統計数理（1996）
第44巻 第1号 19-30

特集「計算統計学の発展」
［研究ノート］

遺伝的アルゴリズムとモンテカルロフィルタ

統計数理研究所　樋　口　知　之

（1995年4月　受付）

1. はじめに

遺伝的アルゴリズム（Genetic Algorithm，以後GAと略す）と通常呼ばれる方法は，自然淘汰と発生のメカニズムを模倣する最適化の一手法で，70年代初めにHolland（1975）によって提案され，主に彼の学生ら（Goldberg（1989））によって積極的に研究が進められた．この方法は，最適化に確率的規則を取り込む最近の流れ（茨木（1993））をくみ，Simulated Annealing法（以後，SAと略す．Kirkpatrick et al.（1983），Aarts and Korst（1989））とともに応用が盛んに行われている（Goldberg（1989），北野（1993），デービス(1994)）．これらのアルゴリズムを用いる主たる目的は，あくまでも，与えられた条件の下で目的関数を最大（あるいは最小）にする最適解の探索であり，本質的に組合せ的な複雑度が高く厳密な最適解を効率よく求めることが難しいような場合に特にその有効性が発揮される（Sen and Stoffa（1992），茨木（1993），デービス(1994)）．

GAやSAなどの方法を問題解決に必要とする背景には，問題対象の数値的表現に自然な未知変量の導入の結果，その未知変量の次元がコンピューターの能力の増大とあいまって高くなる傾向がある．データ解析の場合では，自在にデータを表現できる柔軟性に富むモデルの採用の傾向，つまり数多くのパラメーターを持つ大規模なモデルの採用への動きがこれに対応する．このような，大規模モデルにより対象を表現し，その最適なパラメーターを探索する最適化問題に共通した概念は，対象の固有の知識や経験を数値的に実現する形で，未知変量に対して常に確率分布を想定するベイズ統計学によって与えられる（Akaike（1980），田辺・田中（1983），北川（1993a），伊庭（1993））．

ベイズモデルの最も簡単な場合，つまりガウス型線形モデル（Lindley and Smith（1972），Harrison and Stevens（1976），Akaike（1980），West and Harrison（1989））は，応用対象の適切なparametrizationにより非常に有効であり，実際さまざまな分野の数々の問題に応用されている（Gersch and Kitagawa（1988）の参考文献，あるいは 北川（1993a）を参照）．ただ，線形・ガウスモデルでは取り扱うことができない，あるいは取り扱ったとしても*ad-hoc*な処理が全体の処理の中で重要な役割を果たす問題もある（Kitagawa（1987））．これらの問題を自然な形で取り扱う方法として，パラメーター間に一次元的な特殊な構造をもつような場合に限って，ガウス型線形ベイズ時系列モデルの拡張・一般化である非ガウス型非線形ベイズ時系列モデルを忠実に数値的に実現する方法が提案された（Kitagawa（1987），北川（1989））．ただこの方法による解法では，状態ベクトル（state vector）の次元が高々2～4の場合しか現実的でないことから，強い制約をモデリングに課し，問題対象の自在なparametrizationという観点からは満足できなかった．

最近，この制約を可能な限り取り除いたもとで，非ガウス型非線形ベイズ時系列モデルを解く方法が提案された（Kitagawa（1996））．このアルゴリズムは，分布表現を分布からの実現値でもって行っていることから，モンテカルロフィルタ（Monte Carlo Filter，以後MCFと略す）

と名付けられている．現在，実際の様々な時系列の問題への応用が可能かどうか，汎用性・有効性の観点からの萌芽的研究が進められている（北川（1993b），Higuchi (1995)）．実は，MCF のアルゴリズムはGA の枠組みとほぼ同じ構造をもつことが分かっている（樋口（1994））．MCF は，GA の基本的なコンセプト・イメージの援用ではなく，全く違う目的の為に構成された独立のものであるが，結果として両者に多くの類似点が認められる．本稿では，MCF をGA と対応づけながら解説することで，両者にみられる基本的なアルゴリズムについての共通点を明らかにし，そののち，GA がベイズ統計の枠組みで解釈できることを示す．

2. アルゴリズムの概説

本章ではまず，GA と MCF の説明を，両者の対応付けを念頭に置いて非常に簡単に行う．以下に説明するGA は，GA の根幹をなす手続きのみで構成されたもので，最適化を効率よく行う目的のために導入される modification（改良（茨木（1993））という言葉で表現されるような手続き）をいっさい含まない．詳しい具体的説明は，Goldberg (1989) による教科書本，あるいは安居院・長尾（1993），デービス（1994）等を参考にして頂きたい．

2.1 広義の遺伝的アルゴリズム

前述したようにGA の目的は，目的関数 $f(z)$ を最大（もしくは最小）にするような解 z^* の探索である．以後記述の簡略化のために，問題設定を**目的関数の最大化**にする．GA では z を，一個体を表す意味で *artificial creature* と呼ぶ．生物のDNAを意識して，全ての z の属性は適当に符号化され一本のひもに全情報がのっているとすることから，単に *string* と普通呼ぶ．アルゴリズムは，符号化の仕方によりそのperformanceにかなりの多様性をもつが（Whitley (1994)），それに関する議論は本稿の主目的にはずれるので，割愛する．

このアルゴリズムの特徴は，逐次的に最適解を求めていくアルゴリズムの各ステップにおいて，常に m 個の**最適解の候補**を保持する点である．その目的は，$f(z)$ の値がよく，かつ解の多様性も残るように最適解の候補の更新を行う点にある．これ以後，m 個の解各々を $\{z_i^{(n)}|i=1,\ldots,m\}$ と添え字をつけて，またこの集団 *population* を大文字で $Z^{(n)}$ と表す．上付き添え字の (n) は，第 n 番目の世代，つまり n 回の進化を経た最適解の候補であることを明示的に示す．

