Int. J. Open Problems Compt. Math., Vol. 4, No. 3, September 2011
ISSN 1998-6262;
www.i-csrs.org

Why Are Clonal Selection Algorithms MCMC?

K.E. El-Nady, U.H. Abou El-Enien, and A. A. Badr

Department of Math, Faculty of Science, Alexandria University, Alexandria, Egypt
e-mail:pr_dr_kh_elnady@yahoo.com
Department of Math, Faculty of Science, Alexandria University, Alexandria, Egypt
e-mail:ossama.hanafy77@gmail.com
Department of Computer Science, Faculty of Computers and Information,
Cairo University, Giza, Egypt
e-mail:pr.dr.a.badr@gmail.com

Abstract

Clonal selection algorithms are considered. Two algorithms are designed and executed to obtain purely empirical analysis conclusions in order to turn to purely theoretical analysis results about the behavior of clonal selection algorithms as Markov chains, which confirm the conjectures from these experiments and in order to introduce a complete framework toward a new philosophy of MCMC method and of statistical inference method about Markov chains. First, we model clonal selection algorithms using Markov chains. Second, we carry on a particle analysis and analyze the convergence properties of these algorithms. Third, we propose the unified MCMC theorem and unique chromosomes method for a purely successful optimization of these algorithms.

Keywords: *Clonal selection algorithms; MCMC; Classification; central limit theorem; Stationary multivariate normal distribution; Unique chromosomes*

1. Introduction

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution.

There are works in the literature that attempt to introduce an architecture for the construction of artificial immune systems (e.g., Hofmeyr and Forrest [15]), a physical model for it (e.g., Zak [23]).

Genetic algorithms within the framework of Markov chains have been studied by (e.g., Eiben et al. [1], Fogel [12], Davis et al. [8], Rudolph [22], Nix et al. [21], De Jong et al. [10], and Spears et al. [11]).

Clark et al. [5] have introduced a Markov chain model of the B-cell algorithm. The convergence of immune algorithms has been studied by (e.g., Villalobos-Arias [6]and Cutello et al. [7]. There are works that attempt to contribute for AIS (e.g.,Hon et al. [17]and Timmis et al. [16]).

These works do not develop a unified stochastic model for genetic and clonal selection algorithms, analyze their performance and convergence properties to provide a general framework.

There are works that attempt to combine three areas (simulation, optimization, and analyzing sets of data)(e.g., Conley [24]).

The rest of the paper is organized as follows. In Section 2, we give the formulation of the problem. In Section 3, we state the main result. Then in Section 4, the proof of the main result is given in nine steps. In Section 5, we propose the two algorithms. In Section 6, we give two numerical examples. In Section 7, we give some concluding remarks. In Section 8, we we give some open problems.

2. Formulation of the problem

In this paper, we consider a problem, namely: Why are clonal selection algorithms MCMC?.

Throughout this paper, we consider any objective real valued function of n-variables $f(x_1, x_2, ..., x_n)$, where $a_i \leq x_i \leq b_i$ for $i = 1, 2, ..., n$ are domains of each variable x_i and a_i and b_i are real numbers, de Castro and Von Zuben [9]proposed a clonal selection algorithm, named CLONALG.

Proposition 2.1. We restrict an arbitrary uncountable set S=$\{a_i \leq x_i \leq b_i$ for $i = 1, 2, ..., n\}$ to be a subset of n-space $\mathbf{R}^n$ as a sample space, restrict an arbitrary countable set T to be set of all $(x_1, x_2, ..., x_n)$ in S=$\{a_i \leq x_i \leq b_i\}$ for which $P(x_1, x_2, ..., x_n) > 0$ as a sample space (see Apostol [3]).

Proposition 2.2. We restrict an arbitrary countable set U to be set of all possible simple random samples with replacement in the i^{th} trail

$$D_y^{(i)} = (\ (x_{11}, x_{12}, \ldots, x_{1n}), (x_{21}, x_{22}, \ldots, x_{2n}), \ldots, (x_{m1}, x_{m2}, \ldots, x_{mn})\ )$$

for which

$$P((x_{11}, x_{12}, \ldots, x_{1n}), (x_{21}, x_{22}, \ldots, x_{2n}), \ldots, (x_{m1}, x_{m2}, \ldots, x_{mn})) > 0$$

as a sample space (see [20] and [18]), where for each $(x_{j1}, x_{j2}, \ldots, x_{jn})$ in a common sample space T, $j = 1, 2, \ldots, m$ and $y = 1, 2, \ldots, h$, where m = a sample size and h = number of samples

3. Main result

In this section, we shall state the main theorem.

Theorem 3.1. For any clonal selection algorithm, the following holds:

(1) The sequence of $(m \times n)$-dimensional random matrices $\mathbf{X}_0$, $\mathbf{X}_1$, $\mathbf{X}_2$, ... converges in distribution to the $(m \times n)$-dimensional random matrix $\mathbf{X}$(has unique stationary distribution) if and only if for each $(a_1, a_2, \ldots, a_{(m\times n)})$

$$\phi_{\mathbf{X}_i}(a_1, a_2, \ldots, a_{(m\times n)}) \rightarrow \phi_{\mathbf{X}}(a_1, a_2, \ldots, a_{(m\times n)}) \text{ as } i \rightarrow \infty$$

, uniqueness of characteristic function (see Goldman [14]), where

n = number of variables

m = number of sets of measurements on n variables = sample size = number of chromosomes

$\phi_{\mathbf{X}}(a_1, a_2, \ldots, a_{(m\times n)})$ is the characteristic function of $\mathbf{X}$ of $m \times n$ real variables.
(a) If $\mathbf{P}$ is a transition matrix for any clonal selection algorithm (regular chain) and the probability vector t is a fixed point of the matrix $\mathbf{P}$, then

$$\sum_{j=1}^{2^{(k\times m)}} \sum_{i=1}^{2^{(k\times m)}} t_i P_{ij}^n \rightarrow (\sum_{j=1}^{2^{(k\times m)}} t_j = 1) \text{ as } n \rightarrow \infty.$$

