Rigorous Time Complexity Analysis of Univariate Marginal Distribution Algorithm with Margins

Tianshi Chen, Ke Tang, Guoliang Chen, and Xin Yao

Abstract—Univariate Marginal Distribution Algorithms (UMDAs) are a kind of Estimation of Distribution Algorithms (EDAs) which do not consider the dependencies among the variables. In this paper, on the basis of our proposed approach in [1], we present a rigorous proof for the result that the UMDA with margins (in [1] we merely showed the effectiveness of margins) cannot find the global optimum of the TRAPLEADINGONES problem [2] within polynomial number of generations with a probability that is super-polynomially close to 1. Such a theoretical result is significant in shedding light on the fundamental issues of what problem characteristics make an EDA hard/easy and when an EDA is expected to perform well/poorly for a given problem.

I. INTRODUCTION

Estimation of Distribution Algorithms (EDAs) [10] maintain probabilistic models to generate new solutions and consistently update the probabilistic models during the optimization process. Recently, various kinds of EDAs have been proposed, however, the fundamental theoretical time complexity investigations of EDAs are still few.

Droste [4] presented the first rigorous time complexity analysis of EDA. He analyzed rigorously the first hitting time of the compact Genetic Algorithm (cGA) [6] with population size 2 on linear functions. Later, using the analytical Markov chain framework [8], González analyzed the general worst-case first hitting time of different EDAs on the pseudo-boolean injective functions in her doctoral dissertation [5]. She proved that the worst-case mean first hitting time is exponential in the problem size for four commonly used EDAs. However, in addition to the above general result, she has not analyzed any specific problem.

In [2], we provided a preliminary investigation of Univariate Marginal Distribution Algorithms (UMDAs) [12]. First we showed that the UMDA with truncation selection will spend linear (in the problem size) number of generations to find the optimum of the well-known LEADINGONES problem [7], [13]. After that, we constructed the TRAPLEADINGONES problem based on LEADINGONES, and we proved that the UMDA with 2-tournament selection should spend at least exponential number of generations to find the global optimum of the problem. However, our proofs in [2] are actually based on the “no-random-error assumption”, i.e., the stochastic operators of the UMDA will not bring any

random errors. This assumption cannot characterize the real optimization process of the stochastic algorithms, since the algorithm will always bring random errors. Hence, our preliminary investigations in [2] are not rigorous.

Later, to cope with the random errors occurred in the optimization processes of EDAs, we developed a new approach to analyze the time complexity of EDAs rigorously, and UMDAs are again involved in case studies [1]. Our approach contains two steps: First, we build an easy-to-analyze deterministic system and we exploit the time complexity result of the corresponding EDA from this deterministic system. After that, we estimate the gap between the deterministic system and the real stochastic algorithm by some analytical tools, such as Chernoff bounds [11]. By this approach, we have proven *rigorously* that the UMDA can solve LEADINGONES efficiently. Furthermore, we have also proven rigorously a pair of interesting results, showing that the naive (original) UMDA will fail to optimize a unimodal problem called BVLEADINGONES while the UMDA with margins can avoid premature convergence and thus find the optimum of BVLEADINGONES easily. It is worth noting that there are still many open questions to answer in addition to the investigations in [1], e.g., can we find some problems that are hard for the UMDA with margins? Can we find some problem that is hard for the $(1 + 1)$ EA while it is easy for the UMDA without margins?

This paper serves as an extended and complementary investigation of [1], in which we aim at answering the first question above by rigorous theoretical analysis, confirming that the UMDA cannot find the optimum of TRAPLEADINGONES within a polynomial number (with respect to the problem size) of generations with an overwhelming probability, even if the UMDA is further improved by margins. Moreover, the result of this paper is an extra example of applying our approach to analyze rigorously the behaviors of EDAs in addition to the three theorems presented in [1]. Moreover, recently we have also provided an answer to the second open question mentioned above: In [3], we prove that the so-called SUBSTRING problem is hard for the $(1 + 1)$ EA while it is easy for the UMDA (without margins).

The rest of the paper is organized as follows: Section II introduces the preliminaries of the paper; Section III presents our main result and the corresponding proof; Section IV concludes the whole paper.

The authors are with the Nature Inspired Computation and Applications Laboratory (NICAL), Department of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui 230027, China. Xin Yao is also with the Centre of Excellence for Research in Computational Intelligence and Applications (CERCIA), School of Computer Science, University of Birmingham, UK. (Email: cetacy@mail.ustc.edu.cn, ketang@ustc.edu.cn, glchen@ustc.edu.cn, x.yao@cs.bham.ac.uk)

TABLE I
UMDA WITH TRUNCATION SELECTION

$p_{0,i}(x_i) \leftarrow$ Initial values ($\forall i=1, \dots, n$)
$\xi_1 \leftarrow N$ individuals are sampled according to the distribution $p_0(\mathbf{x}) = \prod_{i=1}^n p_{0,i}(x_i)$
REPEAT
$\xi_t^{(s)} \leftarrow$ The best M individuals are selected from the N individuals in ξ_t ($N > M$)
$p_{t,i}(1) \leftarrow \sum_{\mathbf{x} \in \xi_t^{(s)}} \delta(x_i 1)/M$, $p_{t,i}(0) \leftarrow 1 - p_{t,i}(1)$ ($\forall i=1, \dots, n$)
$\xi_{t+1} \leftarrow N$ individuals are sampled according to the distribution $p_t(\mathbf{x}) = \prod_{i=1}^n p_{t,i}(x_i)$
UNTIL THE STOPPING CRITERION IS MET.

II. PRELIMINARIES

A. Algorithm

A general procedure of UMDA for binary search space is presented in Table I, where $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \{0, 1\}^n$ represents an individual (solution), ξ_t and $\xi_t^{(s)}$ represent the populations before and after the selection at the t^{th} generation ($t \in \mathbb{N}^+$) respectively, $p_{t,i}(1)$ ($p_{t,i}(0)$) is the estimated marginal probability of the i^{th} bit of an individual to be 1 (0) at the t^{th} generation, and the indicator $\delta(x_i|1)$ is further defined as follows:

$$\delta(x_i|1) = \begin{cases} 1, & x_i = 1, \\ 0, & x_i = 0. \end{cases}$$

The marginal probabilities $p_{t,i}(1)$ and $p_{t,i}(0)$ are given by

$$p_{t,i}(1) = \frac{\sum_{\mathbf{x} \in \xi_t^{(s)}} \delta(x_i|1)}{M}, \quad p_{t,i}(0) = 1 - p_{t,i}(1).$$

Let $\mathbf{P}_t(\mathbf{x}) = (p_{t,1}(x_1), p_{t,2}(x_2), \dots, p_{t,n}(x_n))$, where $\mathbf{P}_t(\mathbf{x})$ is a vector of random variables. Then the probability of generating a specific individual $\mathbf{x}$ in the t^{th} generation is

$$p_t(\mathbf{x}) = \prod_{i=1}^n p_{t,i}(x_i).$$

Besides, the UMDA studied in this paper adopts the truncation selection: At the t^{th} generation the selection operator selects the best M individuals among the N individuals in ξ_t , and then $\xi_t^{(s)}$ is obtained for estimating the probability distribution of the t^{th} generation.

