

確率計画問題のための GA の統計的選択手法

所 健 一†

本論文では目的関数や制約式のパラメータが複数のシナリオとして与えられる、複雑な確率計画問題を効率的に解く新しい遺伝的アルゴリズムを提案する。提案アルゴリズムでは、ランダムに選択したシナリオを用いて個体を評価し、統計的な検定 (Welch 検定) により最良個体とはなりえないと判断された個体のみを淘汰する。これにより個体評価に要する計算時間を削減するとともに、サンプリング誤差により適合度の高い個体が淘汰される危険率を一定値以下に抑えることが可能となる。確率的施設配置問題に提案アルゴリズムを適用したところ、期待利益の高い配置案を非常に短い計算時間で求めることができ、提案アルゴリズムの有効性が確認できた。

A Statistical Selection Mechanism of GA for Stochastic Programming Problems

KEN-ICHI TOKORO†

We propose a new genetic algorithm to solve complex stochastic programming problems, in which possible combinations of random parameters are provided as scenarios. In the algorithm, individuals are evaluated based on mean fitness in some scenarios sampled at random, and selection is carried out based on statistical theory (i.e., Welch's test). Thus, the computational demand for evaluating the individuals is reduced, and the probability that good individuals are excluded by sampling error is limited. Our approach significantly reduces computing time required to find high quality solutions for stochastic facility location problems.

1. はじめに

不確実性を考慮した最適化が行える確率計画法は、現実の意思決定に非常に有用である。線形計画法など、確定環境下における最適化を対象とした一般的な最適化手法では、与えられた目的関数や制約に関する 1 組の入力パラメータに対して最適解を求めている。しかし現実の意思決定においては、パラメータの実現値が得られる以前の段階で意思決定をしなければならないケースも多く、必ずしもパラメータが一意に確定するとは限らない。そこで、こうした不確定環境下における最適化を行う手法として、与えられる入力パラメータの実現可能な複数の組合せ (シナリオ) に対して、平均的に高い最適性を発揮する解を求める、確率計画法の研究が行われている。

確率計画法については、これまでもさまざまな研究が行われてきたが、現実の意思決定をよりいっそう有効に支援するためには、より複雑な問題を効率的に解く手法の開発が望まれる。確率計画法の伝統的な手

法の多くは、一般的な線形、あるいは非線形計画法が適用可能な問題を対象としている。しかし、現実の意思決定問題の多くは非線形、不連続、離散的な問題として定式化され、より複雑である。また多くの研究では、3 から 5 程度のシナリオが与えられる問題を対象としているが、現実の意思決定を行ううえでは、より多くのシナリオを考慮した最適化が必要となる。

本論文では複雑な確率計画問題を効率的に解く手法として、複雑な最適化問題に対する有効な解法の 1 つである、遺伝的アルゴリズム (Genetic Algorithm: GA)^{1),2),4),9)} の適用を検討する。これまでも GA を用いて、複雑な確率計画問題を解く研究が行われている^{3),7),10)~12)}。これらの研究では、効率的に解を得るための方法としてサンプリングアプローチが採用されており、モンテカルロサンプリングなどにより選び出したシナリオの一部を用いて個体を評価することで、個体評価に要する計算時間を削減している。しかし、これまでに行われた研究では、選び出した一部のシナリオで個体を評価・選択することの理論的妥当性は示されていない。また、サンプル数を少なくすることで計算効率は高まるものの、個体評価の誤差が増し、得られる解の精度が悪化する。

† 電力中央研究所情報研究所
Communication & Information Research Laboratory,
Central Research Institute of Electric Power Industry

本論文ではサンプリングアプローチにより計算の効率化を図りつつ、統計的検定を用いた選択を行うことで、得られる解の精度を維持する GA を提案する。提案アルゴリズムでは、個体評価のたびに新たなシナリオをサンプルする。そして、現在までの各世代における個体の適合度の分布を基に統計的な検定 (Welch 検定) を行い、最良個体とはなりえないと判断された個体のみを淘汰の対象とする。淘汰の対象とならなかった個体については次世代へと生き残らせ、新たにサンプルしたシナリオでの評価を付け加えていく。これにより選り出した一部のシナリオで個体を評価することによる誤差 (サンプル誤差) が原因で、GA の探索過程で生成された最良個体が誤って淘汰されてしまう確率は一定値以下に抑えられる。また、より多くの世代を生き残る個体ほど、多くのシナリオで評価されることになり、その個体の評価値はサンプル誤差の少ない、より正確な値となっていく。