(b) For any clonal selection algorithm, a real valued function $f(x_1, x_2, ..., x_n)$, where $a_i \leq x_i \leq b_i$ for $i = 1, 2, ..., n$ is one that contains an infinite number of Markov chains(every chain has different Unique chromosomes for purely successful optimization and has different globally optimum value(s)).
(2) The sequence of n-dimensional random vectors $\overline{\mathbf{X}}_0, \overline{\mathbf{X}}_1, \overline{\mathbf{X}}_2, \ldots$ converges in distribution to the n-dimensional random vector $\overline{\mathbf{X}}$(has unique stationary multivariate normal distribution) if and only if for each $(a_1, a_2, \ldots, a_n)$

$$\phi_{\overline{\mathbf{X}}_i}(a_1, a_2, \ldots, a_n) \rightarrow \phi_{\overline{\mathbf{X}}}(a_1, a_2, \ldots, a_n) \text{ as } i \rightarrow \infty,$$

uniqueness of characteristic function, where

n = number of variables.

$\phi_{\overline{\mathbf{X}}}(a_1, a_2, \ldots, a_n)$ is the characteristic function of $\overline{\mathbf{X}}$ of n real variables.

(a) All possible conditional multivariate normal distributions of transition probability matrix have the following form: (see proof of the main theorem)

The conditional distribution of $\overline{\mathbf{X}}_{q+1}$, given that $\overline{\mathbf{X}}_q = L_\star$, is multivariate normal and has

Mean $= \mu_\star = \mu + \Sigma_{t_1 t_2}\Sigma^{-1}(L_\star - \mu)$

and

Covariance $= \Sigma_\star = \Sigma - \Sigma_{t_1 t_2}\Sigma^{-1}\Sigma_{t_2 t_1}$.

The distribution of $\overline{\mathbf{X}}$ is stationary multivariate normal distribution.
(b) For any clonal selection algorithm, a real valued function $f(x_1, x_2, ..., x_n)$, where $a_i \leq x_i \leq b_i$ for $i = 1, 2, ..., n$ is one that contains an infinite number of lumped Markov chains(every chain has different Unique chromosomes for purely successful optimization and has different globally optimum value(s)).

4. Proof of the main result

In this section, we prove the main result in Theorem 3.1. We start with a useful theorem.

Theorem 4.1. Let (S, β, P) be a probability space and let T denote the set of all x in S for which $P(x) > 0$. Then T is countable. (see [3])

We shall prove Theorem 3.1 in nine steps.

Proof of Theorem 3.1. Step 1. For clonal selecton algorithms, we define a probability space(S, β, P).

For clonal selecton algorithms, let a real valued function $f(x_1, x_2, ..., x_n)$, where $a_i \leq x_i \leq b_i$ for $i = 1, 2, ..., n$. we restrict an arbitrary uncountable set S to be a subset of n-space $\mathbf{R}^n$ as a sample space, we shall assume that this set is a Borel set. The Borel subsets of S themselves form a Boolean σ-algebra β.

A nonnegative completely additive set function P defined on β with $P(S)$=1 is called a probability measure. We have a probability space (S, β, P).

Step 2. We prove that T is a countable subset of S, and define n-dimensional random variable defined on a sample space T.

Proof By Theorem 4.1, we restrict an arbitrary countable set T to be set of all $(x_1, x_2, ..., x_n)$ in S for which $P(x_1, x_2, ..., x_n) > 0$ as a sample space. T denotes a countable subset of S (whenever we use a set T in n-space as a sample space, we shall assume that this set is a Borel set).

Let a sample space T of possible solutions to be coded as strings of k bits {0,1} and let each possible configuration have a fitness $f_i, i = 1, .., 2^k$, let $f^\star$ be the globally optimum value. Hence T is countable.

Let T be a set in n-space for some $n \geq 1$ and if τ consists of all subsets of T, the probability function P is completely determined on τ. We have a probability space (T, τ, P).

Next, we define an n-dimensional random variable defined on a sample space T.

Let $\mathbf{X}$ be an n-dimensional random variable defined on a sample space T.

$$\mathbf{X} = (X_1(x_1) = x_1, X_2(x_2) = x_2, \ ..., \ X_n(x_n) = x_n)' = (x_1, x_2, ..., x_n)$$

for each $(x_1, x_2, ..., x_n)$ in T.

We will use notation x_{jk} to indicate the particular value of the k^{th} variable that is observed on the j^{th} item, or trail. That is, x_{jk} = measurement (commonly called data) of the k^{th} variable on the j^{th} item consequently, m measurements on n variables can be displayed as a rectangular array. We need to make assumptions about the variables whose observed values constitute the data set.

Step 3. We prove that U is a countable set of all possible simple random samples with replacement in the i^{th} trail, and define an $(m \times n)$-dimensional random variable defined on a sample space U.

Proof. Let the jk^{th} entry in the data matrix be the random variables X_{jk}.Each set of measurements $\mathbf{X}_j$ on n variables is a random vector, and we have the random matrix $\mathbf{X}$. A random sample can now be defined. $\mathbf{X}_1$, $\mathbf{X}_2$, ...,$\mathbf{X}_j$, ..., $\mathbf{X}_m$ form a random sample if their joint probability mass function is given by the product $P(\mathbf{X}_1)P(\mathbf{X}_2)$ ... $P(\mathbf{X}_j)$ ... $P(\mathbf{X}_m)$, where $P(\mathbf{X}_j) = P(x_{j1}, x_{j2}, \ldots, x_{jk}, \ldots, x_{jn})$ is the probability mass function for the j^{th} row vector.