Furthermore, in this paper we will concern an improved version of the UMDA: the UMDA with *margins*. The idea of margins is implemented as follows:

The highest level the marginal probabilities can reach is $1 - 1/M$ and the lowest level the marginal probabilities can drop to is $1/M$. Any marginal probabilities higher than $1 - \frac{1}{M}$ is set to be $1 - 1/M$, and any marginal probabilities lower than $1/M$ is set to be $1/M$ [1].

The reason that we employ the above improved UMDA in our analysis is that the original UMDA cannot avoid premature convergence at all, and has already been proven to be inefficiently on even a unimodal problem (e.g., BVLEADINGONES problem [1]). To exploit the ability of the UMDA to the full extent, we allow the UMDA to be improved slightly while the basic framework of the algorithm does not change.

B. Problem

The *maximization* problem we consider in this paper is called TRAPLEADINGONES [2]:

$$\text{TRAPLEADINGONES}(\mathbf{x}) \triangleq \begin{cases} b(\mathbf{x}), & b(\mathbf{x}) \leq n, \\ -n, & b(\mathbf{x}) > n, \end{cases} \quad (1)$$

where $\mathbf{x} = (x_1, \dots, x_n)$, $b(\mathbf{x}) = nx_n + \sum_{i=1}^{n-1} \prod_{j=1}^i x_j$, and $\forall k \in \{1, \dots, n\} : x_k \in \{0, 1\}$. The global optimum of TRAPLEADINGONES function is $\mathbf{x}^* = (0, \dots, 0, 1)$. TRAPLEADINGONES has similar structure to LEADINGONES. However, the leading 1-bits will eventually lead to the local optimum $(1, \dots, 1, 0)$ instead of leading to the global optimum $(0, \dots, 0, 1)$. In other words, TRAPLEADINGONES is a deceptive multimodal problem.

C. Analytical Approach and Concrete Tools

In this paper, we will utilize the approach introduced in [1] to analyze the algorithm. The approach can be summarized as the following two major steps according to [1]:

- 1) *Build an easy-to-analyze discrete dynamic system for the EDA. The idea is to de-randomize the EDA and build a deterministic dynamic system.*
- 2) *Analyze the deviations (errors) caused by de-randomization. Note that EDAs are stochastic algorithms. Concretely, tail probability techniques, such as Chernoff bounds, can be used to bound the deviations.*

Concretely, we define a function $\gamma : [0, 1]^n \rightarrow [0, 1]^n$ to represent the updating rule of the algorithm. Given the initial parameter values of the algorithm, we can obtain a deterministic discrete dynamic system $\{\hat{\mathbf{P}}_t(\mathbf{x}^*); t = 0, 1, \dots\}$ related to the marginal probabilities of generating the global optimum:

$$\hat{\mathbf{P}}_0(\mathbf{x}^*) = \mathbf{P}_0(\mathbf{x}^*), \quad (2)$$

$$\hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) = \gamma(\hat{\mathbf{P}}_t(\mathbf{x}^*)), \quad (3)$$

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) = \gamma^t(\hat{\mathbf{P}}_0(\mathbf{x}^*)), \quad (4)$$

where $\hat{\mathbf{P}}_t(\mathbf{x}) = (\hat{p}_{t,1}(x_1), \dots, \hat{p}_{t,n}(x_n))$ is the marginal probability vector of the deterministic system at the t^{th} generation.

The deterministic system is relatively easy to analyze: the time complexity of the system (e.g., the time for the de-randomized marginal probabilities to reach some specific values) totally depends on γ and $\{\hat{\mathbf{P}}_t(\mathbf{x}^*); t = 0, 1, \dots\}$.

What we need to do is to study quantitatively the deviation (difference) between the deterministic system and the real optimization process of the algorithm. More precisely, we need to study the deviations between $\{\hat{\mathbf{P}}_t(\mathbf{x}^*); t = 0, 1, \dots\}$ and $\{\mathbf{P}_t(\mathbf{x}^*); t = 0, 1, \dots\}$. Fortunately, there have already been some potential analytical tools enabling the estimation of such deviations. Below these tools are presented in terms of two lemmas.

Lemma 1 (Chernoff bounds [11]): Let $X_1, X_2, \dots, X_k \in \{0, 1\}$ be k independent random boolean variables with a same distribution:

$$\forall i \neq j : \mathbb{P}(X_i = 1) = \mathbb{P}(X_j = 1),$$

where $i, j \in \{1, \dots, k\}$. Let X be the sum of those random variables, i.e., $X = \sum_{i=1}^k X_i$, then we have

- $\forall 0 < \delta < 1 : \mathbb{P}(X < (1 - \delta)\mathbb{E}[X]) < e^{-\mathbb{E}[X]\delta^2/2}$.
- $\forall 0 < \delta \leq 2e - 1 : \mathbb{P}(X > (1 + \delta)\mathbb{E}[X]) < e^{-\mathbb{E}[X]\delta^2/4}$.

Lemma 2 ([1], [3], [9], [14]): Consider sampling without replacement from a finite population $\{X_1, \dots, X_N\} \in \{0, 1\}^N$. Let $\{X_1, \dots, X_M\} \in \{0, 1\}^M$ be a sample of size M drawn randomly without replacement from the whole population, $X^{(M)}$ and $X^{(N)}$ be the sums of the random variables in the sample and population respectively, i.e., $X^{(M)} = \sum_{i=1}^M X_i$ and $X^{(N)} = \sum_{i=1}^N X_i$, then we have:

$$\begin{aligned} \mathbb{P}\left(X^{(M)} - \frac{MX^{(N)}}{N} \geq M\delta\right) &< e^{-2M\delta^2}, \\ \mathbb{P}\left(\left|X^{(M)} - \frac{MX^{(N)}}{N}\right| > M\delta\right) &< 2e^{-2M\delta^2}, \end{aligned}$$

where $\delta \in [0, 1]$ is some constant (For details of the lemma, one can refer to Corollary 1.1 and Eq. 3.3 of [14]).

III. TIME COMPLEXITY ANALYSIS OF UMDA WITH MARGINS

Before our theoretical analysis, we introduce the following concept:

Definition 1 (b-promising individual [1]): In the population that contains N individuals, the b -promising individuals are those individuals with fitness no smaller than a threshold b .

