

遺伝的アルゴリズムにおける最適解出現確率に及ぼす突然変異の効果

張 玉安[†] 坂本 真人[†] 古谷 博史[†]

本研究では、最適解が得られない確率 F (失敗確率とよぶ) と GA パラメータの関係を解析した。GA における確率的揺らぎについて、積型適応度関数を用いて調べた。マルコフ過程理論とその近似である拡散方程式を用いて 1 次スキーマの確率分布を解析した。数値実験では、失敗確率と集団サイズ N 、交叉率 p_c 、突然変異率 p_m の関係を調べた。交叉率 $p_c = 1$ で突然変異のない場合、失敗確率は拡散方程式による予測とよく一致した。また、交叉と突然変異は失敗確率を大幅に減少させる働きをすることが分かった。

Probability of Obtaining Optimum Solutions in Genetic Algorithm and Roles of Mutation

YU-AN ZHANG,[†] MAKOTO SAKAMOTO[†] and HIROSHI FURUTANI[†]

We have studied the relationship between the failure probability F of obtaining no optimum solution and some GA parameters. We examined effects of stochastic fluctuation in a GA on the multiplicative landscape. We used Markov chain model and its diffusion approximation to calculate the distribution of the first order schemata. In numerical experiments, we studied the effects of the population size N , crossover rate p_c and mutation rate p_m on the failure probability F . We found in GA with $p_c = 1.0$ and $p_m = 0$ that the diffusion equation can reproduce the failure probability F . We also noticed the roles of crossover and mutation in greatly reducing F .

1. はじめに

遺伝的アルゴリズム (GA) を現実の問題に適用する際に問題となるのが、交叉率 p_c 、突然変異率 p_m 、集団の個体数 N などの選択である。しかし、これらのパラメータを決定するための信頼すべき理論はほとんどなく、我々は直観的、または経験によって選ぶしかない。そのとき注意することの一つは、最終的に最適解もしくは準最適解を得ることができるか否か、という点である。本研究では、失敗確率 F を、進化の最終段階において集団中に最適解が一つも得られない確率、として定義し、計算パラメータとの関係について調べた。そのため、確率的解析が比較的容易な積型適応度関数を用い、個体数 N が計算に与える影響を中心に、理論的解析と数値実験を比較しながら分析を行った。

有限集団の進化を数学的に解析するためには、確率論的枠組みが必要となる。集団遺伝学では、この問題を取り扱うため Wright や Fisher らによって開発さ

れたマルコフ過程理論が用いられ、Wright-Fisher モデルとよばれている¹⁾。しかし、Wright-Fisher モデルは解の解析的表現を得ることが難しく、単純な問題でさえ未だに完全な解が得られていない。そのため集団遺伝学の研究者は、拡散方程式を用いてマルコフ過程を近似し、数種類の問題についてはその解析解を得ることに成功した²⁾。

集団遺伝学では、1 遺伝子座の進化を考えることがほとんどで、GA のように多遺伝子座 (多ビット) の問題には直接適用することができない。したがって、これらの確率論的方法を GA の解析に応用する場合、なんらかの近似が必要となる。Asoh と Mühlenbein は、平らな適応度関数において集団が連鎖平衡にあると仮定し、GA の進化を 1 次スキーマを用いて確率論的に記述した³⁾。我々は彼らの確率論的なアプローチを積型適応度の問題に適用し、有限な N における 1 次スキーマの進化を調べた。集団が 1 次スキーマのみを用いて表現するためには、連鎖平衡にあることが必要である。交叉は集団を連鎖平衡に導く役割をするため、交叉が十分有効に働けば手段は連鎖平衡状態になる。また、積型適応度関数による進化では連鎖が弱く、そのため本研究ではこの適応度関数を採用した。

[†] 宮崎大学工学部
Faculty of Engineering, University of Miyazaki

2. 数学モデル

2.1 数学的表現

本研究では選択、交叉、および突然変異の過程を考慮し、計算の効率について集団サイズ N の影響を中心に調べた。選択は、適応度比例選択を用いた。集団は世代ごとに親と子が入れ替わるものとし、集団の個体数は世代によらず一定で N とする。個体の遺伝子型を固定長 ℓ の 2 進数列によって表す。したがって、遺伝子型の総数は $n = 2^\ell$ となる。整数 $i (0 \leq i \leq n-1)$ を 2 進ビット列と $i = (i(\ell), \dots, i(1))$ のように対応させる。