各ステップ（進化）は，以下に示す 3 つの手続きから構成されている．

1. Crossover または Crossing-over（交叉）
   (a) $Z^{(n-1)}$ から $m/2$ 組の *parent* 役をはたすペアをつくる．
   (b) 各ペアに対して Crossover と呼ばれる操作（染色体の交叉を模倣するような形式で行うため，このような名前が付けられている（Goldberg (1989)）．）を確率 P_C で施す（この結果，確率 $1-P_C$ で child は parent と全く同じになる）．
   (c) 次の世代 $z_i^{(n)}$（つまり *child*）を各ペアごとに 2 個ずつ *generate* する．
2. Mutation（突然変異）
   (a) 得られた各 $z_i^{(n)}$ に対し，確率 P_M で適当なランダムな変化を与える．
3. Selection または Reproduction（淘汰）
   (a) 各 $z_i^{(n)}$ の目的関数の値 $f(z_i^{(n)})$ を求める．
   (b) $z_i^{(n)}$ を

$$p(z_i^{(n)}) = \frac{f(z_i^{(n)})}{\sum_{i=1}^{m} f(z_i^{(n)})} \tag{2.1}$$

の確率で *resampling* することで，淘汰後の $Z^{(n)}$ を構成する．

4．1へもどる．

GA の MCF との対応付けを明確にするために，ステップ1－3の操作が施された直後の $Z^{(n)}$ を $Z_C^{(n)}$, $Z_M^{(n)}$, $Z_R^{(n)}$ と明示的に各々表しておく．したがって，ステップ1の操作で parent 役を務めるのは，$Z_R^{(n-1)}$ と表される．同様に，各 string $z_i^{(n)}$ を $z_{C,i}^{(n)}$, $z_{M,i}^{(n)}$, $z_{R,i}^{(n)}$ と表記する．

GA では，目的関数 $f(\cdot)$ を *evaluation function* あるいは *objective function* という．また $\sum_{i=1}^{m} f(z_i^{(n)})$ を *population fitness* と呼ぶ．この population fitness で規格化された evaluation function の値，$p(z_i^{(n)})$ のことを *fitness* と呼ぶ（Whitley (1994)）．自明であるが注意を促しておくと，$f(z_i^{(n)})$ は string $z_i^{(n)}$ 自身だけで定まる量であるが，fitness は集団を構成する他のメンバーに関連して相対的に定まる量である．前述の表記法にしたがうと，evaluation function の値は $f(z_{M,i}^{(n)})$，ステップ3の $p(z_i^{(n)})$ は，$p(z_{M,i}^{(n)})$ となる．

以上は，具体的な手続き操作の説明を省いた**広義の** GA に対する概説で，*simple genetic algorithm*（以後 SGA と呼ぶ）と呼ばれる標準的なものの紹介（安居院・長尾 (1993)）は，具体的な $z_i^{(n)}$ の符号化の説明の後行う．実際には，ステップ1－3でもって行われる操作と等価で，かつ計算時間が少なくてすむような工夫がなされたものが使われている．

ステップ3において $p(z_i^{(n)})$ が本来意図するような意味を持つためには，$f(z_i^{(n)})$ はすべて正値でなければならない．実際の適応にあたっては，全ての $f(z_i^{(n)})$ が正値になるような適当な変換を考え，変換後の $f(z_i^{(n)})$ を用いて $p(z_i^{(n)})$ を計算する．例えば，SA とのアナロジーで，

$$p(z_i^{(n)}) = \frac{\exp(f(z_i^{(n)})/T)}{\sum_{i=1}^{m} \exp(f(z_i^{(n)})/T)} \tag{2.2}$$

という方法も提案されている（Sen and Stoffa (1992)）．T は，*annealing temperature*．一意な最良の変換方法は存在せず，明らかに，変換方法に進化の仕方は大きく依存する．逆に言えば，各適用する問題に即してこの変換を工夫することが，最適化の効率を促進する意味での GA の改良となる．

2.2 Simple Genetic Algorithm (SGA)

SGA はもっとも基本的な操作だけで構成された GA で，*Canonical* GA とも呼ばれている (Whitley (1994))．$z_i^{(n)}$ はベクトルでもかまわないが，スカラーとしても SGA の本質的な部分の説明を損なわないので，簡便化のためにスカラーとする．

2.1.1 符号化

符号化の方法は，GA の最適化の performance を著しく左右するので，問題に応じた細心の注意と工夫が必要とされる（Goldberg (1989)）．まず，z は正整数の値をとるものとする．符号化には普通の2進表現を採用し，出現可能な値を充分表現できる bit 数 l を用意する（Stoffa and Sen (1991)）．

2.2.2 SGA の具体的手続き

1．Crossover または Crossing-over

(a) まず $Z_R^{(n-1)}$ からランダムに2つの string を選び出し（非復元抽出かどうかは問わない），$(z_{R,i}^{(n-1)}, z_{R,j}^{(n-1)})$ からなるペアを $m/2$ 個つくる．

(b) 次に，$U([0,1))$ の一様乱数を発生し，その値が P_C より大きいならば次の操作2へ進む．それ以下ならば，以下の操作を施す．

i．string の切断が生じる部分（*crossover site* と呼ばれる）を $1 \sim l-1$ の整数値からランダムに選ぶ．得た乱数を l_C とする $(1 \leq l_C \leq l-1)$．

ii. $z_{R,i}^{(n-1)}$ と $z_{R,j}^{(n-1)}$ の下位 l_C bit を交換する.