Given that the UMDA adopts the truncation selection, we have the following lemma:

Lemma 3 ([1]): For the UMDA with truncation selection, the proportion of the b -promising individuals after selection at the t^{th} generation satisfies:

$$Q_{t,b}^{(s)} = \begin{cases} \frac{Q_{t,b}N}{M}, & Q_{t,b} \leq \frac{M}{N}, \\ 1, & Q_{t,b} > \frac{M}{N}, \end{cases} \quad (5)$$

where $Q_{t,b} \leq 1$ is the proportion of the b -promising individuals before the truncation selection.

The main result of the paper is presented as follows:

Theorem 1: Given the polynomial population sizes $N = \omega(n^{2+\alpha} \log n)$, $M = \omega(n^{2+\alpha} \log n)$ (where n is the problem size and α can be any positive constant) and $M = \beta N$

($\beta \in (\frac{1}{4}, 1)$ is some constant), the UMDA with truncation selection and margins cannot find the global optimum of the TRAPLEADINGONES problem within polynomial (in the problem size n) number of generations with a probability that is super-polynomially close to 1 (i.e., an overwhelming probability).

Proof: Given that $\mathbf{x}^* = (x_1^*, \dots, x_{n-1}^*, x_n^*) = (0, \dots, 0, 1)$ is the global optimum of the TRAPLEADINGONES problem, we let $\bar{x}_i^* = 1 - x_i^*$ ($i \in \{1, \dots, n\}$). Let $\hat{t}_0$ and $\hat{t}_i$ ($1 \leq i \leq n - 1$) be defined as follows,

$$\begin{aligned} \hat{t}_0 &\triangleq \min \left\{ t; \hat{p}_{t,n}(\bar{x}_n^*) = 1 - \frac{1}{M} \right\}, \\ \hat{t}_i &\triangleq \min \left\{ t; \hat{p}_{t,i}(\bar{x}_i^*) = 1 - \frac{1}{M} \right\}. \end{aligned}$$

On the basis of the above notations and definitions, we are able to decompose the optimization process into $n + 1$ different stages: The 1^{st} stage begins when the optimization process begins, and ends at the $\hat{t}_0^{th}$ generation; The 2^{nd} stage begins after the $\hat{t}_0^{th}$ generation, and ends at the $\hat{t}_1^{th}$ generation; The i^{th} stage ($i \in \{3, \dots, n\}$) begins after the $\hat{t}_{i-2}^{th}$ generation, and ends at the $\hat{t}_{i-1}^{th}$ generation; The $(n + 1)^{th}$ stage begins after the $\hat{t}_{n-1}^{th}$ generation.

Next we will introduce the deterministic system used in the first n stages. Consider the 1^{st} stage, and we let generation $t + 1$ belong to the 1^{st} stage, then the marginal probabilities at the generation is obtained from the marginal probabilities at generation t and the mapping γ_1 .

$$\begin{aligned} \hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) &= \gamma_1(\hat{\mathbf{P}}_t(\mathbf{x}^*)) = \\ &= (R\hat{p}_{t,1}(x_1^*), \dots, R\hat{p}_{t,n-1}(x_{n-1}^*), 1 - G\hat{p}_{t,n}(\bar{x}_n^*)), \end{aligned}$$

where we aim at describing two different situations:

- 1) $j = \{1, \dots, n - 1\}$: In the deterministic system above, we consider that the j^{th} bits of individuals are not exposed to selection pressure, and use the factor $R = (1 + \eta)(1 + \eta')$ ($\eta < 1$ and $\eta' < 1$ are positive functions of the problem size n) to demonstrate the impact of genetic drift on these marginal probabilities, and we let $\eta = \eta' = (\frac{1}{n})^{1+\frac{\alpha}{2}}$.
- 2) $j = n$: In the deterministic system above, the marginal probability $\hat{p}_{t,n}(\bar{x}_n^*) = 1 - \hat{p}_{t,n}(x_n^*)$ will increase, and we use the factor $G = (1 - \delta)(1 - \frac{1}{M})^{\frac{n}{M}}$ ($\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{2e(n)}}\sqrt{\frac{M}{N}})$ is a constant) to demonstrate the impact of selection pressure on the increasing marginal probability $\hat{p}_{t,n}(\bar{x}_n^*)$ ($\hat{p}_{t+1,n}(\bar{x}_n^*) = G\hat{p}_{t,n}(\bar{x}_n^*)$, thus $\hat{p}_{t+1,n}(x_n^*) = 1 - G\hat{p}_{t,n}(\bar{x}_n^*)$ holds).

If generation $t + 1$ belongs to the i^{th} stage ($i \in \{2, \dots, n\}$), then the marginal probabilities at the generation is obtained from the marginal probabilities at generation t and the mapping γ_i .

$$\begin{aligned} \hat{\mathbf{P}}_{t+1}(\mathbf{x}^*) &= \gamma_i(\hat{\mathbf{P}}_t(\mathbf{x}^*)) = (\hat{p}_{t,1}(x_1^*), \dots, \hat{p}_{t,i-2}(x_{i-2}^*), \\ &= 1 - G\hat{p}_{t,i-1}(\bar{x}_{i-1}^*), R\hat{p}_{t,i}(x_i^*), \dots, R\hat{p}_{t,n-1}(x_{n-1}^*), \hat{p}_{t,n}(x_n^*)), \end{aligned}$$

where we aim at describing several different situations:

- 1) $j \leq i - 2$ and $j \in \mathbb{N}^+$: In the deterministic system above, the j^{th} bits of individuals have been exposed to selection pressure for enough time, and $\hat{p}_{t,j}(x_j^*)$ and $\hat{p}_{t,j}(\bar{x}_j^*)$ remain to be $1/M$ and $1 - 1/M$ respectively.

- 2) $j = i - 1$: In the deterministic system above, the marginal probability $\hat{p}_{t,j}(\bar{x}_j^*) = 1 - \hat{p}_{t,j}(x_j^*)$ will increase, and we use the factor $G = (1 - \delta)(1 - \frac{1}{M})^n \frac{N}{M}$ ($\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{2\epsilon(n)}} \sqrt{\frac{M}{N}})$ is a constant) to demonstrate the impact of selection pressure on the increasing marginal probability $\hat{p}_{t,j}(\bar{x}_j^*)$ ($\hat{p}_{t+1,j}(\bar{x}_j^*) = G\hat{p}_{t,j}(\bar{x}_j^*)$, thus $\hat{p}_{t+1,j}(x_j^*) = 1 - G\hat{p}_{t,j}(x_j^*)$ holds).
- 3) $j = \{i, \dots, n-1\}$: In the deterministic system above, the j^{th} bits of individuals are not exposed to selection pressure, and we use the factor $R = (1 + \eta)(1 + \eta')$ ($\eta < 1$ and $\eta' < 1$ are positive functions of the problem size n) to demonstrate the impact of genetic drift on these marginal probabilities, and in the proof we let $\eta = \eta' = (\frac{1}{n})^{1+\frac{\alpha}{2}}$.
- 4) $j = n$: In the deterministic system above, the n^{th} bits of individuals have been exposed to selection pressure for enough time, and $\hat{p}_{t,n}(x_n^*)$ and $\hat{p}_{t,n}(\bar{x}_n^*)$ remain to be the value of $1/M$ and $1 - 1/M$ respectively.