We restrict an arbitrary countable set U to be set of all possible simple random samples with replacement in the i^{th} trail

$$D_y^{(i)} = (\ (x_{11}, x_{12}, \ldots, x_{1n}), (x_{21}, x_{22}, \ldots, x_{2n}), \ldots, (x_{m1}, x_{m2}, \ldots, x_{mn})\ )$$

for which

$$P((x_{11}, x_{12}, \ldots, x_{1n}), (x_{21}, x_{22}, \ldots, x_{2n}), \ldots, (x_{m1}, x_{m2}, \ldots, x_{mn})) > 0$$

as a sample space, where for each $(x_{j1}, x_{j2}, \ldots, x_{jn})$ in a common sample space T, $y = 1, 2, \ldots, h(h =$ number of samples) and $j = 1, 2, \ldots, m(m =$ a sample size.

All possible simple random samples with replacement in the i^{th} trail can be defined by every possible configuration of an entire population of m bit strings. There are $h = 2^{k \times m}$ such samples. Hence U is countable.

Let U be a set in $(m \times n)$-space for some $(m \times n) \geq (2 \times 1 = 2)$ and if ν consists of all subsets of U, the probability function P is completely determined on ν. We have a probability space (U, ν, P).

Next, we define an $(m \times n)$-dimensional random variable defined on a sample space U .

Let $\mathbf{X}$ be an $(m \times n)$-dimensional random variable defined on a sample space U.

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m)$$

$$= (\ (x_{11}, x_{12}, \ldots, x_{1n}), (x_{21}, x_{22}, \ldots, x_{2n}), \ldots, (x_{m1}, x_{m2}, \ldots, x_{mn})\ )$$

for each

$$(\ (x_{11}, x_{12}, \ldots, x_{1n}), (x_{21}, x_{22}, \ldots, x_{2n}), \ldots, (x_{m1}, x_{m2}, \ldots, x_{mn})\ )$$

in U, where $\mathbf{X}_j = (x_{j1}, x_{j2}, \ldots, x_{jn})$ for each $(x_{j1}, x_{j2}, \ldots, x_{jn})$ in T.

Step 4. We prove that U_1 is a countable set of all possible repeated dependent two trails without replacement, define an $2(m \times n)$-dimensional random variable defined on a sample space U_1, and describe Markov chain process.

Proof. We restrict an arbitrary countable set U_1 to be set of all possible repeated dependent two trails without replacement $(D_r^{(1)} \cap D_s^{(2)})$ for which $P(D_r^{(1)} \cap D_s^{(2)}) \geq 0$ as a sample space, where for each $D_y^{(i)}$ in a common sample space U, $y = r$ or s, $r = 1, 2, \ldots, h$ and $s = 1, 2, \ldots, h$. Each conditional probability can be defined and can be obtained by the equation

$$P(D_s^{(2)} | D_r^{(1)}) = \frac{P(D_r^{(1)} \cap D_s^{(2)})}{P(D_r^{(1)})} \geq 0 \text{ such that } P(D_r) \neq 0.$$

There are $2(2^{k\times m})$ such repeated dependent two trails without replacement. Hence U_1 is countable.

Let U_1 be a set in $2(m \times n)$-space for some $2(m \times n) \geq (2(2\times 1) = 4)$ and if ν_1 consists of all subsets of U_1, the probability function P is completely determined on ν_1. We have a probability space (U_1, ν_1, P).

We define an $2(m \times n)$-dimensional random variable defined on a sample space U_1.

Consider $q \in Q$ where Q is the discrete parameter space of the Markov chain process $\{\mathbf{X}_q, q = 0, 1, 2, \ldots\}$(parameter homogeneous).

Let $\mathbf{X}_q$ be an $(m \times n)$-dimensional random variable defined on a sample space U during q (initial value of time parameter), and $\mathbf{X}_{q+1}$ be an $(m \times n)$-dimensional random variable defined on a sample space U during $q + 1$. The possible outcomes for $\mathbf{X}_q$ are $D_1, D_2, \ldots, D_h$, and the same holds for $\mathbf{X}_{q+1}$.

Let $\mathbf{X}$ be an $2(m \times n)$-dimensional random variable defined on a sample space U_1.

$$\mathbf{X} = (\mathbf{X}_q, \mathbf{X}_{q+1}) = (\mathbf{X}_q = D_r, \mathbf{X}_{q+1} = D_S) = (D_r, D_s) = (\mathbf{X}_{b_1}, \mathbf{X}_{b_2})$$

$$=((\mathbf{X}_{b_1 1},\mathbf{X}_{b_1 2},\ldots,\mathbf{X}_{b_1 m}), (\mathbf{X}_{b_2 1},\mathbf{X}_{b_2 2},\ldots,\mathbf{X}_{b_2 m}))$$

for each (D_r, D_s) in U_1, where

$$\mathbf{X}_v = D_y = (\mathbf{X}_{b1}, \mathbf{X}_{b2}, \ldots, \mathbf{X}_{bm}) = \mathbf{X}_b$$

$$=((x_{(b1)1},x_{(b1)2},\ldots,x_{(b1)n}), (x_{(b2)1},x_{(b2)2},\ldots,x_{(b2)n}),\ldots,(x_{(bm)1},x_{(bm)2},\ldots,x_{(bm)n}))$$

for each

$$((x_{(b1)1},x_{(b1)2},\ldots,x_{(b1)n}), (x_{(b2)1},x_{(b2)2},\ldots,x_{(b2)n}),\ldots,(x_{(bm)1},x_{(bm)2},\ldots,x_{(bm)n}))$$

in U, $v = q$ or $q + 1$, $y = r$ or s and $b = b_1$ or b_2.

Next, we describe Markov chain process.