2. Mutation
   (a) すべての $z_{C,i}^{(n)}$ に対し, $U([0,1))$ の乱数を発生し, その値が P_M より大きいならば次の操作3へ進み, それ以下ならば以下の操作を施す.
      i. 突然変異がおこる部分を, $1 \sim l$ の整数値からランダムに選ぶ. 得た乱数を l_M とする $(1 \leq l_M \leq l)$.
      ii. l_M bit 目を, 反転(0なら1へ, 1なら0)させる.
3. Selection or Reproduction
   (a) 2.1節のステップ3の操作に同じ.
4. 1へもどる.

z がベクトルの時は, 各問題に即した配慮が必要(澤井(1993))だが, 普通各成分毎に string を形成し, 1, 2の操作を行えばよい.

2.3 モンテカルロフィルタ

2.3.1 一般化状態空間表現

ベイズ時系列モデルの一般型は, 次のような一般状態空間表現で表せる(北川(1989)).

$$\text{(2.3)} \qquad (\text{システムモデル}) \quad z^{(n)} = g(z^{(n-1)}, v^{(n)})$$

$$\text{(2.4)} \qquad (\text{観測モデル}) \quad y^{(n)} = h(z^{(n)}, w^{(n)})$$

ここで, $z^{(n)}$ は時刻 n の K 次元状態ベクトル, $y^{(n)}$ は時刻 n のスカラー観測値. $v^{(n)}$ は, 分布 $q(v)$ に従う L 次元 noise sequence である. また $w^{(n)}$ は, $r(w)$ の分布にしたがうスカラー noise sequence で, $w^{(n)} = h^{-1}(y^{(n)}, z^{(n)})$ で一意に定まるとする. さらに $y^{(n)}$ と $z^{(n)}$ の関数である $|\partial w^{(n)}/\partial y^{(n)}| = |\partial h^{-1}/\partial y^{(n)}|$ を $w'(y^{(n)}, z^{(n)})$ と記すことにする.

状態ベクトルの推定は, 一般的には次のような分布の漸化式で与えられる(Kitagawa (1987)).

$$\text{(2.5)} \qquad (\text{予測}) \quad p(z^{(n)}|Y^{(n-1)}) = \int p(z^{(n-1)}|Y^{(n-1)}) p(z^{(n)}|z^{(n-1)}) dz^{(n-1)}$$

$$\text{(2.6)} \qquad (\text{ろ波}) \quad p(z^{(n)}|Y^{(n)}) = \frac{p(y^{(n)}|z^{(n)}) p(z^{(n)}|Y^{(n-1)})}{\int p(y^{(n)}|z^{(n)}) p(z^{(n)}|Y^{(n-1)}) dz^{(n)}}$$

ここで $p(x|y)$ は, y が与えられたもとでの x の条件付き分布. また, $Y^{(n)} = [y^{(1)}, \ldots, y^{(n)}]$.

$g(\cdot)$ 及び $h(\cdot)$ が, 各々の変数に関して線形で, その上 $q(v)$ と $r(w)$ が共にガウスという特殊な場合は, 状態空間モデル(あるいは *Dynamic Linear Model* (Harrison and Stevens (1976), West and Harrison (1989)))と呼ばれ, 漸化式(2.5), (2.6)で得られる分布がすべてガウスになる. よって状態ベクトルの推定は, その平均値と分散だけを計算する, 一般に Kalman Filter と総称されるアルゴリズムで簡単に与えられる(Kalman (1960), 片山 (1983), 北川 (1993a)). 一方, 非ガウス型非線形のモデルの状態ベクトルの推定は, たとえ線形であったとしても非ガウス性から(2.5), (2.6)の積分を実際に行わなければならず, 単純な数値的積分で実用的なのは, 状態ベクトルが極めて低次元 $(K \leq 3)$ の場合のみである(Kitagawa (1987), 北川 (1993a)). MCF は K が比較的大きく(20以下)かつ, L が小さい場合($L = 1 \sim 5$)の解法手段として提案された.

MCF において, m 個の *particle* (GA では string に対応) $\{z_i^{(n)}|i = 1, \ldots, m\}$ は, $z^{(n)}$ の分布 $p(z^{(n)}|\cdot)$ からの実現値とする. つまり, $Z^{(n)}$ で $p(z^{(n)}|\cdot)$ を近似する. 従って, $z_i^{(n)}$ の出現確率

$p(z^{(n)}=z_i^{(n)}|\cdot)$ は i によらず，すべて $1/m$ である．

このアイデアは，気体運動論（リフシッツ・ピタエフスキー（1982））と呼ばれる気体（プラズマ）運動の統計的記述に現れる，**分布関数** $f(p,q,t)$ の時間発展を解くために，分布を $f(p,q,t)=1/m\sum_{i=1}^{m}\delta(z-z_i)$ とするアプローチ（*Klimontovich formalism* と呼ばれる（Klimontovich（1967）））と同じである．もちろん，$\delta(\cdot)$ はデルタ関数，z は一般化座標 q と一般化運動量 p で構成される位相空間座標を，また z_i は，その空間内での i 番目の particle の座標を示す．有限個 m の particle の運動を追跡し，その分布の時間発展で分布関数の時間的発展を推定する *particle simulation* と呼ばれる simulation も同じアイデアに基づいている．