Let us investigate the property of the deterministic system $\hat{\mathbf{P}}_t(\mathbf{x}^*)$ at the 1^{st} stage first, where the time index t satisfies that $0 < t \leq \hat{t}_0$. At this stage, we concern the 0-promising individuals, since the n^{th} bits of individuals are exposed to the selection pressure. As a consequence, we study the n^{th} component of $\hat{\mathbf{P}}_t(\mathbf{x}^*)$, i.e., the deterministic marginal probability $\hat{p}_{t,n}(x_n^*)$. Given the initial value that $\hat{p}_{0,n}(x_n^*) = \frac{1}{2}$ holds, the condition that $\forall t < \hat{t}_0 - 1$: $\frac{1}{G}(1 - \frac{1}{M}) > \hat{p}_{t,n}(\bar{x}_n^*) = 1 - \hat{p}_{t,n}(x_n^*) = G\hat{p}_{t-1,n}(\bar{x}_n^*)$ implies Eqs 6 and 7:

$$G^{\hat{t}_0-2}\hat{p}_{0,n}(\bar{x}_n^*) < \frac{1}{G}\left(1 - \frac{1}{M}\right), \quad (6)$$

$$G^{\hat{t}_0-1}\hat{p}_{0,n}(\bar{x}_n^*) \geq \frac{1}{G}\left(1 - \frac{1}{M}\right). \quad (7)$$

Hence we obtain that:

$$\hat{t}_0 \leq \frac{\ln \frac{2(M-1)}{N} - \ln(1 - \delta) + \frac{1}{\epsilon(n)}}{\ln(1 - \delta) + \ln\left(\frac{N}{M}\right) - \frac{1}{\epsilon(n)}} + 2, \quad (8)$$

where $\epsilon(n) = M/n$. Given the polynomial population sizes $N = \omega(n^{2+\alpha} \log n)$, $M = \omega(n^{2+\alpha} \log n)$ (where α can be any positive constant) and $M = \beta N$ ($\beta \in (\frac{1}{4}, 1)$ is some constant), we know that $\hat{t}_0 = \Theta(1)$.

On the other hand, for the marginal probability $\hat{p}_{t,j}(x_j^*)$ ($j \in \{1, \dots, n-1\}$) which characterizes the j^{th} bits of individuals at the 1^{st} stage, the definition of the deterministic system also implies an common upper bound for them. Given $1 < t \leq \hat{t}_0 = \Theta(1)$, we have

$$\forall j \in \{1, \dots, n-1\} : \hat{p}_{t,j}(x_j^*) = \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^{2t} \hat{p}_{0,j}(x_j^*) < \frac{3}{4} \quad (9)$$

holds. By similar calculation as we have done in Eqs. 6, 7 and 9, we can obtain the following two results for the i^{th} stage ($i \in \{2, \dots, n\}$):

$$\hat{t}_{i-1} \leq \frac{i \ln \frac{4(M-1)}{N} - i \ln(1 - \delta) + \frac{i}{\epsilon(n)}}{\ln(1 - \delta) + \ln\left(\frac{N}{M}\right) - \frac{1}{\epsilon(n)}} + 2i, \quad (10)$$

and the condition of the theorem ($M = \beta N$ and δ is a positive constant) implies that $\hat{t}_{i-1} = O(n)$ holds for $i \in$

$\{2, \dots, n\}$. Thus for any t that satisfies $\hat{t}_{i-2} < t \leq \hat{t}_{i-1} = O(n)$, we have

$$\forall j \in \{i, \dots, n-1\} : \hat{p}_{t,j}(x_j^*) = \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^{2t} \hat{p}_{0,j}(x_j^*) < \frac{3}{4}. \quad (11)$$

It is worth noting that in Eq. 10, the coefficient of the item $\frac{M-1}{N}$ is 4 while that in Eq. 8 is 2. The reason is that the initial value of the marginal probability which is under selection pressure at the i^{th} stage ($\hat{p}_{\hat{t}_{i-2}, i-1}(\bar{x}_{i-1}^*)$) is no smaller than $1/4$ (it is implied by an inequality similar to Eq. 11, but differs from Eq. 11 that it holds for the $(i-1)^{th}$ stage) while that of the 1^{st} stage ($\hat{p}_{0,n}(\bar{x}_n^*)$) equals $1/2$.

By restricting our analysis to the 1^{st} stage first, we now utilize *induction* to prove that $\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*)$ holds at the 1^{st} stage with an overwhelming probability. We have to prove that the components of $\mathbf{P}_t(\mathbf{x}^*)$ are no larger than the corresponding components of $\hat{\mathbf{P}}_t(\mathbf{x}^*)$ respectively. At this stage we need to consider two kinds of bits: The first kind contains the n^{th} bits of individuals, and they are exposed to the selection pressure at the 1^{st} stage if the global optimum has not been generated; The second kind contains the j^{th} bits of individuals ($j \in \{1, \dots, n-1\}$), and we assume that they have not been exposed to the selection pressure at the 1^{st} stage if the global optimum has not been generated, which is regarded as a best case analysis for $p_{t,j}(x_j^*)$ ($j \in \{1, \dots, n-1\}$).

The induction begins with the first generation. As the first step, we need to prove that $\mathbf{P}_1(\mathbf{x}^*) \leq \hat{\mathbf{P}}_1(\mathbf{x}^*)$. Let us study the first kind of bits mentioned above. Concerning with the marginal probability $p_{1,n}(x_n^*)$ that characterizes the n^{th} bits of individuals at the 1^{st} generation, we apply Chernoff bounds and obtain the inequalities in Table II, where $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{2\epsilon(n)}} \sqrt{\frac{M}{N}})$ is a positive constant. Since the population size N is polynomial and the initial value $\hat{p}_{0,k}(x_k^*) = 1/2$ holds for any $k \in \{1, \dots, n\}$, we know that the probability estimated in Table II is an overwhelming one.