A Markov chains requires a finite collection of states, denoted by $U = \{D_1, D_2, \ldots, D_h\}$, an $h \times h$ probability matrix $\mathbf{P}$ is called a transition matrix and a probability vector $\mathbf{p}^{(0)} = (\ P(D_1), P(D_2), \ldots, P(D_h)\ ) = (p_1^{(0)}, p_2^{(0)}, \ldots, p_h^{(0)})$ is called the initial probability vector. We can also interpret $\mathbf{p}^{(0)}$ as stationary distribution.

Step 5. We prove that clonal selecton algorithms are regular chains.

Proof. From Step 2, We get unique chromosomes $(= 2^k)$. From Step 3 and Step 4, we generate all possible combinations of states of unique chromosomes $(= 2^{(k\times m)})$ and give each state a number.

Now we recall a theorem in [19].

Theorem 4.2. If $\mathbf{P}$ is a transition matrix for ergodic chain, then:

(1) There is a unique probability vector fixed point $\mathbf{t} : \mathbf{t} = \mathbf{tP}$.
(2) All components of $\mathbf{t}$ are positive.
(3) If $h_j^{(n)}$ is the average number of times the process is in state j in the first n steps, then for any $\epsilon > 0$,

$$P(\mid h_j^{(n)} - t_j \mid > \epsilon) \to 0 \text{ as } n \to \infty$$

no matter what the starting state is.

We apply clonal selecton algorithms on each generated state for n-iterations, where n is a large number. We count for each state with specific number the number of times it appeared, calculate the probability of each state (P), where

$$P = \frac{\text{number of times it appeared}}{n}$$

and get the stationary distribution ordered by the state number and its probability P. All states in this chain will be ergodic.

We conclude and have that all stationary distributions ($= 2^{(k\times m)}$) are the same unique probability vector fixed point. Hence clonal selecton algorithms are ergodic chains.

We take each generated state and its chain, get the position numbers sequences for each unique state (of $2^{k\times m}$) in the chain, starting with index zero in the initial state. We conclude and have that all sequences of positions are even positions or odd positions or random positions sequences. Hence clonal selecton algorithms are regular chains.

Step 6. We describe the partition L of all possible combinations of states of unique chromosomes, prove that L is a countable set and define an n-dimensional random variable defined on a sample space L.

Let $L = \{L_1, L_2, \ldots, L_g\}$ be a partition of all possible combinations of states of unique chromosomes. We get L_z for each state in combination and combine states according their equal L_z's ($(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$'s)-Taking into account that L_z's are unique for each set of states.

We prove that L is a countable set.

We restrict an arbitrary countable set L to be set of all possible outcomes in the i^{th} trail $L_z^{(i)} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ for which $P(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n) > 0$ as a sample space, where $z = 1, 2, \ldots, g$

$$\bar{x}_k = \frac{x_{1k}+x_{2k}+\ldots+x_{mk}}{m} = \sum_{j=1}^{m} \frac{x_{jk}}{m} \text{ for } k = 1, 2, \ldots, n$$

and x_{jk} = measurement of the k^{th} variable on the j^{th} item. Hence L is a countable.

Let L be a set in n-space for some $n \geq 1$ and if ζ consists of all subsets of L, the probability function P is completely determined on ζ. We have a probability space (L, ζ, P).

Next, we define an n-dimensional random variable defined on a sample space L.

Let $\overline{\mathbf{X}}$ be an n-dimensional random variable defined on a sample space L. For each $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ in L,

$$\overline{\mathbf{X}} = (\overline{X}_1(\bar{x}_1) = \bar{x}_1, \overline{X}_2(\bar{x}_2) = \bar{x}_2, \ldots, \overline{X}_n(\bar{x_n}) = \bar{x}_n)$$

$$= (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n).$$

Step 7. We prove that L_1 is set of all possible repeated dependent two trails without replacement, define an $2n$-dimensional random variable defined on a sample space L_1 and describe lumped Markov chains process.

Proof. We restrict an arbitrary countable set L_1 to be set of all possible repeated dependent two trails without replacement $(L_e^{(1)} \cap L_d^{(2)})$ for which $P(L_e^{(1)} \cap L_d^{(2)}) \geq 0$ as a sample space, where for each $L_z^{(i)}$ in a common sample space L, $z = e$ or d, $e = 1, 2, \ldots, g$ and $d = 1, 2, \ldots, g$. Each conditional probability can be defined and can be obtained by the equation

$$P(L_d^{(2)}|L_e^{(1)}) = \frac{P(L_e^{(1)} \cap L_d^{(2)})}{P(L_e^{(1)})} \geq 0 \text{ such that } P(L_e) \neq 0.$$

Hence L_1 is countable.

Let L_1 be a set in $2n$-space for some $2n \geq (2(1) = 2)$ and if ζ_1 consists of all subsets of L_1, the probability function P is completely determined on ζ_1. We have a probability space (L_1, ζ_1, P).

We define an $2n$-dimensional random variable defined on a sample space L_1.

Consider $q \in Q$ where Q is the discrete parameter space of the Markov chain process $\{\overline{\mathbf{X}}_q, q = 0, 1, 2, \ldots\}$(parameter homogeneous).

Let $\overline{\mathbf{X}}_q$ be an (n)-dimensional random variable defined on a sample space L during q (initial value of time parameter), and $\overline{\mathbf{X}}_{q+1}$ be an (n)-dimensional random variable defined on a sample space L during $q + 1$. The possible outcomes for $\overline{\mathbf{X}}_q$ are $L_1, L_2, \ldots, L_g$, and the same holds for $\overline{\mathbf{X}}_{q+1}$.