2.3.2 アルゴリズム

MCF の各ステップ（進化）は，以下に示す 2 つの手続きから構成されている．

1．予測
- $z_i^{(n-1)}$ に対して以下の操作を行い，次の世代 $z_i^{(n)}$ を generate する．

(2.7)
$$z_i^{(n)}=g(z_i^{(n-1)},v_i^{(n)})$$

ここで，$v_i^{(n)}$ は分布関数 $q(v)$ から得られた乱数．後述のために，予測後の $z_i^{(n)}$ を，$z_{P,i}^{(n)}$ と記す．$g(\cdot)$ は，上述した（2.3）のシステムモデルに現れた関数．この操作で定まる $\{z_{P,i}^{(n)}|i=1,\ldots,m\}$ は，$\{z_i^{(n-1)}|i=1,\ldots,m\}$ と $\{v_i^{(n)}|i=1,\ldots,m\}$ が，$p(z^{(n-1)}|Y^{(n-1)})$ 及び $q(v)$ からおのおの発生した独立な sequence であるから，（2.5）のシステムモデルのプロセスで定義される予測分布，$p(z^{(n)}|Y^{(n-1)})$ からの実現値と考えられる．よって，$p(z^{(n)}=z_{P,i}^{(n)}|Y^{(n-1)})=1/m$．

2．ろ波
- （2.6）式の $p(y^{(n)}|z^{(n)})$ は，

(2.8)
$$\begin{aligned}p(y^{(n)}|z^{(n)})&=p(w^{(n)}|z^{(n)})\cdot\left|\frac{\partial w^{(n)}}{\partial y^{(n)}}\right|\\&=r(w^{(n)})w'(y^{(n)},z^{(n)})\end{aligned}$$

と書ける．これより，

(2.9)
$$\begin{aligned}p(z^{(n)}=z_{P,i}^{(n)}|Y^{(n)})&=\frac{r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})\cdot p(z^{(n)}=z_{P,i}^{(n)}|Y^{(n-1)})}{\sum_{i=1}^{m}r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})\cdot p(z^{(n)}=z_{P,i}^{(n)}|Y^{(n-1)})}\\&=\frac{r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})\cdot(1/m)}{\sum_{i=1}^{m}r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})\cdot(1/m)}\\&=\frac{r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})}{\sum_{i=1}^{m}r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})}\end{aligned}$$

が得られる．ここで，$w_i^{(n)}=h^{-1}(y^{(n)},z_{P,i}^{(n)})$ とした．
- $z_{P,i}^{(n)}$ を，$r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})/\sum_{i=1}^{m}r(w_i^{(n)})w'(y^{(n)},z_{P,i}^{(n)})$ の確率で resampling し，新しい m 個の $z_i^{(n)}$ を得る．これを，$z_{F,i}^{(n)}$ と記す．$\{z_{F,i}^{(n)}|i=1,\ldots,m\}$ は，$Y^{(n)}$ を観測したもとでの $z^{(n)}$ の分布 $p(z^{(n)}|Y^{(n)})$ からの実現値と見なせ，$p(z^{(n)}=z_{F,i}^{(n)}|Y^{(n-1)})=1/m$ となる．よって，式（2.7）の $z_i^{(n-1)}$ は，$z_{F,i}^{(n-1)}$ である．

3．1 へもどる．

実際の解析で多く使われるモデルでは，$w'(y^{(n)},z^{(n)})$ が $z^{(n)}$ に依存しない場合も多く，そのよ

うな場合 (2.9) 式中の $w'(y^{(n)}, z_{P,i}^{(n)})$ の項は分子，分母で打ち消しあい省略できる.

3. アルゴリズムの対応関係

3.1 MCF の GA の観点からの説明

3.1.1 予測

Crossover とMutation の一連の手続きを意味的に捉えれば，parent $Z_R^{(n-1)}$ から，次の世代，つまり child $Z_M^{(n)}$ への世代交代である．一方，予測の操作の内容も，$Z_F^{(n-1)}$ から，$Z_P^{(n)}$ への世代交代である．従って，予測の手続きはCrossover と Mutation によってなされる操作に意味的に対応している．Crossover と Mutation で加えられる randomness は，世代交代の時の情報伝達の際に生じる，**情報の揺らぎ**（あるいは乱れ）に相当する．MCF では，予測で加えられるシステムノイズがその役割を担っている.

3.1.2 ろ波

ろ波がSelection に完全に対応しているのは，(2.9) の $r(\cdot)w'(\cdot)/m$ を，(2.1) の evaluation function $f(\cdot)$ と置き換えれば明らかである．これにより，$(1/m)\sum r(\cdot)w'(\cdot)$ で近似される (2.9) の分母の $p(y^{(n)}|Y^{(n-1)})$ は，第 n 世代の population fitness と解釈できる．ただし MCF において evaluation function は，$r(\cdot)w'(\cdot)/m$ の関数形自体は不変だが，与えられるデータ $y^{(n)}$ のために n とともに変化する.

ろ波で行っている操作をGA の観点から説明すると，child $z_{P,i}^{(n)}$ を，与えられた**環境** $y^{(n)}$ への各child の適応度に応じて淘汰させ，淘汰後の parent の集団 $Z_F^{(n)}$ を作り出す手続きとなる.

淘汰のプロセスの主たる手続きである Resampling 法の明確な類別法はないが，大まかには 6 個にわけられ (Goldberg (1989))，関数の最大値を求めるいくつかの問題への適用を通して得られた数値的経験が報告されている (Brindle (1981))．MCF における Resampling 法としては，バイアス・分散の小さい尤度の推定値を与えるものが望ましい．いくつかの数値実験をもとに，SGA で普通採用されているルーレット法よりも，Stochastic Universal Sampling (Baker (1987)) と呼ばれる手続きのほうが，計算効率の面 (Whitley (1994)) も含めて望ましい結果を与えることが示されている (樋口 (1994)).