提案アルゴリズムの有効性を確認するため、代表的な確率計画問題の 1 つである確率的施設配置問題を対象とした数値実験を行った。27 個のシナリオが与えられる有限シナリオのケースと、シナリオが確率分布として与えられる無限シナリオのケースの、2 つのケースについて数値実験を行ったところ、いずれのケースにおいても提案アルゴリズムの有効性が確認できた。有限シナリオのケースにおいては提案アルゴリズムにより、すべてのシナリオを用いて個体を評価した場合と期待利益がほぼ等しい解を、約 1/3 の計算時間で求めることができた。また、無限シナリオのケースにおいては、100 個のシナリオを用いて個体を評価した場合と同程度の期待利益の解が、約 1/10 の計算時間で求められた。

本論文では次章において対象とする確率計画問題と、これまでに行われた GA の適用に関する研究について説明する。次に 3 章において、サンプルしたシナリオにおける適合度を基に、統計的手法を用いて個体選択を行う、新たな GA を提案する。そして 4 章において、この提案アルゴリズムを確率的施設配置問題へと適用した結果について報告する。

2. 確率計画問題と GA の適用

本論文では図 1 に示す、不確実なパラメータの実現可能な組合せがシナリオとして与えられる、リコース問題と呼ばれるタイプの確率計画問題を対象に検討を行う。ここでリコースとは、シナリオ確定前に行った意思決定に対して、シナリオ確定後にシナリオの実現値と意思決定との差異を埋めるために必要となるコス

$$\max_x z = f(x) - \sum_{k \in K} \pi_k \cdot Q(x, \xi_k) \quad (1)$$

$x = (x_1, x_2, \dots, x_n)$ 決定変数

$f(x)$: 目的関数のシナリオに依存しない部分

K : シナリオの集合

ξ_k : シナリオ k が発生した場合のパラメータの実現値

π_k : シナリオ k の実現確率

$$Q(x, \xi_k) = \min\{g(y, \xi_k) \mid h_j(y, \xi_k) \geq T_j(x, \xi_k), j = 1, 2, \dots, m\}$$

図 1 対象とする確率計画問題

Fig. 1 Formulation of a stochastic programming problem.

トを意味する。図 1 の定式化では $Q(x, \xi_k)$ が、シナリオ確定前に行った決定 x に対して、シナリオ k が実現した場合のリコースを表している。

図 1 の確率計画問題へ GA を適用する単純な方法は、決定変数 x を個体の染色体にコード化し、式 (1) の z で個体を評価することである。しかし、このためにはすべてのシナリオに対して $Q(x, \xi_k)$ を計算しなければならず、与えられたシナリオが少数の場合を除き、個体評価に多くの計算時間が必要となる。また無限シナリオのケースでは、シナリオの分布が特別な確率分布に従う場合を除き、全シナリオにおける平均適合度 (z) を計算することはできない。

このため、これまでに行われた確率計画問題へ GA を適用した多くの研究では、モンテカルロサンプリングなどにより選り出したシナリオの一部を用いて個体を評価する、サンプリングアプローチが採用されている^{3), 7), 10)~12)}。しかし、これまでの研究では、サンプルしたシナリオを用いて個体を評価・選択することの理論的妥当性は示されていない。また、サンプル数を少なくすることで計算効率は高まるものの、個体評価の誤差が増し、得られる解の精度が悪化する。

サンプルした 100 から 200 のシナリオを個体評価に用いている文献^{3), 10), 11)}の研究に対して、Loughlin らは各個体を評価するたびに 10 個のシナリオをサンプルし、これを用いて個体を評価することで効率的に解を求める GA を提案している⁷⁾。このアルゴリズムにより、個体評価に要する計算時間は大幅に削減される。しかし少数のサンプルしか用いないことで個体評価の誤差が増し、文献⁷⁾の数値実験では、厳密には問題の制約条件を満たさない個体が最良個体として評価されている。Loughlin らの研究では、ある一定確率以内であれば制約条件を満たさないシナリオの存在を許容する、機会制約条件問題と呼ばれる確率計画問題を対象としているが、得られた解において制約を

満たさないシナリオの割合は、必ずしも許容値を満たしていない。このため Loughlin らは、制約を満たさないシナリオの許容割合を 20% と厳しくすることで、制約を満たさないシナリオの割合が 25% 以下となる解を求めている。しかし、本論文で扱うリコース問題においては、このような方法で解の精度を改善することはできず、個体評価の誤差はそのまま目的関数値の悪化につながる。また、Yoshitomi らは各世代ごとに 1 つのシナリオを選び出し、このシナリオに対する適合度で個体を評価し、最終的に全世代を通して最も出現頻度の高かった個体を最良個体とする GA を提案しているが、この方法でもシナリオ数が 2 の単純な問題においても最適解が得られないケースがあるなど、得られる解の精度に問題がある¹²⁾。

3. アルゴリズム

複雑な確率計画問題を効率的に解くアルゴリズムとして、少数のサンプルしたシナリオにおける適合度で個体を評価しつつ、統計的検定を用いた選択を行うことで解の精度を維持する、以下に示す GA を提案する。