We now carry out best case analysis for the marginal probability $p_{1,j}(x_j^*)$ ($j \in \{1, \dots, n-1\}$) which characterizes the j^{th} bits of individuals. Since we consider the 0-promising individuals, we do not need to concern the augment of other marginal probabilities $p_{1,j}(\bar{x}_j^*)$. Instead, the genetic drift has to be taken into account, since given the condition that the global optimum has not been generated at the 1^{st} stage, in the best case there will be no selection pressure on the j^{th} bits of individuals. Recall that in the deterministic system we have defined a factor $R = (1 + \eta)(1 + \eta')$ ($\eta < 1$ and $\eta' < 1$ are positive functions of the problem size n) to demonstrate the impact of genetic drift on these marginal probabilities, where $\eta = \eta' = (\frac{1}{n})^{1+\frac{\alpha}{2}}$ holds, next we will show that with an overwhelming probability $p_{1,j}(x_j^*)$ is bounded by $\hat{p}_{1,j}(x_j^*)$ at the first generation.

For the marginal probability $p_{1,j}(x_j^*)$ ($j \in \{1, \dots, n-1\}$), we apply Chernoff bounds to study the deviations brought

TABLE II

$$\begin{aligned}
& \mathbb{P}\left(p_{1,n}(x_n^*) \leq \hat{p}_{1,n}(x_n^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) = \mathbb{P}\left(p_{1,n}(\bar{x}_n^*) \geq \hat{p}_{1,n}(\bar{x}_n^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(p_{1,n}(\bar{x}_n^*) \geq G p_{0,n}(\bar{x}_n^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) > \mathbb{P}\left(M p_{1,n}(\bar{x}_n^*) \geq (1-\delta) p_{0,n}(\bar{x}_n^*) \left(1 - \frac{1}{M}\right)^n N, \mathbf{x}^* \notin \xi_1 \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \mathbb{P}\left(M p_{1,n}(\bar{x}_n^*) \geq (1-\delta) p_{0,n}(\bar{x}_n^*) \left(1 - \frac{1}{M}\right)^n N \mid \mathbf{x}^* \notin \xi_1, \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \mathbb{P}\left(\mathbf{x}^* \notin \xi_1 \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(M p_{1,n}(\bar{x}_n^*) \geq (1-\delta) p_{0,n}(\bar{x}_n^*) \left(1 - \frac{1}{M}\right)^n N \mid \mathbf{x}^* \notin \xi_1, \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \mathbb{P}\left(\mathbf{x}^* \notin \xi_1 \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right) \left(1 - \prod_{k=1}^n \hat{p}_{0,k}(x_k^*)\right)^N. \tag{12}
\end{aligned}$$

by the random sampling procedures, we have:

$$\begin{aligned}
& \mathbb{P}\left(N_{1,j}(x_j^*) \leq (1+\eta) p_{0,j}(x_j^*) N \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > 1 - e^{-\hat{p}_{0,j}(x_j^*) N \eta^2 / 4},
\end{aligned}$$

where η is a parameter, and $N_{1,j}(x_j^*)$ is the number of individuals that take the value x_j^* in their j^{th} bit in the population before selection at the 1^{st} generation.

Some random deviation will also be brought by the truncation selection operator, since it has to deal with some individuals with the same fitness (genetic drift [15]). Noting that in our best case analysis the j^{th} bits of individuals are not exposed to the selection pressure, then for these bits the selection procedure can be regarded as Simple Random Sampling without replacement. Next we use Lemma 2 to estimate the probability that the number of individuals taking the value x_j^* on their j^{th} bits after selection of the 1^{st} generation (let this number be $N_{1,j}^{(s)}(x_j^*)$):

$$\begin{aligned}
& \mathbb{P}\left(N_{1,j}^{(s)}(x_j^*) < (1+\eta')(1+\eta) p_{0,j}(x_j^*) M \mid N_{1,j}(x_j^*) \leq (1+\eta) p_{0,j}(x_j^*) N, \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(N_{1,j}^{(s)}(x_j^*) - (1+\eta) p_{0,j}(x_j^*) M < \eta'(1+\eta) p_{0,j}(x_j^*) M \mid N_{1,j}(x_j^*) \leq (1+\eta) p_{0,j}(x_j^*) N, \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > 1 - e^{-2(1+\eta)^2 \hat{p}_{0,j}^2(x_j^*) \eta'^2 M}.
\end{aligned}$$

where η' is a parameter, and the definition of $N_{1,j}^{(s)}(x_j^*)$ implies that $N_{1,j}^{(s)}(x_j^*) = p_{1,j}(x_j^*) M$. Let $\eta' = \eta = \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}$, the condition $M = \omega(n^{2+\alpha} \log n)$ further implies that

$$\begin{aligned}
& \mathbb{P}\left(p_{1,j}(x_j^*) \leq \hat{p}_{1,j}(x_j^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(p_{1,j}(x_j^*) \leq \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 p_{0,j}(x_j^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - n^{-p_{0,j}(x_j^*) \omega(1)}\right) \left(1 - n^{-\left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*) \omega(1)}\right) \\
& > \left(1 - n^{-\left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*) \omega(1)}\right)^2
\end{aligned}$$

holds for any $j \in \{1, \dots, n-1\}$ (there are $n-1$ marginal probabilities belonging to the above kind). Combining the above inequality with Eq. 12, we obtain the following

inequality:

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_1(\mathbf{x}^*) \leq \hat{\mathbf{P}}_1(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right) \\
& \quad \left(1 - n^{-\left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*) \omega(1)}\right)^{2(n-1)} \\
& \quad \left(1 - \prod_{k=1}^n \hat{p}_{0,k}(x_k^*)\right)^N,
\end{aligned}$$

where $\hat{p}_{0,j}(x_j^*) = 1/2$ is the initial value. The above inequality implies that $\mathbf{P}_1(\mathbf{x}^*) \leq \hat{\mathbf{P}}_1(\mathbf{x}^*)$ holds with an overwhelming probability.

Now we assume that at the $(t-1)^{th}$ generation ($1 < t \leq \hat{t}_0$), the following inequality holds:

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right)^{t-1} \\
& \quad \left(1 - n^{-\left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*) \omega(1)}\right)^{2(n-1)(t-1)} \\
& \quad \prod_{t'=0}^{t-2} \left(1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)\right)^N. \tag{13}
\end{aligned}$$

The aim of induction is to prove the following inequality for the t^{th} generation ($1 < t \leq \hat{t}_0$):

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right)^t \\
& \quad \left(1 - n^{-\left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*) \omega(1)}\right)^{2(n-1)t} \\
& \quad \prod_{t'=0}^{t-1} \left(1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)\right)^N.
\end{aligned}$$

Now we decompose the probability of $\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*)$, conditional on $\mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)$:

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*), \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*), \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& \quad \mathbb{P}\left(\mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& \quad \mathbb{P}\left(\mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right),
\end{aligned}$$

where we utilize the Markov property of the UMDA. Noting that Eq. 13 holds, to finish our induction we only need to prove the following inequality:

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right) \\
& \quad \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*) \omega(1)}\right)^{2(n-1)} \\
& \quad \left(1 - \prod_{k=1}^n \hat{p}_{t-1,k}(x_k^*)\right)^N,
\end{aligned}$$

where $1 < t \leq \hat{t}_0$ holds, i.e., the t^{th} generation belongs to the 1^{st} stage.