3.1.3 尤度

モデルの尤度 $p(Y^{(N)})$ は，次のように分解できる (Kitagawa (1987)).

$$
\begin{aligned}
p(Y^{(N)}) &= p(y^{(N)}, Y^{(N-1)}) \\
&= p(y^{(N)}|Y^{(N-1)})\cdot p(Y^{(N-1)}) \\
&= \prod_{n=1}^{N} p(y^{(n)}|Y^{(n-1)}) \\
&= \prod_{n=1}^{N} (\text{第 } n \text{ 世代の } \textit{population fitness}) \qquad (3.1) \\
&= \text{一族 } (\textit{family}) \text{ の } \textit{fitness} \qquad (3.2)
\end{aligned}
$$

よって，モデルの尤度とは，(3.1) で定義される *family fitness* の事である．$Y^{(N)}$ を環境という言葉に言い換えると，最尤法とは，**与えられた環境のもとで**，family fitness **を最大にするようなモデルの選択**と解釈できる．3.1.1–3.1.3 に示した対応のつくアナロジーが 樋口(1994) にまとめてある.

3.1.4 平滑化

データ $Y^{(n)}$ を観測したもとで，過去に溯って，$z^{(n')}$ $(n' < n)$ の分布，つまり $p(z^{(n')}|Y^{(n)})$ を求める手続きを平滑化という．MCF の枠組みでは，*Storing the State Vector* の方法と，*Two-filter Formula* による方法（Kitagawa (1994)）が提案されている（北川 (1993b), Kitagawa (1996)）．後者は，$z^{(n)}$ のとる値が離散である時の (2.3)，(2.4) に対応するモデル（**隠れマルコフモデル**と呼ばれる）において事後分布を求める手続き（普通 Baum-Welsh アルゴリズムと呼ばれる —— 正確には，Baum-Welsh アルゴリズムと forward-backward アルゴリズムを組み合わせた手続き ——）と本質的には同じである（Baum et al. (1970), Huang et al. (1990)）．ここでは，前者についてGA 的な観点から説明を行う．理解を容易にするために，string に対応する $z_i^{(n)}$ を第 n 世代の人間一個体として説明する．

説明のために，いくつかの変数を定義しておく．まず，平滑化の手続きで求められた，第 $n-1$ 世代の i 番目の人を，$z_{S,i}^{(n-1)}$ で示す．また，その人とその先祖がたどってきた履歴，意味的には家系図に相当する次のような量を

$$S_i^{(n-1)} = \left[z_{S,i}^{(1)}, z_{S,i}^{(2)}, \ldots, z_{S,i}^{(n-1)}\right] \tag{3.3}$$

で定義する．状態ベクトル $z^{(n)}$ を第 $n-1$ 世代まで並べた量を

$$X^{(n-1)} = \left[z^{(1)}, z^{(2)}, \ldots, z^{(n-1)}\right] \tag{3.4}$$

で表す．$X^{(n-1)} = S_i^{(n-1)}$ の意味を，

$$z^{(1)} = z_{S,i}^{(1)},\ z^{(2)} = z_{S,i}^{(2)}, \ldots, z^{(n-1)} = z_{S,i}^{(n-1)} \tag{3.5}$$

とする．$p(X^{(n-1)}|Y^{(n-1)})$ は $S_i^{(n-1)}$ $(i = 1,\ldots,m)$ で代表されているとする．従って，$p(X^{(n-1)} = S_i^{(n-1)}|Y^{(n-1)}) = 1/m$ である．

Smoothing by storing the state vector による方法は，以下の 2 つの手続きで構成されている．

1．新世代の誕生
   (a) $S_i^{(n-1)}$ に対して次の操作を行い，第 n 世代の人間 $t_i^{(n)}$ をつくる．

$$t_i^{(n)} = g(z_{S,i}^{(n-1)}, v_i^{(n)}) \tag{3.6}$$

   この操作は，予測と同じ．
   (b) 次のような量をつくる．

$$T_i^{(n)} = \left[z_{S,i}^{(1)}, z_{S,i}^{(2)}, \ldots, z_{S,i}^{(n-1)}, t_i^{(n)}\right] \tag{3.7}$$

   これは，i 番目の前世代までの家系図に，第 n 世代を加えた新しい家系図をつくることを意味する．$Y^{(n-1)}$ のもとでの $X^{(n)}$ の分布は，$T_i^{(n)} (i = 1,\ldots,m)$ で代表され，$p(X^{(n)} = T_i^{(n)}|Y^{(n-1)}) = 1/m$ である．

2．淘汰・再生産
   • 環境 $y^{(n)}$ のもとでの各家系の生存確率は，ろ波と同じような計算で次のように求められる．

$$p(X^{(n)} = T_i^{(n)}|Y^{(n)}) = \frac{p(y^{(n)}|T_i^{(n)}, Y^{(n-1)}) \cdot p(X^{(n)} = T_i^{(n)}|Y^{(n-1)})}{p(y^{(n)}|Y^{(n-1)})} \tag{3.8}$$

$$= \frac{p(y^{(n)}|t_i^{(n)}) \cdot p(X^{(n)} = T_i^{(n)}|Y^{(n-1)})}{p(y^{(n)}|Y^{(n-1)})}$$

$$= \frac{r(w_i^{(n)})w'(y^{(n)},t_i^{(n)}) \cdot (1/m)}{\sum_{i=1}^{m} r(w_i^{(n)})w'(y^{(n)},t_i^{(n)}) \cdot (1/m)}$$

ただし，$w_i^{(n)} = h^{-1}(y^{(n)},t_i^{(n)})$．また $w'(\cdot)$ は (2.8) 式と同じく，2.3.1で定義された関数である．