Let $\overline{\mathbf{X}}$ be an $2n$-dimensional random variable defined on a sample space L_1.

$$\overline{\mathbf{X}} = (\overline{\mathbf{X}}_q, \overline{\mathbf{X}}_{q+1}) = (\overline{\mathbf{X}}_q = L_e, \overline{\mathbf{X}}_{q+1} = L_d) = (L_e, L_d) = (\overline{\mathbf{X}}_{t_1}, \overline{\mathbf{X}}_{t_2})$$

$= ((\overline{X}_{t_1 1}, \overline{X}_{t_1 2}, \ldots, \overline{X}_{t_1 n}),(\overline{X}_{t_2 1}, \overline{X}_{t_2 2}, \ldots, \overline{X}_{t_2 n}))$

for each (L_e, L_d) in L_1, where $\overline{\mathbf{X}}_v = L_z = \overline{\mathbf{X}}_t$

$= (\overline{X}_{t1}(\bar{x}_{t1}) = \bar{x}_{t1}, \overline{X}_{t2}(\bar{x}_{t2}) = \bar{x}_{t2}, \ldots, \overline{X}_{tn}(\bar{x}_{tn}) = \bar{x}_{tn})$

$= (\bar{x}_{t1}, \bar{x}_{t2}, \ldots, \bar{x}_{tn})$ for each $(\bar{x}_{t1}, \bar{x}_{t2}, \ldots, \bar{x}_{tn})$ in L, $v = q$ or $q+1$, $z = e$ or d and $t = t_1$ or t_2.

Next, we describe the lumped Markov chains process.

A Markov chains requires a finite collection of states, denoted by $L = \{L_1, L_2, \ldots, L_g\}$, an $g \times g$ probability matrix $\mathbf{P}$ is called a transition matrix and a probability vector $\mathbf{p}^{(0)} = (\ P(L_1),\ P(L_2),\ \ldots,\ P(L_g)\ ) = (p_1^{(0)}, p_2^{(0)}, \ldots, p_g^{(0)})$ is called the initial probability vector. We can also interpret $\mathbf{p}^{(0)}$ as stationary distribution.

Step 8. By Theorems for the mean and covariance of the sampling distribution of $\overline{\mathbf{X}}$ and central limit Theorem (see[18], [20]), we conclude and have the following modified forms of the Theorems for an n-dimensional random variable $\overline{\mathbf{X}}$ defined on a sample space L.

Theorem 4.3. If $D_1, D_2, \ldots, D_h$ represent all possible simple random samples with replacement from a common joint probability mass function $P(x_1, x_2, ..., x_n)$ (the parent population, whatever its form, have a mean $\mu = (\mu_1, \mu_2, ..., \mu_n)'$ and a finite covariance Σ_1), then an n-dimensional density for the random vector $\overline{\mathbf{X}} = (\overline{X}_1, \overline{X}_2, ..., \overline{X}_n)'$(sampling distribution of $\overline{\mathbf{X}}$) $r(\overline{\mathbf{X}})$ is centered around the population mean, regardless of sample size ($E(\overline{\mathbf{X}}) = \mu$).

Theorem 4.4. Each covariance $\sigma_{ik},\ i, k = 1, 2, \ldots, n$ of the distribution of $\overline{\mathbf{X}}$ decreases with increasing sample size; that is, the distribution of $\overline{\mathbf{X}}$ becomes more concentrated around the population mean as the sample size gets larger (a covariance of $\overline{\mathbf{X}}$ is denoted by $\text{cov}(\overline{\mathbf{X}}) = \frac{\Sigma_1}{m}$, where m (number of chromosomes) is a sample size.

For practical problem, often $m = 100 =$ chromosomes and bit $= k > 50$; thus there may be more than 2^{5000} possible states in the chain(see [13]). The distribution of $\overline{\mathbf{X}}$ becomes more symmetrical as the sample size gets larger and is approximately $N_n(\mu, \Sigma = \frac{\Sigma_1}{m})$ for large sample size. An n-dimensional normal density for the random vector $\overline{\mathbf{X}}$ has the form

$$f(\overline{\mathbf{x}}) = \frac{1}{(2\Pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{\frac{-(\overline{\mathbf{X}}-\mu)'\Sigma^{-1}(\overline{\mathbf{X}}-\mu)}{2}} \text{ where } -\infty < \bar{x}_i < \infty,\ i = 1, 2, \ldots, n.$$

Theorem 4.5. Let $\mathbf{X}_1,\ \mathbf{X}_2,\ \ldots,\ \mathbf{X}_m$ be independent observations from any

population with mean μ and finite covariance Σ_1. Then $\overline{\mathbf{X}}$ has an approximate $N_n(\mu, \frac{\Sigma_1}{m})$ distribution for large sample sizes. Here m should also be large relative to n. The approximation provided by the central limit theorem applies to discrete, as well as continuous, multivariate populations.

The modified Theorems apply to sampling of finite populations if the sampling fraction is 5 percent or smaller (the sample size m is small relative to the population size M; that is, with fraction small) (see [20]).

Step 9. We prove that clonal selecton algorithms are lumped Markov chains.

Proof. From Step 2, we get unique chromosomes $(= 2^k)$. From Step 3 and Step 4, we generate all possible combinations of states of unique chromosomes$(= 2^{(k\times m)})$ and give each state a number.

From Step 6, we get L_z for each state in combination and combine states according to their equal L_z's - Taking into account that L_z's are unique for each set of states. Let each set of states to be a separate new state. From Step 6 and Step 7, we identify new state space of L_z's based on set of states of unique L_z's, define $L = \{L_1, L_2, \ldots, L_g\}$ and define a random variable $\overline{\mathbf{X}}$ on state space L.

From Step 8, we get sampling distribution of $\overline{\mathbf{X}}$ $(N_n(\mu, \frac{\Sigma_1}{m} = \Sigma))$, get expectation of sampling distribution $E(\overline{\mathbf{X}}) = \mu$, where μ is the expectation of parent population and get covariance matrix of $\overline{\mathbf{X}}$ $(= \Sigma = \frac{\Sigma_1}{m})$, where $m =$ sample size and $\Sigma_1 =$ covariance matrix of parent population. We conclude and have that the distribution of $\overline{\mathbf{X}}_q$ is $N_n(\mu, \Sigma)$, and the same holds for $\overline{\mathbf{X}}_{q+1}$.