世代 t における遺伝子型 i の個体数を $N_i(t)$ とし

$$N = \sum_{i=0}^{n-1} N_i(t),$$

となる、遺伝子型 i の相対頻度を

$$x_i(t) = N_i(t)/N,$$

と表す。

集団の平均適応度は次式で表される

$$\bar{f}(t) = \sum_{j=0}^{n-1} f_j x_j(t). \quad (1)$$

f_i は遺伝子型 i の適応度である。我々は、積型適応度関数を

$$f_i = \prod_{k=1}^{\ell} \{1 + i(k)s\}, \quad (s \geq 0) \quad (2)$$

と定義する。パラメータ s は、選択の強さを表す。

2.2 連鎖平衡

連鎖は、集団内での異なる遺伝子座間の統計的相関を意味し、遺伝子座間に何らかの相関があるとき、その集団は連鎖不平衡状態にあるという⁴⁾。逆に相関がない状態を連鎖平衡とよぶ。連鎖不平衡は選択の過程で遺伝子座間に強い相関をもたらす (epistatic) 適応度を用いた場合、および遺伝的揺らぎの結果引き起こされる。交叉は連鎖不平衡の程度を減らすオペレーターとして働き、突然変異も同様な働きをする。

集団が連鎖平衡にあるならば、相対頻度 $x_i(t)$ は積型になり、次式で表される。

$$x_i(t) = \prod_{k=1}^{\ell} h_{i(k)}(t). \quad (3)$$

ここで $h_{i(k)}$ は k 番目のビットにおける 1 次スキーマの相対頻度を表す。式 (2) と式 (3) を式 (1) 代入すると、平均適応度も $i(k) = 1$ の 1 次スキーマ頻度の積で表される⁵⁾。ここで $i(k) = 1$ を $1(k)$ と表した。

$$\bar{f}(t) = \prod_{k=1}^{\ell} \{1 + sh_{1(k)}(t)\}. \quad (4)$$

さらにまた、位置依存性がなければ $h_1 = h_{1(k)}$ とし

$$\bar{f}(t) = \{1 + sh_1(t)\}^\ell \quad (5)$$

となる。

2.3 決定論的方程式

ここでは選択のための決定論的方程式を導く^{5),6)}。相対頻度 $x_i(t)$ を用いて適応度比例選択の過程を表現すると

$$x_i(t+1) = \frac{f_i x_i(t)}{\bar{f}(t)}, \quad (6)$$

となる。集団が連鎖平衡にあるならば、 $x_i(t)$ は次式で表される

$$x_i(t) = \prod_{k=1}^{\ell} h_{i(k)}(t).$$

これらの表現と方程式 (4) を用いて、進化方程式 (6) を書き直すと

$$\prod_{k=1}^{\ell} h_{i(k)}(t+1) = \prod_{k=1}^{\ell} \frac{(1 + si(k))h_{i(k)}(t)}{1 + sh_{1(k)}(t)} \quad (7)$$

となる。ここで $i(k) = 1$ とおき、1 次スキーマの決定論的進化方程式を導くことができる。

$$h_{1(k)}(t+1) = \frac{(1+s)h_{1(k)}(t)}{1 + sh_{1(k)}(t)}. \quad (8)$$

3. 確率論的モデル

3.1 Wright-Fisher モデル

決定論的方程式は個体数を無限大に近づけた場合に相当し、集団サイズ N が十分大きい場合には良い近似を与える。しかし、比較的少数の個体を扱う場合には適切でないことがある。そのため、Wright と Fisher は独立に、有限の個体数を考慮した進化モデルを提案した¹⁾。このモデルでは遺伝子座は一つ ($\ell = 1$) で、対立遺伝子を A および a と表す。遺伝子 A をビット 0、遺伝子 a をビット 1 に対応させる。遺伝子 a をもつ個体の数を N_1 とする。遺伝子 $i = 0, 1$ の適応度を次式で定義する。

$$f_i = \begin{cases} 1 & (i = 0) \\ 1 + s & (i = 1). \end{cases} \quad (9)$$

$s > 0$ とし、ビット 1 が有利な遺伝子である。

Wright-Fisher モデルの遷移確率 $P(j | i)$ はランダムサンプリングによる集団分布の変化を表すために用いられる⁷⁾。

$$P(j|i) = \binom{N}{j} b^j (1-b)^{N-j}, \quad (10)$$

$$b = \frac{(1+s)i}{(1+s)i + N - i}$$

ここで i は世代 t における遺伝子 a の数、 j は世代 $t+1$ における遺伝子 a の数とする。

集団が世代 t において $N_1 = i$ となる確率を $q_i(t)$ とする。集団の進化は次式で与えられる。

$$q_j(t+1) = \sum_{i=0}^N P(j|i) q_i(t) \quad (11)$$

規格化条件から

$$\sum_{i=0}^N q_i(t) = 1 \quad (12)$$

となる。遺伝子 a が集団中から失われる状態（消滅状態）は $i = 0$ 、逆に遺伝子 a が集団をすべて占める状態（固定状態）は $i = N$ に対応する。