- $T_i^{(n)}$ を $p(X^{(n)} = T_i^{(n)}|Y^{(n)})$ の確率で resampling し，$S_i^{(n)}$ を選ぶ．この結果，$p(X^{(n)} = S_i^{(n)}|Y^{(n)}) = 1/m$ となる．この手続きは，2.3.2のろ波と同じ．

3．1に戻る．

この平滑化の方法を $p(z^{(n')}|Y^{(N)})$ を求めるのに適用するのは，$p(z^{(n')}|Y^{(n)})$ を近似する m 個の家系が n が大きくなると最後にはすべて同じ家系になってしまうため，あまり実用的ではない (Kitagawa (1996))．この理由は簡単である．家系が続くかどうかは，ステップ 1 (a)の操作で定まる第 n 世代の人間 $t_i^{(n)}$ の環境 $y^{(n)}$ への適応度ですべて決まってしまう．卑近なたとえになってしまうが，いかに先祖の家系が立派であろうと，第 n 世代 $t_i^{(n)}$ が，“不肖の息子的”に環境への適応が悪いと，この家系はこの世代で淘汰されて終わってしまう可能性が非常に高くなる．n が大きくなっても家系が続くためには，各世代でつねに環境への適応度が他の人間と比較して高い事が必要である．従って n の増大とともに，このような都合のいい対応ができる家系だけが再生産され増殖し続ける．

3.2 GA のベイズ的解釈

3.2.1 システムノイズ

parent から child への世代交代の時，MCF にはシステムモデルで記述される deterministic な成長・退化の効果があるが，一方 GA の方にはそのようなものはない．例えば具体的にシステムモデルを与える関数 $g(\cdot)$ として，次のようなシステムノイズが加法的に付加されるものを考える．

$$z^{(n)} = \phi(z^{(n-1)}) + v^{(n)} \tag{3.9}$$

$\phi(\cdot)$ が，世代交代の際に deterministic な影響を与える．ところが，GA にはそのような明示的な“世代間の駆動力”はない．アナロジーの説明の補足をしておくと，GA で *genetic drift* という，解の良し悪しと無関係に生じ結果として最適化の効率を下げるようなものがこの駆動力に対応する．

SGA の Crossover と Mutation によって与えられる情報の揺らぎを，MCF のシステムノイズの観点から考える．まず，簡単な Mutation の方から始める．

Mutation は，$z_{C,i}^{(n)}$ にノイズを加えて $z_{M,i}^{(n)}$ を生み出すプロセスなので，完全に MCF の予測の特殊なケースと考えられる．システムノイズの分布型の推察には，システムモデルの具体的な形が必要なので，MCF で最も簡単な $\phi(z^{(n-1)}) = z^{(n-1)}$ のランダムウォークモデルを考える．Mutation の場合 $v_i^{(n)}$ が $z_{C,i}^{(n)}$ に依存しているため，ある分布 $q(v)$ から生まれた乱数という形には書けない．しかし，$z_{C,i}^{(n)}$ と，操作の結果決まった $z_{M,i}^{(n)}$ を使って定められる $v_i^{(n)}$ の分布，$\hat{q}(v)$ の形は次のように明示的に表せる．

$$\hat{q}(v) = (1-P_M) \cdot \mu(v) + P_M \cdot \frac{1}{l} \sum_{j=1}^{l} \mu(|v|-2^{j-1}) \tag{3.10}$$

$\mu(x)$ は，$x=0$ のとき 1，それ以外では 0 の値をとる関数．$\hat{q}(v)$ は，P_M の 1 個のパラメーターで規定されている．

Crossover の場合は，一般には $\hat{q}(v)$ でさえ明示的に表現することは難しいが，次のような仮定のもとで得ることができる．

A1）l_C の取れる範囲を，$1 \le l_C \le l_{\max}$ に制限する．もちろん，$l_{\max} \le l-1$．

A2）parent 役を務める $z_{R,i}^{(n-1)}$ の $l_{\max}$ bit 以下で表現される値は，一様分布になっているとする．つまり，各 $z_{R,i}^{(n-1)}$ の $l_{\max}$ bit 以下の下位の bit で表現されてる数値を $x_i^{l_{max}}$ とすると，$P(x_i^{l_{\max}}) = 1/(2^{l_{\max}})$（ただし，$0 \le x_i^{l_{\max}} \le 2^{l_{\max}}-1$ の整数）と仮定する．

まず l_C が，具体的に $l_C = k$ の値をとった時の，$\hat{q}(v)$ の形を求める．今，$z_{R,i}^{(n-1)}$ の k bit 以下の下位の bit で表現される数値を x_i^k とする．$z_{R,i}^{(n-1)}$ が $z_{R,j}^{(n-1)}$ との Crossover によって受ける変化は，ランダムウォークモデルの定義から，$v_{ij} = x_j^k - x_i^k$ である．上の仮定より，x_i^k と x_j^k の分布はおのおの $0 \le x \le 2^k-1$ の範囲で一様であるから，v_{ij} の分布 $p(v_{ij})$ は，

$$(3.11)\qquad p(v_{ij}=d \mid l_C=k, P_C=1) = \frac{2^k-|d|}{2^{2k}} \quad (-(2^k-1) \le d \le 2^k-1 \text{ の整数})$$

となる．

l_C は，$1 \le k \le l_{\max}$ の間の整数値を一様に取るので，

$$(3.12)\qquad p(v_{ij}=d \mid P_C=1) = \frac{1}{l_{\max}} \sum_{k=l_d}^{l_{\max}} \frac{2^k-|d|}{2^{2k}} \quad (-(2^{l_{\max}}-1) \le d \le 2^{l_{\max}}-1 \text{ の整数})$$