Let $\overline{\mathbf{X}} = (\overline{\mathbf{X}}_q, \overline{\mathbf{X}}_{q+1})'$ be distributed as $N_{2n}(\mu^\star, \Sigma^\star)$ with

$$\mu^\star = (\mu = \mu_{t_1}, \mu = \mu_{t_2})', \; \Sigma^\star = \begin{pmatrix} \Sigma_{t_1t_1} = \Sigma & \Sigma_{t_1t_2} \\ \Sigma_{t_2t_1} & \Sigma_{t_2t_2} = \Sigma \end{pmatrix},$$

and $\mid \Sigma_{t_2t_2} \mid > 0$. Then the conditional distribution of $\overline{\mathbf{X}}_{q+1}$, given that $\overline{\mathbf{X}}_q = L_\star$, is normal and has

$$\text{Mean} = \mu_\star = \mu + \Sigma_{t_1t_2}\Sigma^{-1}(L_\star - \mu)$$

and

$$\text{Covariance} = \Sigma_\star = \Sigma - \Sigma_{t_1t_2}\Sigma^{-1}\Sigma_{t_2t_1}.$$

We select one state from U, get $L_\star$ for the selected state and replicate the selected state $2^{(k\times m)}$ times.

We apply clonal selecton algorithms for one transition (iteration) only on the replicated state to produce new states.

On the generated new states, we compute L_z. On the set of L_z's computed, we compute conditional normal distribution $N(\mu_\star, \Sigma_\star)$. We compute $\Sigma_{t_1t_2}$, where $\mu_\star = \mu + \Sigma_{t_1t_2}\Sigma^{-1}(L_\star - \mu)$.

By a similar argument, we compute $\Sigma_{t_1t_2}$ for all states of L, have that $\Sigma_{t_1t_2}$ is the same and have that $\Sigma_\star$ is the same. Hence clonal selecton algorithms are lumped Markov chains. We compute all possible conditional normal distributions of transition matrix and can also interpret the distribution of $\overline{\mathbf{X}}$ as stationary normal distribution.

On the basis of Steps 1-9, we complete the proof of Theorem 3.1.

5. Proposed algorithms

We prepared programs by using MATLAB 7.5. Afterwards, two algorithms are designed and executed to observe the characteristics and to know the behavior of the following:

5.1. Clonal selecton algorithms (as Markov chains) We named the first proposed algorithm regular optimization analysis (ROA), the basic steps of the ROA algorithm are as follows:

1. Input number of bits k.
2. Get unique chromosomes $= 2^k$.
3. Input number of chromosomes m.
4. Get number of states $= 2^{(k\times m)}$.
5. Generate all possible combinations of states of unique chromosomes (= also $2^{(k\times m)}$).
6. Give each state a number.
7. Pick one state randomly.
8. Apply any clonal selection algorithm on the state for n-iterations, where n is a large number.
9. Count for each state with specific number the number of times it appeared.
10. Calculate the probability of each state (P), where

$$P = \frac{\text{number of times it appeared}}{n}.$$

11. Get the stationary distribution ordered by the state number and its probability P.
12. Taking the randomly chosen state and its chain. Get the position numbers sequences for each unique state (of $2^{k\times m}$) in the chain, starting with index zero in the initial state.

Check if all sequences of positions are either

(A) Even positions or

(B) Odd positions only then the chain is cyclic else if also random positions sequences appear then the chain is regular.

5.2. Clonal selection algorithms (as lumped Markov chains) We named the second proposed algorithm Central MCMC optimization analysis(C-MCMC-OA),the basic steps of the C-MCMC-OA algorithm are as follows:

1. Input number of bits k.
2. Get unique chromosomes $= 2^k$.
3. Input number of chromosomes m.
4. Get number of states $= 2^{(k\times m)}$.
5. Get all possible combinations of states of unique chromosomes $= U$ (= also $2^{(k\times m)}$).
6. Get L_z for each state in combination.
7. Combine states according to their equal L_z's - Taking into account that L_z's are unique for each set of states.
8. Let each set of states to be a separate new state.
9. Identify new state space of L_z's based on set of states of unique L_z's.
10. Define $L = \{L_1, L_2, \ldots, L_g\}$.
11. Define a random variable $\overline{\mathbf{X}}$ on state space L.
12. Get sampling distribution of $\overline{\mathbf{X}}$ $(N_n(\mu, \frac{\Sigma_1}{m} = \Sigma))$.
13. Get expectation of sampling distribution $E(\overline{\mathbf{X}}) = \mu$, where μ is the expectation of parent population.
14. Get covariance matrix of $\overline{\mathbf{X}}$ $(= \Sigma = \frac{\Sigma_1}{m})$, where m = sample size and Σ_1 = covariance matrix of parent population.
15. Let$(\overline{\mathbf{X}}_q, \overline{\mathbf{X}}_{q+1})'$be distributed as $(N_{2n}(\mu^\star, \Sigma^\star))$ with

$$\mu^\star = (\mu = \mu_{t_1}, \mu = \mu_{t_2})',\ \Sigma^\star = \begin{pmatrix} \Sigma_{t_1t_1} = \Sigma & \Sigma_{t_1t_2} \\ \Sigma_{t_2t_1} & \Sigma_{t_2t_2} = \Sigma \end{pmatrix},$$

and $\mid \Sigma_{t_2t_2} \mid > 0$, where q is the initial value of time parameter Q of the Markov chain process $\{\overline{\mathbf{X}}_q, q = 0, 1, 2, \ldots\}$ that the possible outcomes for $\overline{\mathbf{X}}_q$ are $L_1, L_2, \ldots, L_g$ and the same holds for $\overline{\mathbf{X}}_{q+1}$.
16. Get $E(\overline{\mathbf{X}}_q) = \mu$.
17. Get covariance matrix of $\overline{\mathbf{X}}_q = \Sigma$.
18. Get $E(\overline{\mathbf{X}}_{q+1}) = \mu$.
19. Get covariance matrix of $\overline{\mathbf{X}}_{q+1} = \Sigma$.
20. Group states on U such that

(1) states are globally optimal (all members of states are identical).