3.1 個体評価

シナリオの発生確率に比例した確率で、モンテカルロサンプリングなどの方法によりシナリオを選び出し、式 (2) で計算される、サンプリングされたシナリオにおける平均適合度で個体を評価する。

$$z'(S) = f(x) - 1/|S| \sum_{k \in S} Q(x, \xi_k) \quad (2)$$

ここで $S \subset K$ はシナリオの発生確率 π_k に比例した確率で、与えられたシナリオ集合 K から選び出したシナリオの集合を表す。また、 $|S|$ は集合 S の要素数である。

3.2 $z'(S)$ の分布

$z'(S)$ の値は選び出したシナリオ S に依存する。ただし、 $z'(S)$ の値は個々のシナリオ $k \in S$ における、ある個体の適合度の平均であるので、サンプルするシナリオ数 ($|S|$) が十分に大きければ、中心極限定理により、 $z'(S)$ の分布は正規分布に従うと仮定することができる。さらに大数の法則により、シナリオをサンプルし、 $z'(S)$ を計算することを繰り返すと、得られる $z'(S)$ の分布の平均は全シナリオにおける個体の平均適合度 z に収束すると仮定できる。なお、ここでは上述のように、中心極限定理により $z'(S)$ の分布に正規分布を仮定するため、シナリオを選び出す確率にのみシナリオの発生確率を反映し、 $z'(S)$ の計算にはシナリオの発生確率を用いない。

たとえば与えられたシナリオに対して、ある個体の

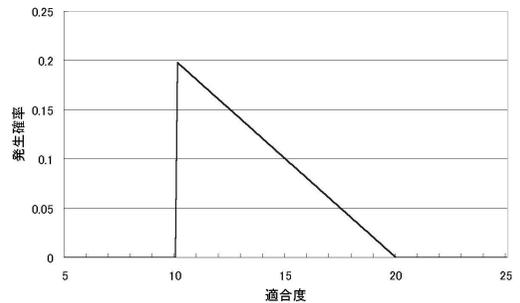


図 2 個々のシナリオにおける適合度 ($z'(k)$) の分布 (1)
Fig. 2 Distribution of $z'(k)$ in each scenario (1).

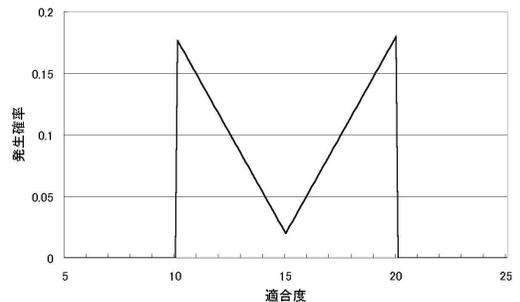


図 3 個々のシナリオにおける適合度 ($z'(k)$) の分布 (2)
Fig. 3 Distribution of $z'(k)$ in each scenario (2).

適合度が図 2、図 3 のように分布しているとする。このとき、シナリオの発生確率に比例した確率で、それぞれ 5, 10, 20 個のシナリオを選び出し、この個体の平均適合度 ($z'(S)$) を計算することを 100 回繰り返すと、得られる $z'(S)$ の分布は表 1、表 2 のようになる。ここで表 1、表 2 の各列の括弧内の数値は、 $z'(S)$ が平均 z の正規分布に従った場合の理論値を表している。このように無限シナリオで、しかも適合度が平均とは大きく異なるシナリオの発生確率が高い図 2 や図 3 のケースにおいても、 $z'(S)$ の分布はほぼ正規分布と一致する。

3.3 統計的選択方法

シナリオをサンプルし、各個体の $z'(S)$ を計算することを繰り返すことで得られる $z'(S)$ の分布が、平均 z の正規分布に従うとすると、以下に示す Welch 検定を用いることで、2 つの異なる個体の z の差の有意性を、いくつかのサンプルシナリオ S に対する $z'(S)$ の値を基に推定することができる。

Welch 検定

$x_i, i = 1, 2, \dots, m$ と $y_j, j = 1, 2, \dots, n$ をそれぞれ $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ から選び出した m, n 個のサンプルとする。このとき、式 (3) で計算される t は、自由度 ν の t 分布に従う。

表 1 図 2 のシナリオにおける $z'(S)$ の分布

Table 1 Distribution statistics of $z'(S)$ when fitness of the individual is distributed as Fig. 2.

| サンプル数 (S) | 平均 | | 分散 | | 歪度 | | 尖度 | |
|-------------|-------|---------|------|--------|-------|--------|-------|-------|
| 5 | 13.31 | (13.33) | 1.08 | (1.11) | 0.15 | (0.00) | -0.55 | (3.0) |
| 10 | 13.40 | (13.33) | 0.51 | (0.56) | 0.12 | (0.00) | 0.10 | (3.0) |
| 20 | 13.32 | (13.33) | 0.33 | (0.28) | 0.001 | (0.00) | -0.74 | (3.0) |

