつホワイトノイズで，μ は解コストとの期待値であり，$\rho(1)$ は 1 ステップの自己相関関数とする．このとき，$\{F_t\}$ が，

$$F_t = \mu + \rho(1)(F_{t-1} - \mu) + \Delta \tag{1}$$

の差分方程式で表現でき AR(1) としてモデル化でき，そこから，解コストの平均 μ，解コストの分散 σ^2，解 x_0 のコストとその近傍解コストの相関係数 ρ を推定できる．

さらに，解空間が AR(1) として近似できるならば，与えられた解 x_0 の近傍のコストは，

$$\phi(x) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} \exp\left(-\frac{(x-\hat{\mu})^2}{2\hat{\sigma}^2}\right), \tag{2}$$

に従う正規分布としてモデル化できることが示されており，

$$\hat{\mu} = \mu + \rho(f(x_0) - \mu)), \tag{3}$$
$$\hat{\sigma}^2 = \sigma^2(1-\rho^2) \tag{4}$$

と計算できる．

3. 極値統計による近傍構造解析に対するアプローチ

Kaji[3] による近傍の最小値推定は精度が高いが，複雑な計算を要求され簡便に利用することが難しい．そこで，本節で，極値統計の考えに基づき近傍の最小値を容易に推定できることを示す．

極値統計とは，独立な同一な分布に従う N 個の確率変数 $X_1, \ldots, X_N$ に対し，最大値 $x_{\max} = \max\{X_1, \ldots, X_N\}$，あるいは最小値 $x_{\min} = \min\{X_1, \ldots, X_N\}$ の分布を考察する研究分野である．その応用として，川の氾濫の度合いを定量化して，その最大値を推定することで堤防の高さを決める問題などで利用されている．

この極値統計における 1 つの方法 [4] を用いて，近傍の最小値を検討してみる．ここで，$\{X_1, \ldots, X_i, \ldots, X_N\}$ を近傍集合とし，各 X_i は近傍解のコストを示す確率変数であり，それらは同一の確率密度関数 $f(x)$ に従い，かつ独立であると仮定する．この仮定のもと，$x_{\min} = \min\{X_1, \ldots, X_N\}$ を導出すれば，$x_{\min}$ が近傍の最小値に対応する．また，$f(x)$ は近傍をモデル化した確率密度関数とし，$f(x)$ の累積分布関数を $F(x)$ と書き，$F'(x) = f(x)$ となる．

表 1: eil101.tsp における極値統計による推定値

f(x)	サンプル	Kaji	極値統計
3173.2	3073.6	3108.2	3081.2
2543.1	2465.2	2490.6	2463.5
2001.9	1947.5	1960.1	1933.1
1474.4	1437.3	1443.0	1415.9
697.3	696.8	681.3	654.2

$x_{\min}$ が x 以下になる確率 $H(x)$ を求めると，

$$\begin{aligned}H(x) &= \text{「少なくとも 1 つは } x \text{ 以下になる確率」}\\ &= 1 - \text{「}X_1, \ldots, X_N\text{が全て } x \text{ 以上になる確率」}\\ &= 1 - \left\{\big(1 - P(X_1 \le x)\big)\cdots\big(1 - P(X_N \le x)\big)\right\}\\ &= 1 - \{1 - F(x)\}^N \qquad (5)\end{aligned}$$

$x_{\min}$ の分布の確率密度関数 $h(x)$ は，$H(x)$ を x で微分すれば求まり，

$$h(x) = H'(x) = N\{1 - F(x)\}^{N-1} \cdot f(x) \qquad (6)$$

となる．さらに，最小値を満たす条件は，

$$\begin{aligned}h'(x) = (N-1)\{1 - F(x)\}^{N-2} \cdot \{-f^2(x)\}\\ + \{1 - F(x)\}^{N-1} \cdot f'(x) \qquad (7)\end{aligned}$$

の値が 0 となる x を導けばよい．すなわち，確率密度関数 $h(x)$ の最頻値を求める．また，(7) 式において，$N \gg 1$ のときは，$F(x_{min}) \approx 0$，および $N - 1 \approx N$ であるから，$x_{\min}$ は，

$$f'(x_{\min}) - Nf^2(x_{\min}) = 0 \qquad (8)$$

の条件を満たし，おおよその最小値として $x_{\min}$ を算出できる．

4.　数値計算

(8) 式を用いて x_{min} を推定するために，$f(x)$ は (2) 式による正規分布の確率密度関数とし，N は近傍の大きさとする．微分の計算は，SciPy(Python) に組み込まれている derivative() 関数を用いて解く．(8) 式を解く求根アルゴリズムは，SciPy(Python) の brentq() 関数を用いて，Brent 法で 0 点近くの収束に逆 2 次補間を組み合わせた方法で解く．この手法は，二分法による確実な収束性と，根の囲い込み法による速さを持った特徴を有している．

表 1 に，TSPLIB[2] の事例 eil101 に対しての推定値の結果を示す．この表の項目では，現在の解 x のコストを “$f(x)$” とし，解 x の近傍の実際の最小値を実験的に求めた値を “サンプル” として示す．論文 [3] による最小値の推定値の項目名を “Kaji”，今回提案した極値統計に基づく推定結果の項目名を “極値統計” とし，それぞれの推定値を記しておく．ここで，近傍の平均，分散の値が必要であるが，(3) 式，(4) 式から求めた値を利用する．表 1 において，実際の値に近い結果が得られ，局所解に近い領域以外では，今回提案した推定法が，Kaji による推定より優れた結果を示している．ただし，局所解に近い領域では，従来の方法が勝っているが，簡便に利用するにあたって十分な推定値を導出していると考える．

5.　おわりに

今回は，従来手法と比較し，近傍の最小値を極値統計の考えに基づき推定した．その結果，最小値の推定としては良好な結果が得られたが，実験対象は，TSP に対し 2-opt 近傍における検証のみであり，今後はさらに異なる近傍，異なる問題に実験対象を広げ解析の性能，特徴を明らかにしなければならない．また，極値統計学における逆関数法を用いた方法などがあり，それらの手法など用いてより検討を加えていきたい．

参考文献

[1] T. F. Gonzalez, editor. *Handbook of Approximation Algorithms and Metaheuristics.* Chapman & Hall/CRC, second edition, 2018.

[2] D. S. Johnson and L. A. McGeoch. Experimental analysis of heuristics for the STSP. In G. Gutin and A. P. Punnen, editors, *The Traveling Salesman Problem and Its Variations*, pages 369–443. Springer, 2002.

[3] T. Kaji. A probabilistic analysis of neighborhoods for combinatorial optimization problems and its application. *Journal of Heuristics*, 27(6):1057–1079, 2021.

[4] 蒼馬 竜. **極値統計によろしく**. 暗黒通信団, 2018.