Concerning with the marginal probability $p_{t,n}(x_n^*)$ that characterizes the n^{th} bits of individuals at the t^{th} generation, we apply Chernoff bounds and obtain the inequalities in Table III, where $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{2e(n)}} \sqrt{\frac{M}{N}})$ is a positive constant. In addition to $p_{t,n}(x_n^*)$, we now carry out best case analysis for the marginal probability $p_{t,j}(x_j^*)$ ($j \in \{1, \dots, n-1\}$) which characterizes the j^{th} bits of individuals at the t^{th} generation. By setting $\eta = (\frac{1}{n})^{1+\frac{\alpha}{2}}$ in the deterministic system, we now show that with an overwhelming probability $p_{t,j}(x_j^*)$ is bounded by $\hat{p}_{t,j}(x_j^*)$. We apply Chernoff bounds to study the deviations brought by the random sampling procedures, we have:

$$\begin{aligned}
& \mathbb{P}\left(N_{t,j}(x_j^*) \leq (1+\eta)p_{t-1,j}(x_j^*)N \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > 1 - e^{-\hat{p}_{t-1,j}(x_j^*)N\eta^2/4} > 1 - e^{-\hat{p}_{0,j}(x_j^*)N\eta^2/4},
\end{aligned}$$

where $N_{t,j}(x_j^*)$ is the number of individuals that take the value x_j^* in their j^{th} bit in the population before selection at the t^{th} generation, and we utilize the fact that $\hat{p}_{t-1,j}(x_j^*) > \hat{p}_{0,j}(x_j^*)$ holds for $1 < t \leq \hat{t}_0$ (the consequence of $R > 1$).

As we have done at the 1^{st} generation, we will also deal with the deviations brought by the truncation selection operator (since it has to deal with some individuals with the same fitness). Noting that in our best case analysis the j^{th} bits of individuals are not exposed to the selection pressure at the whole 1^{st} stage, then for these bits the selection procedure can be regarded as Simple Random Sampling without replacement. By Lemma 2, we estimate the probability that the number of individuals taking the value x_j^* on their j^{th} bits after selection of the t^{th} generation (let

this number be $N_{t,j}^{(s)}(x_j^*)$):

$$\begin{aligned}
& \mathbb{P}\left(N_{t,j}^{(s)}(x_j^*) < (1+\eta')(1+\eta)p_{t-1,j}(x_j^*)M \mid N_{t,j}(x_j^*) \leq (1+\eta)p_{t-1,j}(x_j^*)N, \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(N_{t,j}^{(s)}(x_j^*) - (1+\eta)p_{t-1,j}(x_j^*)M < \eta'(1+\eta)p_{t-1,j}(x_j^*)M \mid N_{t,j}(x_j^*) \leq (1+\eta)p_{t-1,j}(x_j^*)N, \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > 1 - e^{-2(1+\eta)^2 \hat{p}_{0,j}^2(x_j^*) \eta'^2 M}.
\end{aligned}$$

where the definition of $N_{t,j}^{(s)}(x_j^*)$ implies that $N_{t,j}^{(s)}(x_j^*) = p_{t,j}(x_j^*)M$. By setting $\eta' = \eta = (\frac{1}{n})^{1+\frac{\alpha}{2}}$, the conditions $N = \omega(n^{2+\alpha} \log n)$ and $M = \omega(n^{2+\alpha} \log n)$ further implies that

$$\begin{aligned}
& \mathbb{P}\left(p_{t,j}(x_j^*) \leq \hat{p}_{t,j}(x_j^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(p_{t,j}(x_j^*) \leq \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\alpha}{2}}\right)^2 p_{t-1,j}(x_j^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > \left(1 - n^{-p_{0,j}(x_j^*)\omega(1)}\right) \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*)\omega(1)}\right), \\
& > \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*)\omega(1)}\right)^2
\end{aligned}$$

holds for any $j \in \{1, \dots, n-1\}$. Combining the above inequality with Eq. 14, we obtain the following inequality:

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right) \\
& \quad \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*)\omega(1)}\right)^{2(n-1)} \\
& \quad \left(1 - \prod_{k=1}^n \hat{p}_{t-1,k}(x_k^*)\right)^N,
\end{aligned}$$

where the initial value of the UMDA $\hat{p}_{0,j}(x_j^*) = 1/2$ holds. Hence, we have proven that, given that the t^{th} generation belongs to the 1^{st} stage ($1 < t \leq \hat{t}_0$), the following inequality always holds:

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right)^t \\
& \quad \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \hat{p}_{0,j}^2(x_j^*)\omega(1)}\right)^{2(n-1)t} \\
& \quad \prod_{t'=0}^{t-1} \left(1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)\right)^N. \tag{15}
\end{aligned}$$

Since the initial value $\hat{p}_{0,n}(\bar{x}_n^*) = \hat{p}_{0,j}(x_j^*) = 1/2$ holds, the above inequality implies:

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 8}\right)^t \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2(n-1)t} \\
& \quad \prod_{t'=0}^{t-1} \left(1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)\right)^N. \tag{16}
\end{aligned}$$

TABLE III

$$\begin{aligned}
& \mathbb{P}\left(p_{t,n}(x_n^*) \leq \hat{p}_{t,n}(x_n^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) = \mathbb{P}\left(p_{t,n}(\bar{x}_n^*) \geq \hat{p}_{t,n}(\bar{x}_n^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(p_{t,n}(\bar{x}_n^*) \geq G\hat{p}_{t-1,n}(\bar{x}_n^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \geq \mathbb{P}\left(p_{t,n}(\bar{x}_n^*) \geq Gp_{t-1,n}(\bar{x}_n^*) \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > \mathbb{P}\left(Mp_{t,n}(\bar{x}_n^*) \geq (1-\delta)p_{t-1,n}(\bar{x}_n^*)\left(1 - \frac{1}{M}\right)^n N, \mathbf{x}^* \notin \xi_t \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > \mathbb{P}\left(Mp_{t,n}(\bar{x}_n^*) \geq (1-\delta)p_{t-1,n}(\bar{x}_n^*)\left(1 - \frac{1}{M}\right)^n N \mid \mathbf{x}^* \notin \xi_t, \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \mathbb{P}\left(\mathbf{x}^* \notin \xi_t \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& = \mathbb{P}\left(Mp_{t,n}(\bar{x}_n^*) \geq (1-\delta)p_{t-1,n}(\bar{x}_n^*)\left(1 - \frac{1}{M}\right)^n N \mid \mathbf{x}^* \notin \xi_t, \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \mathbb{P}\left(\mathbf{x}^* \notin \xi_t \mid \mathbf{P}_{t-1}(\mathbf{x}^*) \leq \hat{\mathbf{P}}_{t-1}(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{t-1,n}(\bar{x}_n^*) N \delta^2 / 2}\right) \left(1 - \prod_{k=1}^n \hat{p}_{t-1,k}(x_k^*)\right)^N > \left(1 - e^{-(1-\frac{1}{M})^n \hat{p}_{0,n}(\bar{x}_n^*) N \delta^2 / 2}\right) \left(1 - \prod_{k=1}^n \hat{p}_{t-1,k}(x_k^*)\right)^N. \quad (14)
\end{aligned}$$