表 2 図 3 のシナリオにおける $z'(S)$ の分布

Table 2 Distribution statistics of $z'(S)$ when fitness of the individual is distributed as Fig. 3.

| サンプル数 (S) | 平均 | | 分散 | | 歪度 | | 尖度 | |
|-------------|-------|---------|------|--------|-------|--------|-------|-------|
| 5 | 14.95 | (15.00) | 2.17 | (2.33) | 0.02 | (0.00) | -0.38 | (3.0) |
| 10 | 15.18 | (15.00) | 1.06 | (1.17) | 0.03 | (0.00) | 0.32 | (3.0) |
| 20 | 15.00 | (15.00) | 0.68 | (0.58) | -0.16 | (0.00) | -0.51 | (3.0) |

$$t = -\frac{X - Y}{\sqrt{s_x^2/m + s_y^2/n}} \quad (3)$$

ここで、 X, Y はそれぞれ x_i, y_j の平均を表し、 s_x^2, s_y^2 はそれぞれ x_i, y_j の分散を表している。また、 ν は式 (4) で計算される ν' に最も近い整数である。

$$\nu' = \frac{(s_x^2/m + s_y^2/n)^2}{\frac{(s_x^2/m)^2}{m-1} + \frac{(s_y^2/n)^2}{n-1}} \quad (4)$$

提案アルゴリズムでは、この Welch 検定を用いて、現世代の最良個体と比較して、有意に適合度が低いと判断された個体のみを淘汰の対象とし、この中から各世代ごとにある一定数を上限に適合度の低い個体を淘汰する。

3.4 提案アルゴリズムのフロー

提案アルゴリズムでは以下に示す手順により、サンプルしたシナリオにおける適合度から個体を評価・選択し、最適な解を探索する。

(1) 初期化

- (1-1) ランダムに N 個の個体からなる初期集団を生成する。
- (1-2) 個体 i の履歴データを $g_i = Z_i = \sigma_i^2 = 0$ とする。ここで g_i は個体 i が生き延びた世代数を表し、 Z_i と σ_i^2 はそれぞれ個体 i が生き延びた過去の世代における適合度 $z'(S)$ の平均と分散を表す。

(2) 個体の評価

- (2-1) 各シナリオの発生確率に比例した確率で、与えられたシナリオから m 個のシナリオをランダムに選び出す。
- (2-2) 選び出した m 個のシナリオの集合を S とし、シナリオ集合 S に対する個体 i の平均適合度 $z'_i(S)$ を計算する。
- (2-3) 個体 i の履歴データを更新する。

$$g_i = g_i + 1$$

$$Z_i = Z_i + (z'_i(S) - Z_i)/g_i$$

$$\sigma_i^2 = \frac{(g_i - 2)(z'_i(S) - Z_i)^2}{g_i} + \frac{(g_i - 2)\sigma_i^2}{g_i - 1}$$

(3) 個体の選択

(3-1) 現在の世代の中で Z_i が最大となる個体を選び、これを i^* とする。

(3-2) 個体の集合 $D = \{j \mid t_j > t_\alpha(\mu)\}$ を選ぶ。ここで

$$t_j = \frac{Z_{i^*} - Z_j}{\sqrt{\sigma_{i^*}^2/g_{i^*} + \sigma_j^2/g_j}}$$

である。また、 $t_\alpha(\mu)$ は自由度 μ 、信頼区間 $\alpha\%$ の t 値であり、 μ は

$$\mu' = \frac{(\sigma_{i^*}^2/g_{i^*} + \sigma_j^2/g_j)^2}{\frac{(\sigma_{i^*}^2/g_{i^*})^2}{g_{i^*}-1} + \frac{(\sigma_j^2/g_j)^2}{g_j-1}}$$

に最も近い整数を表す。

(3-3) 集合 D に含まれる個体数が n_w 以上なら、適合度の低い順に n_w 個の個体を淘汰する。そうでなければ集合 D に含まれるすべての個体を淘汰する。

(4) 再生

- (4-1) 現世代の集団から $\min\{n_w, |D|\}$ 個の個体を選び出す。
- (4-2) 選び出した個体からランダムに個体ペアを作り、交叉の遺伝的操作を加えて新しい個体を生成する。
- (4-3) 新しい個体に突然変異の遺伝的操作を加える。
- (4-4) 新しい個体の履歴データを初期化する。

(5) 終了判定

終了条件を満たしたならアルゴリズムを終了する。

g_i の値が一定数以上で、 Z_i の値が最大となる個体を最良個体として選択する．そうでなければ (2) からの操作を繰り返す．

3.5 提案アルゴリズムの特徴

この提案アルゴリズムでは、現世代で平均適合度 (Z_i) が最大となる個体と比べ、Welch 検定により有意に平均適合度が低いと判断された個体のみを淘汰の対象とし、他の個体については次の世代へと生き残らせる．これにより GA の探索過程で生成された最良の個体が、サンプル誤差により誤って淘汰されてしまう確率は、あらかじめ定めた一定値以下に抑えられる．理論的には各世代の選択で最良の個体が淘汰されてしまう確率は $\alpha\%$ 以下となる．また生き残った個体については、次の世代で新たにサンプルされたシナリオでの評価が加えられることで、その平均適合度 (Z_i) はより正確な (z に近い) 値となり、 $z'(S)$ のサンプルが増えることから、より正確な差の検定が行えるようになる．