となる．ここで，$l_d = \text{ceiling}(\log_2(|d|+1))$ で定義される値．ceiling (x) は，x 以上の最も小さい整数を与える関数（つまり天井関数）．ただし，$d=0$ の l_d は 1 とする．この式をさらに書き下すと

$$(3.13)\qquad p(v_{ij}=d \mid P_C=1) = \frac{1}{l_{\max}}\left(\frac{1}{2^{l_d-1}} - \frac{1}{2^{l_{\max}}} - \frac{|d|}{3}\left(\frac{1}{4^{l_d-1}} - \frac{1}{4^{l_{\max}}}\right)\right)$$

$$(-(2^{l_{\max}}-1) \le d \le 2^{l_{\max}}-1 \text{ の整数})$$

となる．従って，

$$(3.14)\qquad p(v_{ij}=d) = (1-P_C)\mu(d) + \frac{P_C}{l_{\max}}\left(\frac{1}{2^{l_d-1}} - \frac{1}{2^{l_{\max}}} - \frac{|d|}{3}\left(\frac{1}{4^{l_d-1}} - \frac{1}{4^{l_{\max}}}\right)\right)$$

$$(-(2^{l_{\max}}-1) \le d \le 2^{l_{\max}}-1 \text{ の整数})$$

が得られる．具体的に，$l_{\max}=7$, $P_C=1$ の時の，$p(v_{ij}=d)$ を図 1 に示す．

GA においては最適化の効率を規定する P_C や P_M の値をどのくらいにしたらよいのかが一つの問題である．問題ごとに *ad-hoc* に適当に与えるのではなく，P_C や P_M の値を動的に変化させつつ最適化を行うような，メタレベルで GA を適用する工夫もされている（Grefenstette (1986)）．MCF における P_C や P_M は，システムノイズの分布型 $q(v)$ を記述するのに必要なパラメーター，つまりベイズ統計の枠組みで hyperparameter と呼ぶものに相当する．

3.2.2 *fitness*

fitness のベイズ統計の観点からの意味づけは，3.1.2の説明によりすでに明らかである．MCF におけるGA の evaluation function に相当するものに，各ステップ毎に入ってくるデータ $y^{(n)}$ が，GA の evaluation function にはない．これをあえて MCF の枠組みで説明を行うと，GA

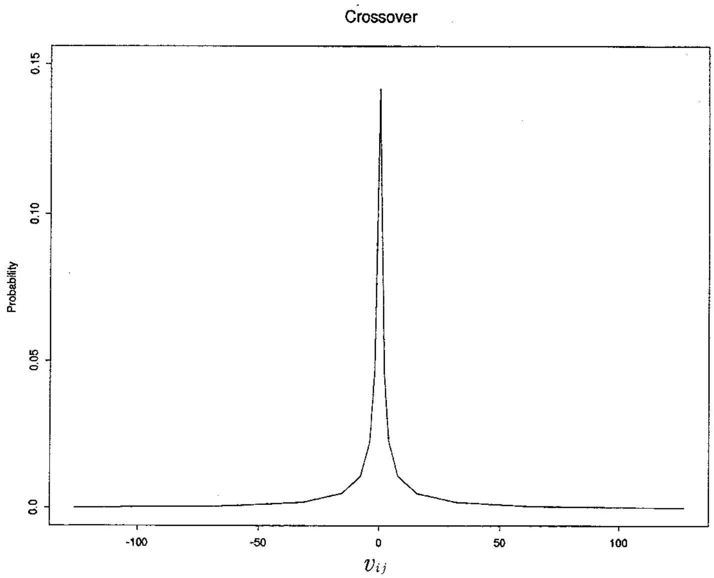

図 1. *Crossover* で与えられる操作を，MCF のシステムノイズで与えられる情報の揺らぎと捉えた時の，その分布.

では毎ステップ同じ環境（データ）が与えられたもとで淘汰の手続きを行っているものと解釈できる．つまり GA の淘汰は，毎ステップ同じデータを観測し，分布 $p(z^{(n)}|Y^{(n)})$ の近似を，$Z_R^{(n)}$ で行っていることになる．この観点からすると，最適化の効率を上げるために導入される evaluation function の関数型の変形，つまり *fitness scaling*（あるいは *fitness streching* と呼ばれる（安居院・長尾（1993）））が GA で果たす役割も明確になる．fitness scaling の一方法として提案された，(2.2) で与えられるような変形においての，コントロールパラメータ T の果たす役割も容易に理解できる．これらの変形はすべて，MCF の枠組みでは，観測ノイズの分布型 $r(w)$ を時間とともに変化させることと解釈できる．

4. まとめ

GA と MCF のアナロジーに焦点をあてることにより，両者の背後にあるアイデアを明確に理解できる．また，各々のアルゴリズムの拡張・改良が示唆できる．例えば，GA を MCF の観点からみると，システムノイズに相当する情報の揺らぎの与え方も，Crossover や Mutation だけでなく，そのさまざまな変形が考えうる．MCF の手続きも，予測の操作に GA の Crossover や Mutation の操作をとりこんだ工夫がありえる．実際にそれらの試みが筆者によってなされつつある（Higuchi（1995））．

参 考 文 献

Aarts, E. and Korst, J. (1989). *Simulated Annealing and Boltzmann Machine*, Wiley, Chichester.
安居院猛，長尾智晴（1993）．『ジェネティックアルゴリズム』，昭晃堂，東京.
Akaike, H. (1980). Likelihood and the Bayes procedure (with discussion), *Bayesian Statistics*, 143-165, University Press, Valencia, Spain.
Baker, J. (1987). Reducing bias and inefficiency in the selection algorithm, *Proceedings of the Second*