On the other hand, for any t' satisfies $1 < t' \leq \hat{t}_0 = \Theta(1)$, to Chernoff bounds, for the i^{th} stage, we have

$$\begin{aligned}
& \forall j \in \{1, \dots, n-1\}: \\
& \hat{p}_{t',j}(x_j^*) = \left(1 + \left(\frac{1}{n}\right)^{1+\frac{\delta}{2}}\right)^{2t'} \hat{p}_{0,j}(x_j^*) < \frac{3}{4} \\
& \mathbb{P}\left(r_t\left((1^{i-2} * \dots * 1)\right) > (1-\delta)\left(1 - \frac{1}{M}\right)^{i-1} \mid \mathbf{x}^* \notin \xi_t, p_{t-1,n}(\bar{x}_n^*) = 1 - \frac{1}{M}, \right. \\
& \quad \left. \forall j \leq i-2: p_{t-1,j}(\bar{x}_j^*) = 1 - \frac{1}{M}\right) \\
& > 1 - e^{-(1-\frac{1}{M})^{i-1} N \delta^2 / 2} > 1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 2}. \quad (18)
\end{aligned}$$

holds. Hence, we know that $1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)$ is super-polynomially close to 1. Noting that the population size N is polynomial, we know that the probability mentioned in Eq. 16 is an overwhelming one. So far we have proven that at the 1^{st} stage ($0 < t \leq \hat{t}_0$), $\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*)$ holds with an overwhelming probability.

Next we must prove that at the i^{th} stage ($i \in \{2, \dots, n\}$, $\hat{t}_{i-2} < t \leq \hat{t}_{i-1}$), $\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*)$ still holds with an overwhelming probability. The above result for us to prove can be formally written as follows: for any t that satisfies $\hat{t}_{i-2} < t \leq \hat{t}_{i-1}$, we have

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 2}\right)^t \left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 8}\right)^t \\
& \quad \left(1 - n^{-(1+(\frac{1}{n})^{1+\frac{\delta}{2}})^2 \omega(1)}\right)^{2(n-i)t+2\sum_{k=0}^{i-3} \hat{t}_k} \\
& \quad \prod_{t'=0}^{t-1} \left(1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)\right)^N. \quad (17)
\end{aligned}$$

The idea of proving the above result has been shown in the proof for the 1^{st} stage, however, the $i-1$ -promising individuals will be considered at the i^{th} stage. However, to prove the above result, an additional result is required: at the i^{th} stage, the marginal probability $p_{.,j}(\bar{x}_j^*)$ has reached $1 - \frac{1}{M}$ ($j \in \{1, \dots, i-2, n\}$), $p_{.,j}(\bar{x}_j^*)$ will not drop to a level that is smaller than $1 - \frac{1}{M}$ with an overwhelming probability. This proposition result in the first factor $\left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 2}\right)^t$ at Eq. 17. Let $r_t((1^{i-2} * \dots * 1))$ be the proportion of individuals $(1^{i-2} * \dots * 1)$ before selection at the t^{th} generation, where we let $*$ be either 0 or 1. According

Combining with the fact that $\delta \in (\max\{0, 1 - \frac{2M}{N}\}, 1 - e^{\frac{1}{2\epsilon(n)}\sqrt{\frac{M}{N}}})$,

$$\begin{aligned}
& r_t((1^{i-2} * \dots * 1)) > (1-\delta)\left(1 - \frac{1}{M}\right)^{i-1} \\
& > (1-\delta)^2 \left(1 - \frac{1}{M}\right)^n > \frac{M}{N}
\end{aligned}$$

holds with an overwhelming probability $1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 2}$. Thus, according to Lemma 3, after the selection the marginal probability $p_{.,j}(\bar{x}_j^*)$ ($j \in \{1, \dots, i-2, n\}$) will still maintain a level of $1 - \frac{1}{M}$ with an overwhelming probability.

Due to the length of the paper, it would be hard to present the detailed proof for the i^{th} stage. Fortunately, the proof for the i^{th} stage is not very different from that for the 1^{st} stage, and by induction for every stage respectively (as we have done to the 1^{st} stage), it is not hard to achieve the following result: given any $0 < t \leq \hat{t}_{n-1}$

$$\begin{aligned}
& \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\
& > \left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 2}\right)^t \left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 8}\right)^t \\
& \quad \left(1 - n^{-(1+(\frac{1}{n})^{1+\frac{\delta}{2}})^2 \omega(1)}\right)^{2(n-i)t+2\sum_{k=0}^{i-3} \hat{t}_k} \\
& \quad \prod_{t'=0}^{t-1} \left(1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)\right)^N, \quad (19)
\end{aligned}$$

which is an overwhelming probability. Since $\hat{t}_{n-1} = O(n)$, then we know that the probability of the event $\forall t \in (0, \hat{t}_{n-1}], t \in \mathbb{N}^+ : \mathbf{P}_t(\mathbf{x}^*) \leq \hat{\mathbf{P}}_t(\mathbf{x}^*)$ holds with an

overwhelming probability. Noting that for $0 < t \leq \hat{t}_{n-1}$,

$$\hat{\mathbf{P}}_t(\mathbf{x}^*) < \left(\frac{3}{4}, \dots, \frac{3}{4}\right)$$

holds, we know that the probability of finding the global optimum within $\hat{t}_{n-1} = O(n)$ is smaller than

$$\left(1 - \left(1 - \left(\frac{3}{4}\right)^n\right)^{O(n)}\right) \left(1 - \frac{1}{\text{SuperPoly}(n)}\right) + \frac{1}{\text{SuperPoly}(n)},$$

where $1 - 1/\text{SuperPoly}(n)$ refers to the probability mentioned in Eq. 19, and $1/\text{SuperPoly}(n)$ refers to the difference between 1 and the probability mentioned in Eq. 19. We see that the probability of finding the global optimum before the end of the n^{th} stage is super-polynomially close to 0.