このように提案アルゴリズムでは GA の世代が進行していく中で、より多くの世代を生き残る個体ほど、その平均適合度はより正確な値となっていく．一方、平均適合度の低い個体については、ある程度の精度の適合度が得られた段階で淘汰されるので、適合度の低い個体の評価に要する計算時間は削減される．

4. 数値実験

提案アルゴリズムの有効性を確認するため、代表的な確率計画問題の 1 つである確率的施設配置問題^{(6),(8)}を対象とする数値実験を行った．施設配置問題とは最適な施設の規模と、与えられた候補地の中から最適な施設の設置場所を決める問題である．ここでは需要、製造コスト、輸送コスト、製品の価格を不確定パラメータとし、これらのパラメータが有限シナリオとして与えられるケースと、確率分布として与えられる無限シナリオのケースの、2 つのケースについて数値実験を行った．

4.1 確率的施設配置問題の定式化

確率的施設配置問題を定式化するにあたり、以下の記号を定義する．

I : 需要地の集合

J : 施設が設置可能な候補地の集合

K : 与えられたシナリオの集合

x_j : 候補地 j に施設が設置されるなら 1, その他の場合は 0 の値をとる変数

$y_{i,j}^k$: シナリオ k が生じた場合、需要地 i の需要のうち、候補地 j に設置された施設から供給される製

品の割合

f_j : 候補地 j に施設を設置するために必要となる固定費

v_j : 候補地 j に設置する施設の生産能力を表す変数

g_j : 候補地 j に施設を設置するために必要となる、工場の生産能力に比例する変動費

d_i^k : シナリオ k が生じた場合の需要地 i の需要

p_i^k : シナリオ k が生じた場合の需要地 i での製品 1 単位あたりの価格

$c_{i,j}^k$: シナリオ k が生じた場合の、候補地 j の施設で製品を製造し、これを需要地 i へと供給するために必要となる、製造コストと輸送コストの合計

π_k : シナリオ k の発生確率

上記の記号を用いると、確率的施設配置問題は以下のように定式化される．

確率的施設配置問題

$$\begin{aligned} \max_{x_j, y_{i,j}^k, v_j} Z_p = & - \sum_{j \in J} g_j \cdot v_j - \sum_{j \in J} f_j \cdot x_j \\ & + \sum_{i \in I, j \in J, k \in K} \pi_k \cdot d_i^k (p_i^k - c_{i,j}^k) y_{i,j}^k \\ \text{s.t.} & \sum_{j \in J} y_{i,j}^k \leq 1, i \in I \\ & \sum_{i \in I} d_i^k \cdot y_{i,j}^k - v_j \leq 0, j \in J, k \in K \\ & y_{i,j}^k - x_j \leq 0, i \in I, j \in J, k \in K \\ & x_i \in \{0, 1\}, i \in I \\ & y_{i,j}^k \geq 0, i \in I, j \in J, k \in K \\ & v_j \geq 0, j \in J \end{aligned}$$

問題のパラメータについては、文献 8) などで行われている代表的なベンチマーク問題 (Kuehn-Hamburger 問題⁵⁾) を参考に、以下のように設定した．需要地数を 50, 施設配置の候補地数を 16 とし、施設の固定コスト f_j を設置場所 j にかかわらず 750, 設備の変動費 g_j を j にかかわらず 150 とした．設備の生産能力については文献 8) では連続値としているが、ここではより現実的な条件として、選択可能な設備の生産能力を 150, 300, 450 の 3 種類とした．

4.2 染色体表現

上述の確率的施設配置問題は、 x_j と v_j を 1 段階目の決定変数、 $y_{i,j}^k$ を 2 段階目の決定変数とする 2 段階問題となる．そして 1 段階目の決定変数 x_j, v_j が確定すれば、これに対して最適な 2 段階目の決定変数 $y_{i,j}^k$ の値は、単純な輸送問題を解くことで求められる．

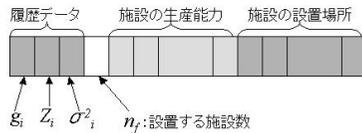


図 4 提案アルゴリズムでの染色体表現
Fig. 4 Chromosome in the proposed GA.