International Conference on Genetic Algorithms and Their Applications (ed. J. Grefenstette), 14-21, Erlbaum Associates, Hillsdale, New Jersey.
Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilitic functions of Markov chains, *Ann. Math. Statist.*, **41**, 164-171.
Brindle, A. (1981). Genetic algorithm for function optimization, Doctorial Dissertation, University of Alberta, Edmonton(unpublished).
デービス, L. 監修 (1994).『遺伝アルゴリズムハンドブック』(嘉数侑昇 他共訳), 森北出版, 東京.
Gersch, W. and Kitagawa, G. (1988). Smoothness priors in time series, *Bayesian Analysis of Time Series and Dynamic Models* (ed. J.C. Spall), 431-476, Marcel Dekker, New York and Basel.
Goldberg, D.E. (1989). *Genetic Algorithm in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Massachusetts.
Grefenstette, J.J. (1986). Optimization of control parameters for genetic algorithm, *IEEE Trans. Systems Man. Cybernet.*, **SMC-16**(1), 122-128.
Harrison, P.J. and Stevens, C.F. (1976). Bayesian forecasting (with discussion), *J. Roy. Statist. Soc. Ser. B*, **38**, 205-247.
樋口知之 (1994). モンテカルロフィルタにおけるランダムサンプリング, 情報処理学会研究報告 (情処研報) 94-HPC-54, **94**(108), 7-14.
Higuchi, T. (1995). Mutation and crossover operators in the genetic algorithm as a replacement for prediction in Kitagawa's Monte Carlo filter, Abstract of seasonal adjustment workshop, Bureau of the Census,Washington, D.C.
Holland, J.H. (1975). *Adaption in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor.
Huang, X.D., Ariki, Y. and Jack, M.A. (1990). *Hidden Markov Models for Speech Recognition*, Edinburgh University Press, Edinburgh.
伊庭幸人 (1993). ベイズ統計と統計物理, Research Memo., No. 492, The Institute of Statistical Mathematics, Tokyo.
茨木俊秀 (1993).『離散最適化法とアルゴリズム』, 岩波講座 応用数学 1 [方法 8], 岩波書店, 東京.
Kalman, R.E. (1960). A new approach to linear filtering and prediction problems, *Transactions of the ASME, Series D, Journal of Basic Engineering*, **82D**, 35-45.
片山 徹 (1983).『応用カルマンフィルタ』, 朝倉書店, 東京.
Kirkpatrick, S., Gelatt, C.D., Jr. and Vecchi, M.P. (1983). Optimization by simulated annealing, *Science*, **220**, 671-680.
Kitagawa, G. (1987). Non-Gaussian state space modeling of nonstationary time series (with discussion), *J. Amer. Statist. Assoc.*, **79**, 1032-1063.
北川源四郎 (1989). 非ガウス型時系列モデリング, オペレーションズ・リサーチ, **34**(10), 541-546.
北川源四郎 (1993a).『時系列解析プログラミング』, 岩波書店, 東京.
北川源四郎 (1993b). モンテカルロフィルタ, 統計数理研究所共同研究リポート, No. 49, 17-22.
Kitagawa, G. (1994). The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother, *Ann. Inst. Statist. Math.*, **46**(4), 605-623.
Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, *Journal of Computational and Graphical Statistics*, **5**(1), 1-25.
北野宏明 編 (1993).『遺伝的アルゴリズム』, 産業図書, 東京.
Klimontovich, Y.L. (1967). *The Statistical Theory of Non-equilibrium Processes in a Plasma* (ed. D. ter Haar; translated by H.S.H. Massey and O.M. Blunn), MIT Press, Cambridge, Massachusetts.
リフシッツ,イェ・エム, ピタエフスキー,エリ・ペ (1982).『物理的運動学I』(ランダウ＝リフシッツ理論物理学教程), 東京図書, 東京.
Lindley, D. V. and Smith, A.F.M. (1972). Bayes estimates of the linear model (with discussion), *J. Roy. Statist. Soc. Ser. B*, **34**, 1-41.
澤井秀文 (1993). 遺伝的アルゴリズムの画像復元への応用,『遺伝的アルゴリズム』(北野宏明 編), 第 8 章, 産業図書, 東京.
Sen, M.K. and Stoffa, P.L. (1992). Rapid sampling of model space using genetic algorithm: examples from seismic waveform inversion, *Geophysical Journal International*,**108**, 281-292.
Stoffa, P.L. and Sen, M.K. (1991). Nonlinear multiparameter optimization using genetic algorithms: inversion of plane wave seismograms, *Geophysics*, **56**(11), 1794-1810.
田辺國士, 田中輝雄 (1983). ベイズモデルによる曲線・曲面のあてはめ, 月刊 地球, **5**(3), 179-186.
West, M. and Harrison, P.J. (1989). *Bayesian Forecasting and Dynamic Model*, Springer, New York.
Whitley, D. (1994). A genetic algorithm tutorial, *Statistics and Computing*, **4**, 63-85.

Genetic Algorithm and Monte Carlo Filter

Tomoyuki Higuchi
(The Institute of Statistical Mathematics)

The "Genetic Algorithm" (GA) has been proposed to get an optimal solution for a nonlinear multiparameter optimization problem, and well applied to many inverse problems as well as the "Simulated Annealing" (SA). These methods are characterlized by using both a random walk in model space and a transition probability rule to help guide their search. Recently, the algorithm with a name of "Monte Carlo Filter" (MCF) has been presented for the generalized state space model. In this study, we investigate the relationships between GA and MCF. The major objective of this paper is to cast GA into the Bayesian framework by its interpretation from the viewpoint of MCF.

Key words: Genetic algorithm, Monte Carlo filter, generalized state space model, dynamic linear model, Bayesian approach.