On the other hand, let us consider the case of $t > \hat{t}_{n-1}$. In this case, all the marginal probabilities $p_{t,j}(\bar{x}_j^*)$ ($j \in \{1, \dots, n\}$) has already reached $1 - 1/M$, and according to similar analysis as we have done in Eq. 18, we have

$$\begin{aligned} & \mathbb{P}\left(\mathbf{P}_t(\mathbf{x}^*) = \hat{\mathbf{P}}_t(\mathbf{x}^*) \mid \mathbf{P}_0(\mathbf{x}^*) = \hat{\mathbf{P}}_0(\mathbf{x}^*)\right) \\ & > \left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 2}\right)^t \left(1 - e^{-(1-\frac{1}{M})^n N \delta^2 / 8}\right)^{\hat{t}_{n-1}} \\ & \quad \left(1 - n^{-\left(1+(\frac{1}{n})^{1+\frac{\alpha}{2}}\right)^2 \omega(1)}\right)^{2 \sum_{k=0}^{n-1} \hat{t}_k} \\ & \quad \prod_{t'=0}^{t-1} \left(1 - \prod_{k=1}^n \hat{p}_{t',k}(x_k^*)\right)^N. \end{aligned}$$

Due to the conditions that $N = \omega(n^{2+\alpha} \log n)$, $M = \omega(n^{2+\alpha} \log n)$ and the definition of the deterministic system, we know that the above probability is an overwhelming one for any polynomial t . Consequently, given any polynomially large generation index $t > \hat{t}_{n-1}$, the probability of finding the global optimum before the t^{th} is super-polynomially close to 0.

Finally, by combining the cases of $0 < t \leq \hat{t}_{n-1}$ and $t > \hat{t}_{n-1}$ together, we have proven the theorem. ■

IV. CONCLUSION

In this paper, we provide a rigorous proof for the time complexity result of the UMDA with margins on TRAPLEADINGONES. Although only a single lower bound result is proven for the UMDA, it is sufficient to show that this deceptive problem is hard for the UMDA according to the problem hardness classification proposed in [1]. Let us recall that in [1] we have already shown that the UMDA without margins cannot solve a unimodal problem (BVLEADINGONES) efficiently; If the UMDA is further improved by margins, the unimodal problem will no longer be hard. Combining these facts with the result obtained in this paper, we know that margins can improve the performance of UMDA sometimes. However, it does not mean that we can deal with all situations by margins.

It can be shown by drift analysis [7] or Yu and Zhou's approach [16] that TRAPLEADINGONES (BVLEADINGONES) is hard (easy) for the basic $(1+1)$ EA. As a result, a problem can be easy (hard) for both the EA and UMDA with margins.

It is interesting to identify problems that are easy (hard) for the EA but hard (easy) for the UMDA with margins. Such studies may lead to more insightful understanding of the behaviors of both EAs and EDAs, and will be considered in depth in our future work.

It is important to note that our ultimate goal is to understand theoretically the relationship between problem characteristics and algorithmic features, which is an enormous challenge. Such an ultimate goal can be achieved step by step through careful and vigorous analysis of different cases that have different complexity behaviors.

ACKNOWLEDGEMENT

This work is partially supported by a National Natural Science Foundation of China grant (No. 60533020), the Fund for Foreign Scholars in University Research and Teaching Programs (Grant No. B07033), the Fund for International Joint Research Program of Anhui Science and Technology Department (No. 08080703016), and an Engineering and Physical Science Research Council grant in UK (No. EP/D052785/1).

REFERENCES

- [1] T. Chen, K. Tang, G. Chen, and X. Yao, "Analysis of Computational Time of Simple Estimation of Distribution Algorithms," submitted to *IEEE Trans. Evol. Comput.* on 26/11/2007.
- [2] T. Chen, K. Tang, G. Chen and X. Yao, "On the Analysis of Average Time Complexity of Estimation of Distribution Algorithms," in *Proc. 2007 IEEE Congr. Evol. Comput. (CEC'07)*, 2007, pp. 453–460.
- [3] T. Chen, P. K. Lehre, K. Tang, and X. Yao, "When Is an Estimation of Distribution Algorithm Better than an Evolutionary Algorithm," in *Proc. 2009 IEEE Congr. Evol. Comput. (CEC'09)*, 2009.
- [4] S. Droste, "A Rigorous Analysis of the Compact Genetic Algorithm for Linear Functions," *Natural Comput.*, vol. 5, no. 3, pp. 257–283, 2006.
- [5] C. González, *Contributions on Theoretical Aspects of Estimation of Distribution Algorithms*, Doctoral dissertation, the University of the Basque Country, 2005.
- [6] G. R. Harik, F. G. Lobo and D. E. Goldberg, "The compact genetic algorithm." In *Proc. 1998 IEEE Int. Conf. Evol. Comput.*, 1998, pp. 523–528.
- [7] J. He and X. Yao, "Drift analysis and average time complexity of evolutionary algorithms," *Artif. Intell.*, vol. 127, no. 1, pp. 57–85, 2001.
- [8] J. He and X. Yao, "Towards an Analytic Framework for Analysing the Computation Time of Evolutionary Algorithms," *Artif. Intell.*, vol. 145, no. 1-2, pp. 59–97, 2003.
- [9] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Assoc.*, vol. 58, pp. 13–30, 1963.
- [10] P. Larrañaga and J. A. Lozano, *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Norwell, MA: Kluwer, 2001.
- [11] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, 1995.
- [12] H. Mühlenbein and G. Paaß, "From recombination of genes to the estimation of distribution i. binary parameters," in *Lecture Notes in Computer Science 1411: PPSN IV*, 1996, pp. 178–187.
- [13] G. Rudolph, "Finite Markov chain results in evolutionary computation: A tour d'horizon," *Fundamenta Informaticae*, vol. 35, no.1-4, pp. 67–89, 1998.
- [14] R. J. Serfling, "Probability inequalities for the sum in sampling without replacement," *Ann. Statist.*, vol. 2, no. 1, pp. 39–48, 1974.
- [15] D. Thierens, D. E. Goldberg and A. G. Pereira, "Domino convergence, drift, and the temporal-salience structure of problems," in *Proc. 1998 IEEE Int. Conf. Evol. Comput.*, 1998, pp. 535–540.
- [16] Y. Yu and Z.-H. Zhou, "A new approach to estimating the expected first hitting time of evolutionary algorithms," *Artif. Intell.*, vol. 172, no. 15, pp. 1809–1832, 2008.