そこで、ここでは 1 段階目の決定変数のみを染色体としてコード化し、個体評価の際には輸送問題を解くことで 2 段階目の変数 $y_{i,j}^k$ の最適な値を求めた。

1 段階目の決定変数の染色体へのコード化については、図 4 に示す表現方法を採用した。染色体は大きく 4 つのブロックにより構成される。このうち最初のブロックは、Welch 検定に利用する個体の履歴情報を保持する 3 つの遺伝子により構成される。最初の遺伝子はその個体が何世代生き残っているかを表している。2 番目と 3 番目の遺伝子にはそれぞれ、その個体が過去の世代において評価を受けた $z'(S)$ の平均と分散が記録される。2 つ目のブロックは 1 つの遺伝子から構成され、設置する施設数 n_f を表す。第 3、第 4 のブロックは、ともに候補地の数と等しい数の遺伝子から構成される。それぞれ最初の n_f 個の遺伝子だけが意味を持ち、施設を設置する候補地と設置する施設の規模を表している。

4.3 シナリオ

無限シナリオのケースについては、需要 d_i^k 、製品価格 p_i^k 、製品の製造・輸送コスト $c_{i,j}^k$ が、それぞれ $30.0 \pm 10\%$ 、 $20.0 \pm 15\%$ 、 $15.0 \pm 25\%$ の一様分布に従うものとした。また有限シナリオのケースについては、それぞれの不確定パラメータの分布が上記の一様分布に従うとしたうえで、27 個のシナリオをモンテカルロサンプリングにより選び出した。

4.4 遺伝的操作とパラメータ設定

ルーレット選択を用いて、適合度 (Z_i) の高さ按比例した確率で 2 個体を選び出し、これらの個体に対して確率 0.8 で二点交叉を加え、新たな 2 個体を作り出す。また、新たに作られた個体に対して、確率 0.1 の一様変異を加える。数値実験では初期集団の個体数を 50 とし、この集団を 100 世代に達するまで進化させ、集団の中の 20 世代以上を生き残った個体の中から、最も Z_i が大きい個体を最良個体として採用した。

4.5 結 果

最良個体が淘汰される危険率 (Welch 検定を行う際の有意水準) と、サンプルするシナリオ数の設定を変えて、提案アルゴリズムを上述の施設配置問題に適用した。危険率の設定としては 10%、15%、20% の 3 種

類、サンプルするシナリオ数については 5、10 の 2 種類の、合計 6 種類のパラメータ設定の組合せについて数値実験を行った。また提案アルゴリズムとの比較のため、個体の評価・選択方法を以下に示す 2 つの方法に変更した場合についても数値実験を行った。

方法 1: アルゴリズムの実行に先立ち、モンテカルロサンプリングにより 100 個のシナリオを選び出し、つねにこのシナリオにおける平均適合度で個体を評価する。

方法 2: 個体評価のたびに新たに 10 個のシナリオを選び出し、このシナリオにおける適合度の平均で個体を評価する。

なお、文献 3)、7)、11) の研究は機会制約条件問題を対象としたものであるが、これらの研究で採用されている個体の評価・選択方法を、そのまま本論文で扱うリソース問題に適用したとすると、方法 1 は文献 3)、10)、11) で採用されている評価・選択方法に、方法 2 は文献 7) で採用されている評価・選択方法にそれぞれ相当する。

4.5.1 有限シナリオのケース

与えられた 27 のシナリオに対する最適な施設配置を、提案アルゴリズムと方法 1、2 の評価・選択方法を採用した場合について、それぞれ 10 回ずつ求めたところ、表 3 の結果となった。表 3 には 10 回の実行における期待利益の平均と分散、得られた期待利益の最大値と最小値、それに 1 回の実行に要した計算時間を示した。また、期待利益の最大値の行の括弧内の数字は、それぞれの方法を 10 回実行したときに、期待利益が 6321.4 となる解が得られた回数を表している。なお、与えられたシナリオの数が 100 以下となるこのケースでは、方法 1 の適用にあたっては、すべてのシナリオにおける適合度の平均で個体を評価した。

表 3 の結果を見ると、提案アルゴリズムを用いることで、よりサンプル誤差の影響の少ない個体選択が可能となっている。単純にサンプルした 10 個のシナリオで個体を評価する方法 2 を用いた場合には、すべてのシナリオで個体を評価する方法 1 を用いた場合と比べ、期待利益は平均で 0.27%、最悪ケースでは 1.09% 減少している。これに対して同じサンプル数でも、提案アルゴリズムを用いた場合の期待利益の減少はより小さく、たとえば危険率を 15% とした場合の期待利益の減少は、平均で 0.03%、最悪ケースでも 0.33% に抑えられている。また 10 回の実行で方法 2 とサンプル数を 10 とした提案アルゴリズムにより得られた期待利益の差を統計的に検定すると、危険率の設定が 10% の場合には有意水準 20% で、危険率の設

表 3 数値実験の結果 (有限シナリオケース)
Table 3 Result of numerical example (finite scenarios case).