3.2 拡散近似

一般に、進化方程式 (11) を解析的に解くことは非常に難しい。そのため研究者は Wright-Fisher モデル²⁾ に対する拡散近似を導入した。この場合、拡散方程式による近似はコルモゴロフ前向き方程式とよばれる偏微分方程式で表される。拡散方程式では、Wright-Fisher モデルにおける i/N を連続変数として近似し、さらに世代 t も連続量として扱われる。

さらに、確率過程理論では別のタイプの拡散方程式があり、コルモゴロフ後向き方程式と呼ばれる。この方程式は固定確率と消滅確率を導くために使用され、

$$\frac{\partial \psi(p, t)}{\partial t} = \frac{V(p)}{2} \frac{\partial^2 \psi(p, t)}{\partial p^2} + M(p) \frac{\partial \psi(p, t)}{\partial p}, \quad (13)$$

$$V(p) = \frac{p(1-p)}{N}, M(p) = sp(1-p)$$

と表される。ここで、変数 p は初期状態における遺伝子 a の相対頻度を表す。

我々は固定確率に関心があり、究極固定確率を次式で定義する。

$$u(p) = \lim_{t \rightarrow \infty} \psi(p, t) \quad (14)$$

方程式 (13) において $\partial \psi(p, t) / \partial t = 0$ とすると

$$\frac{d^2 u(p)}{dp^2} + \frac{2M(p)}{V(p)} \frac{du(p)}{dp} = 0, \quad (15)$$

となり、方程式境界条件は

$$u(0) = 0, \quad u(1) = 1, \quad (16)$$

で与えられる。 $2M/V = 2Ns$ から解は

$$u(p) = \frac{1 - \exp(-2Nsp)}{1 - \exp(-2Ns)}, \quad (17)$$

となる。

失敗確率 F を定常状態において集団中に最適解が一つも存在しない確率と定義する。失敗確率は $u(p)$ と ℓ により

$$F = 1 - u(p)^\ell \quad (18)$$

と表される。

4. 数値計算

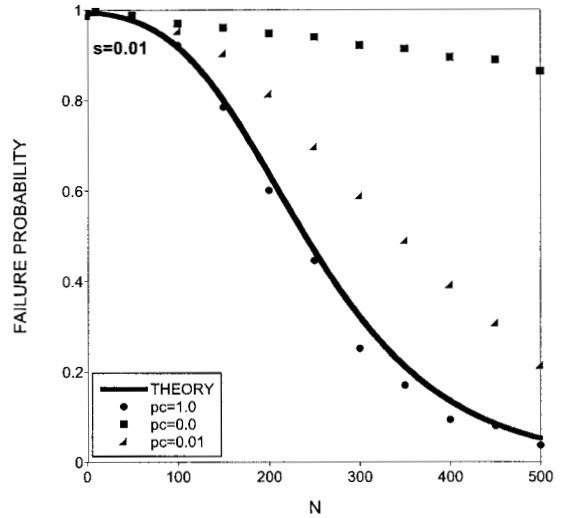


図1 失敗確率 F の N 依存性。弱い選択 ($s = 0.01$)。

数値計算では、適応度比例選択を用い、ビット長は $\ell = 8$ とした。交叉は一様交叉を用いた。1次スキーマの初期値は $p = 0.5$ とした。乱数を変えながら同じ計算を1000回行い、その平均値を求めた。異なる選択の強さ s を用いた2種類の計算を行い、それらの結果を比較した。ここでは、選択の強さ $s = 0.01$ と $s = 0.4$ を弱い選択および強い選択とよぶことにする。突然変異はとくに断らないかぎり $p_m = 0$ とした。

図1は弱い選択 ($s = 0.01$) の場合について、失敗確率 F の N 依存性を示したものである。実線は、式 (18) を用いた理論予測値である。理論値は、交叉率 $p_c = 1.0$ の計算結果とよく一致している。対照として交叉のない場合と弱い交叉 ($p_c = 0.01$) の結果も示した。失敗確率 F では、交叉が非常に重要な役割をしていることが分かる。

図2は強い選択 ($s = 0.4$) の場合について、集団サイズ $N = 20$ と 60 における失敗確率 F の p_c 依存性を示したものである。選択が強く集団サイズが小さい場合も、失敗確率は交叉に大きく依存する。 $N = 60$ では、 $p_c > 0.2$ において失敗確率 F がほとんどゼロになるため p_c に依存しなくなる。

図3は弱い選択 ($s = 0.001$) の場合について、集団サイズ $N = 200$ と 400 における突然変異の効果を示したものである。この図から、弱い選択で集団サイズ $N = 200$ の場合、 $p_m = 0.002$ から 0.003 の領域で