(2) All or Some members of states have globally optimum values.

(3) states are not globally optimal and all members of states are identical.
(4) All or Some members of states have not globally optimum values.

21. Select one state randomly from each group on U.
22. Get $L_\star$ for the selected states.
23. Replicate each selected state (from each group) $2^{(k\times m)}$ times.
24. Apply any clonal selection algorithm for one transition (iteration) only on each replicated selected state to produce new states.
25. On the generated new states, compute L_z.
26. On the set of L_z's computed, compute conditional normal distribution $N(\mu_\star, \Sigma_\star)$.
27. Compute $\Sigma_{t_1t_2}$, where $\mu_\star = \mu + \Sigma_{t_1t_2}\Sigma^{-1}(L_\star - \mu)$.
28. check that $\Sigma_\star$ is the same for all four groups.
29. If $\Sigma_{t_1t_2}$ for the four groups of states are equal then the process is Markov chain.
30. Substitute all possible L_z's to compute all conditional expected values $\mu_\star$. As a result this computes all possible conditional normal distributions and The distribution of $\overline{\mathbf{X}}$ is stationary multivariate normal distribution.

6. Numerical results and discussion

6.1. A numerical example for clonal selection algorithms (as Markov chains)

Consider the following function: $f(x) = x \cdot \sin(10\Pi \cdot x) + 1$, $x \in [1.7, 2]$

if $k = 2$ bits, $m = 6$ chromosomes, $n = 70000$ iterations, then

Unique chromosomes = { 00 = 1.700000, 01 = 1.800000, **10 = 1.900000**, 11 = 2.000000 } = $2^k = 2^2$

and the globally optimum value = 1.900000.

All possible combinations of states of unique chromosomes =

{ 0:(00, 00, 00, 00, 00, 00), 1:(00, 00, 00, 00, 00, 01),

..., 4095:(11, 11, 11, 11, 11, 11) } = $2^{(k\times m)} = 2^{(2\times 6)} = 2^{12}$

Pick one state randomly (state 162), apply CLONALG (for mutation see [25])on the state for $n = 70000$ and probability of mutation = 0.9, get

162, 3835, 3315, ..., 63, 424, 1403

,and then get stationary distribution ordered by the state number

(0, 1, 2, . . . , 4093, 4094, 4095) =

(0.000029, 0.000014, 0.000014, . . . , 0.000129, 0.000129, 0.004600).

Taking the randomly chosen state 162 and its chain. Get the position numbers sequences for each unique state (of $2^{2\times 6}$) in the chain, starting with index zero in the initial state. We conclude and have that all sequences of positions are even positions, odd positions or random positions sequences then the chain is regular.

162 to 3835(neither even nor odd)
1-351-403-1044-1207-1848-2850-3163-3819-4103-4122-4525-4529-4534-4538-4540-4592-4655-4750-5080-
⋮
-53679-53689-53755-55371-55373-55632-56582-56700-56706-56747-56986-56988-57315-57389-57586-58119-58279-58284-58630-58691-58965-59043-59337-59401-59848-61050-61489-62044-62509-63199-63228-63415-63552-64330-64438-64492-64960-65090-65130-65428-65542-66059-66496-66958-67009-67239-68180-68315-68887-69312-69318-69320-69529
⋮
162 to 1396(odd)
67639
162 to 641(even)
68110
162 to 1216(odd)
68627
162 to 500(odd)
68697
162 to 2446(odd)
69883

Consider the same function. Pick one state randomly (state 380), apply CLONALG on the state for $n = 90000$ and probability of mutation = 0.9, get

380, 884, 951, . . . , 2020, 1197, 3213

,and then get stationary distribution ordered by the state number

(0, 1, 2, . . . , 4093, 4094, 4095) =

(0.000122, 0.000022, 0.000011, ..., 0.000089, 0.000144, 0.004889).

Taking the randomly chosen state 380 and its chain. Get the position numbers sequences for each unique state (of $2^{2\times 6}$) in the chain, starting with index zero in the initial state. We conclude and have that all sequences of positions are even positions, odd positions or random positions sequences then the chain is regular.

380 to 951(neither even nor odd)
2-563-4782-13911-14366-15871-16205-16207-20896-21951-27264-27376-28417-31728-34596-36192-37499-38747-41289-44987-49454-49618-52106-52501-55824-57879-58447-59695-60540-63336-64743-65171-68952-70796-78909-80822-84734-85339-85441-85473-86243-89029-89496
⋮
380 to 3226(odd)
88367
380 to 3142(odd)
89085
380 to 1922(even)
89498
380 to 2657(even)
89650

From chains of states 162 and 380, we obtain the same unique stationary distribution, and have regular chain.

6.2. A numerical example for clonal selection algorithms (as lumped Markov chains)

Consider the same function, unique chromosomes, and all possible combinations of states of unique chromosomes as above. We will use the notation $\bar{x}_{\bar{x}value} = \bar{x}value$.

And we define L as the above C-MCMC-OA algorithm.

$L = \{$ {(00, 00, 00, 00, 00, 00)}: $1.700000 = \bar{x}_{1.700000}$,

{(00, 00, 00, 00, 00, 01), (00, 00, 00, 00, 01, 00), (00, 00, 00, 01, 00, 00), (00, 00, 01, 00, 00, 00), (00, 01, 00, 00, 00, 00), (01, 00, 00, 00, 00, 00)}: $1.716667 = \bar{x}_{1.716667}$, ..., {(11, 11, 11, 11, 11, 11)}: $2.000000 = \bar{x}_{2.000000}$ $\}$

Sampling distribution of $\overline{\mathbf{X}}$ $(N_1(1.850000, 0.000004))\equiv$ stationary distribution. Covariance = -0.000004 for the four groups of states are equal then the

process is lumped Markov chain. Conditional variance = 0.000001 is the same for all four groups.