| サンプル数 危険率 | 方法 1 | 方法 2 | 提案アルゴリズム | | | | | |
|--------------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 27 | 10 | 5 | 5 | 10 | 10 | 10 | 10 |
| | - | - | 10% | 15% | 20% | 10% | 15% | 20% |
| 平均 | 6321.4 | 6304.4 | 6305.4 | 6314.6 | 6312.2 | 6316.9 | 6319.3 | 6319.0 |
| 分散 | 0.0 | 675.1 | 366.3 | 241.9 | 262.8 | 89.9 | 44.3 | 56.8 |
| 最大 | 6321.4(10) | 6321.4(6) | 6321.4(5) | 6321.4(8) | 6321.4(7) | 6321.4(8) | 6321.4(9) | 6321.4(9) |
| 最小 | 6321.4 | 6252.4 | 6274.5 | 6274.5 | 6274.5 | 6297.6 | 6300.4 | 6297.6 |
| 計算時間 (秒) | 1,200 | 780 | 450 | 450 | 450 | 850 | 850 | 850 |

表 4 数値実験の結果 (無限シナリオケース)
Table 4 Result of numerical example (infinite scenarios case).

| サンプル数 危険率 | 方法 1 | 方法 2 | 提案アルゴリズム | | | | | |
|--------------|--------|----------|----------|--------|--------|--------|--------|--------|
| | 100 | 10 | 5 | 5 | 5 | 10 | 10 | 10 |
| | - | - | 10% | 15% | 20% | 10% | 15% | 20% |
| 平均 | 6253.2 | 5868.8 | 6255.1 | 6248.8 | 6242.7 | 6243.4 | 6245.3 | 6239.2 |
| 分散 | 220.5 | 127395.9 | 202.4 | 596.8 | 287.0 | 411.6 | 270.3 | 401.3 |
| 最大 | 6274.2 | 6237.8 | 6286.2 | 6296.8 | 6274.8 | 6274.0 | 6269.6 | 6262.1 |
| 最小 | 6228.2 | 5109.25 | 6236.2 | 6210.2 | 6219.6 | 6212.8 | 6213.5 | 6209.0 |
| 計算時間 (秒) | 4,430 | 780 | 450 | 450 | 450 | 850 | 850 | 850 |

定が 15%と 20%の場合には有意水準 15%で、その差は有意なものとなった。なお、各手法により得られた期待利益の分布が偏っており、母集団に正規性を仮定することができないことから、ここではノンパラメトリック検定の一種である Wilcoxon 検定により期待利益の差を検定した。Wilcoxon 検定とは得られたデータを順位に置き換えることで、2つの母集団の差を検定する手法である。

また、この数値実験では危険率を 10%とした場合を除き、サンプル数を 5 としても得られる解の精度はそれほど悪化しておらず、サンプル数を 5 とした提案アルゴリズムの計算効率が非常に高くなっている。サンプル数を 10 とした場合と比べ、サンプル数を 5 とすることで期待利益の平均は減少し、分散は増加する傾向が見られる。しかし、この差は大きなものではなく、Wilcoxon 検定を適用した結果では、危険率を 10%に設定した場合を除き、得られた期待利益の間に統計的に有意な差は認められない。一方でサンプル数を 5 とすることで、解を得るまでに要する計算時間は大幅に削減されている。また、サンプル数を 5 とした場合でも、平均的には提案アルゴリズムを 2 回実行することで、期待利益が 6321.4 となる配置案が求まっている。サンプル数を 5 とした場合には、方法 1 の約 1/3 の計算時間で解を求めることができることを考えると、平均的な性能で見れば、期待利益が 6321.4 となる配置案を得るという観点で評価しても、サンプル数を 5 とした提案アルゴリズムの方が方法 1 より効率的となっている。

なお、この数値実験では危険率を 10%、サンプル数を 5 に設定した場合、他の設定に比べて提案アルゴリズムにより得られる解の精度が悪化している。これはサンプル数を 5 とした場合には $z'(S)$ の分散が大きく、危険率の設定を 10%と厳しくした場合には、Welch 検定により淘汰の対象と判断される個体が非常に少なくなっていることから、新たに生成される個体が少なくなり、集団の進化が阻害されることが原因になっていると考えられる。

4.5.2 無限シナリオのケース

不確定パラメータが確率分布として与えられた無限シナリオのケースについて、提案アルゴリズムと方法 1, 2 の個体の評価・選択を用いた場合について、それぞれ 10 回ずつ最適な配置案を求めた。モンテカルロサンプリングにより選び出した 1,000 個のシナリオで、得られた配置案の期待利益を評価したところ、表 4 に示す結果となった。