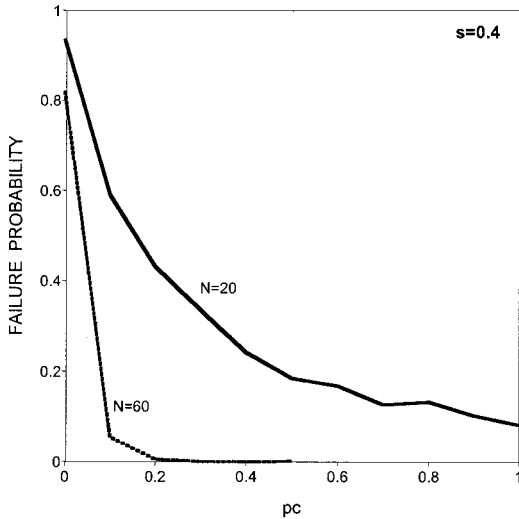


図2 強い選択 ($s = 0.4$). 集団サイズ $N=20,60$ における失敗確率 F の p_c 依存性.

最も良い値を得ることが分かる.

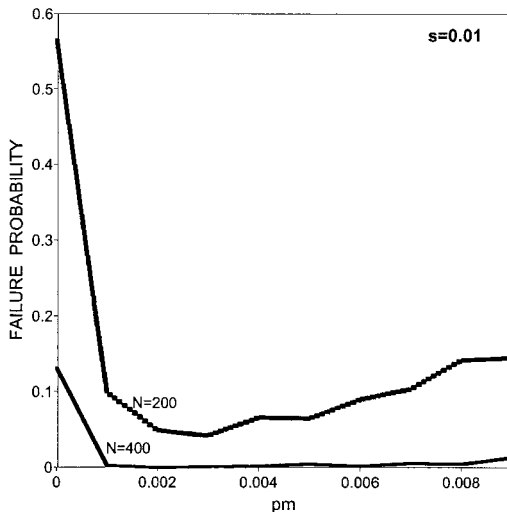


図3 弱い選択 ($s = 0.01$). 集団サイズ $N=200,400$ における失敗確率 F の p_m 依存性.

5. まとめと今後の課題

本研究では、積型適応度関数を対象に、GA の失敗確率を中心に解析を進めてきた。拡散方程式により、積型適応度関数を用いた GA の性質がよく説明できることが示された。

突然変異の無い場合、1 次スキーマに対する Wright-Fisher モデルは 2 つの吸収状態をもつマルコフ過程となる。もう一つの吸収状態は、有利な遺伝子が集団から消滅してしまう状態（望ましくない解）に対応する。失敗確率は、この消滅確率によって決定されるため、消滅確率を通じて N 依存性が現れる。

数値計算の結果から、交叉が失敗確率を減らす上で非常に重要な働きをしていることが分かった。弱い選択 $s = 0.01$ の場合、失敗確率は交叉により大きく変化するが弱い交叉率 p_c で十分である。強い選択 $s = 0.4$ の場合も、失敗確率は交叉により大きく変化する。しかし集団サイズ N が小さい場合は、大きい交叉率を用いる必要があることも分かった。

突然変異もまた GA に大きい効果をもたらす。現在その定量的解析を進めており、別の機会に報告したい。

参考文献

- 1) J. W. J. Ewens: *Mathematical Population Genetics. I. Theoretical Introduction*, Second Edition. Springer-Verlag, New York (2004)
- 2) J. F. Crow and M. Kimura: *An Introduction to Population Genetics Theory*. Harper and Row, New York (1970)
- 3) H. Asoh and H. Mühlenbein: On the Mean Convergence Time of Evolutionary Algorithms without Selection and Mutation. *Parallel Problem Solving from Nature, Lecture Notes in Computer Science*, **866**, Springer-Verlag, New York, (1994) 88–97
- 4) J. Maynard Smith: *Evolutionary Genetics*, 2nd edition, Oxford University Press, Oxford, (1998).
- 5) H. Furutani: Schema Analysis of Genetic Algorithms on Multiplicative Landscape, *Proceedings of the Simulated Evolution and Learning Conference, SEAL'02*, (2002) 230–235
- 6) H. Furutani: Schema Analysis of Average Fitness in Multiplicative Landscape. *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2003, Lecture Notes in Computer Science*, **2723**, Springer-Verlag, New York, (2003) 934–947
- 7) H. Furutani, M. Sakamoto and S. Katayama: Influence of Finite Population Size –Extinction of Favorable Schema, *Proceedings of the First International Conference on Natural Computation, Lecture Notes in Computer Science*, **3611**, (2005) 243–248