All possible conditional normal distributions of transition matrix =

$\{\bar{x}_{1.700000}$:$N_1(1.989819,0.000001)$, $\bar{x}_{1.716667}$:$N_1(1.974284,0.000001)$,

..., $\bar{x}_{2.000000}$:$N_1(1.710181,0.000001)\}$.

7. Discussion

In this paper, the main result is the unified MCMC theorem for clonal selection algorithms.Using this, we propose unique chromosomes method for a purely successful optimization of these algorithms and obtain purely classification of chains, all conditional multivariate normal distributions and stationary multivariate normal distributions for them.

8. Open problems

For genetic algorithms with or without bit mutation, we can prove the unified MCMC theorem and can obtain purely classification of chains, all conditional multivariate normal distributions and stationary multivariate normal distributions for them.

References

[1] A. E. Eiben, E.H.L. Aarts and K.M. Van Hee, *Global Convergence of Genetic Algorithms: A Markov Chain Analysis*, Proceedings of the 1st international PPSN I conference, "Parallel Problem Solving from Nature", Springer 1991, 4-12 (eds. H.-P. Schwefel, R. Männer).

[2] U.H. Abou El-Enien, *Graph Drawing algorithms for Markov Chains*, M.Sc. diss. 2006 (in Egypt), ISSR, Cairo University.

[3] T. Apostol*Calculus, vol.II*, John Wiley, Inc., 1969.

[4] U.N.Bhat*Elements of Applied Stochastic Processes*, John Wiley and Sons, New York, 1972.

[5] E. Clark , A. Hone ,J. Timmis (2005) A markov chain model of the b-cell algorithm. In:C. Jacob ,M. Pilat ,P. Bentley ,J. Timmis (eds) Proc of the 4th International Conference on Artificial Immune Systems (ICARIS), Lecture Notes in Computer Science, vol 3627. Springer, Berlin, pp 318330

[6] M. Villalobos-Arias, C. A. Coello Coello and O. Hernández-Lerma, *Convergence Analysis of a Multiobjective Artificial Immune System Algorithm*, Third International Conference "Artificial Immune Systems", Springer 2004, 226-235 (eds. G. Nicosia, V. Cutello, P. Bentley, J. Timmis).

[7] V. Cutello, G. Nicosia, P. Oliveto and M. Romeo, *On the convergence of immune algorithms*, Proceedings of "Foundations of Computational Intelligence", 2007, 409-416.

[8] T.E.Davis , J.C. Principe , A Markov Chain Framework for the simple Genetic Algorithm. Evolutionary Computation., **1**(1993), no. 3, pp. 269-288.

[9] L. N. De Castro and F. J. Von Zuben, "The clonal selection algorithm with engineering applications," in Proceeding of Artificial Immune System Workshop, Genetic and Evolutionary Computation Conference (GECCO' 2000), A. S. Wu (Ed.), 2000, pp. 36-37.

[10] K. A. De Jong, William M. Spears and D. F. Gordon, *Using Markov Chains to Analyze GAFOs*, The Third International Workshop, "Foundations of Genetic Algorithms", Morgan Kaufmann 1995, 115-137 (eds. L. Darrell Whitley and Michael D. Vose).

[11] W.M. Spears and K. A. De Jong , *Analyzing GAs using Markov models with semantically ordered and lumped states*, 4th International Workshop, "Foundations of Genetic Algorithms", Morgan Kaufmann 1997, 85-100(eds.R. K. Belew and m. D. Vose).

[12] Fogel D., Asymptotic Convergence Properties of Genetic Algorithms and Evolutionary Programming: Analysis and Experiments. Cybernetics and Systems., **25**(2007), no. 3, pp. 389-407.

[13] D.B. Fogel, *Evolutionary Computation : Towards a New Philosophy of Machine Intelligence*, IEEE Press, New York, 1995.

[14] M. Goldman*Introduction to Probability and Statistics*, Harcourt, Brace and World, New York, 1970.

[15] S. Hofmeyr, S. Forrest , Architecture for an Artificial Immune System. Evolutionary Computation., **7**(2000), no.1, pp. 45-68.

[16] A. Hone and J. Kelsey*Optima, Extrema, and Artificial Immune Systems*, International Conference on Artificial Immune Systems, ICARIS, Berlin 2004, 80-90.

[17] J. Timmis , A. Hone , T. Stiborand, E. Clark , Theoretical advances in artificial immune systems. Theoretical Computer Science,Elsevier Science Publishers Ltd, **403**(2008), no. 1, pp. 11-32.

[18] R.A. Johnson, D.W. Wichern*Applied Multivariate Statistical Analysis*, Prentice Hall, New Jersey, 1998.

[19] J.G.Kemeny, J.L.Snell*Finite Markov Chains*, Springer-Verlag, New York, 1976.

[20] Neter et al. *Applied Statistics*, Allyn and Bacon, 1993.

[21] Nix E., Vose D., Modeling genetic algorithms with markov chains. Annals of Mathematics and Artificial Intelligence, **5**(1992), no.1, pp. 79-88.

[22] Rudolph G., Convergence Analysis of Canonical Genetic Algorithms. IEEE Trans. on Neural Networks, **5**(1994), no. 1, pp. 96-101.

[23] M. Zak , Physical Model of Immune Inspired Computing. Information Sciences., Elsevier **129**(2000), no.1, pp. 61-79.

[24] W. Conley, Simulation optimization and correlation with multi stage Monte Carlo optimization. International Journal of Systems Science., Taylor and Francis **38**(2007), no.12, pp. 1013-1019.

[25] Z. Michalewicz *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, 1996