この無限シナリオのケースでは、提案アルゴリズムの有効性がよりいっそう明らかとなっている。単純に 10 個のシナリオで個体を評価する方法 2 を用いた場合には、得られる解の期待利益のばらつきは非常に大きくなっている。これに対して提案アルゴリズムを用いた場合には、サンプル数を 5 とした場合でも、100 個のシナリオで個体を評価する方法 1 を用いた場合と同程度の期待利益の解を約 1/10 の計算時間で得ることができている。この場合についても Wilcoxon 検定を適用したところ、個体評価に方法 2 を用いた場合と提案アルゴリズムにより得られた期待利益の差は、い

ずれのサンプル数と危険率の設定においても有意水準1%で有意となった。また、この無限シナリオのケースでは、提案アルゴリズムのサンプル数の違いによる期待利益の差はまったく見受けられなかった。

5. おわりに

本論文では複雑な確率計画問題を効率的に解く手法として、サンプルした少数のシナリオを用いた個体の評価値を基に、統計的な検定により個体選択を行うGAを提案した。これまでもサンプルしたシナリオで個体を評価するGAの研究が行われている。しかし、これらの研究ではサンプルしたシナリオで個体を評価することの理論的妥当性は示されておらず、またサンプル数を少なくすると計算効率は高まるものの、得られる解の精度が悪化する問題があった。これに対し提案アルゴリズムでは、統計的検定(Welch検定)を用いて、サンプル誤差を考慮した個体選択を行うことで、少数のサンプルでも解の精度を損なうことなく効率的に解を探索することが可能となっている。本論文では代表的な確率計画問題を対象に数値実験を行ったが、現実の問題には、たとえばGAを用いないとリソースが計算できないような、個体評価により多くの計算を必要とする問題もある。こうした問題を解くうえで、提案アルゴリズムの有用性は非常に大きい。また、数値実験から明らかのように、提案アルゴリズムの性能は与えられたシナリオ数、サンプルするシナリオ数、Welch検定の有意水準により変化する。今後は提案アルゴリズムの個体の評価・選択に関するパラメータ(Welch検定に用いる有意水準と選び出すシナリオのサンプル数)について、これとアルゴリズムの収束性の関係を明らかにし、提案アルゴリズムを適用する際のパラメータ設定の目安を明らかにしたい。

謝辞 統計的検定手法に関して有益な助言をいただいた電力中央研究所の篠原上席研究員と、研究の全般にわたりご指導いただいた電力中央研究所の松井上席研究員に感謝の意を表します。

参 考 文 献

- 1) Chambers, L.(Ed.): *Practical Handbook of Genetic Algorithms: Applications Volume I*, CRC Press (1995).
- 2) Chambers, L.(Ed.): *Practical Handbook of Genetic Algorithms: New Frontiers Volume II*, CRC Press (1995).
- 3) Cieniawski, S.E., Eheart, J.W. and Ranjithan, S.: Using genetic algorithms to solve multiob-

jective groundwater monitoring problem, *Water Resources Research*, Vol.31, No.2, pp.399-409 (1995).

- 4) Fogel, D.B.: *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, IEEE Press (1999).
- 5) Kuehn, A.A. and Hamburger, M.J.: A Heuristic Program for Locating Warehouses, *Management Science*, Vol.9, pp.643-666 (1963).
- 6) Laporte, G., Louveaux, F.V. and Hamme, L.V.: Exact Solution to a Location Problem with Stochastic Demand, *Transportation Science*, Vol.28, No.2, pp.95-103 (1994).
- 7) Loughlin, D.H. and Ranjithan, S.: Chance-Constrained Genetic Algorithms, *Proc. Genetic and Evolutionary Computation Conference (GECCO-99)*, Vol.1, pp.369-376, Morgan Kaufmann, Orlando, Florida (1999).
- 8) Louveaux, F.V. and Peeters, D.: A dual-based Procedure for stochastic facility location, *Operations Research*, Vol.40, No.3, pp.564-573 (1992).
- 9) Michalewicz, Z. and Fogel, D.B.: *How to Solve It: Modern Heuristics*, Springer (1999).
- 10) Painton, L. and Chambell, J.: Genetic Algorithms in Optimization of System Reliability, *IEEE Trans. Reliability*, Vol.44, No.2, pp.172-178 (1995).
- 11) Ritzel, B.J., Eheart, J.W. and Ranjithan, S.: Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, *Water Resources Research*, Vol.30, No.5, pp.1589-1603 (1994).
- 12) Yoshitomi, Y., Ikenoue, H., Takeba, T. and Tomita, S.: Genetic Algorithm in Uncertain Environments for Solving Stochastic Programming Problem, *Journal of the Operations Research Society of Japan*, Vol.43, No.2, pp.266-290 (2000).

(平成 14 年 2 月 4 日受付)

(平成 14 年 4 月 10 日再受付)

(平成 14 年 6 月 4 日採録)



所 健一(正会員)

1987年慶應義塾大学理工学部管理工学科卒業。1989年同大学院理工学研究科修士課程修了。同年(財)電力中央研究所入所。現在、同研究所情報研究所主任研究員。電気学会、

日本 OR 学会各会員。