

# On the robustness of median sampling in noisy evolutionary optimization

Chao BIAN<sup>1</sup>, Chao QIAN<sup>1\*</sup>, Yang YU<sup>1</sup> & Ke TANG<sup>2</sup>

<sup>1</sup>State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China;

<sup>2</sup>Shenzhen Key Laboratory of Computational Intelligence, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Received 4 May 2020/Revised 10 July 2020/Accepted 16 October 2020/Published online 8 April 2021

**Abstract** Evolutionary algorithms (EAs) are a sort of nature-inspired metaheuristics, which have wide applications in various practical optimization problems. In these problems, objective evaluations are usually inaccurate, because noise is almost inevitable in real world, and it is a crucial issue to weaken the negative effect caused by noise. Sampling is a popular strategy, which evaluates the objective a couple of times, and employs the mean of these evaluation results as an estimate of the objective value. In this work, we introduce a novel sampling method, median sampling, into EAs, and illustrate its properties and usefulness theoretically by solving OneMax, the problem of maximizing the number of 1s in a bit string. Instead of the mean, median sampling employs the median of the evaluation results as an estimate. Through rigorous theoretical analysis on OneMax under the commonly used onebit noise, we show that median sampling reduces the expected runtime exponentially. Next, through two special noise models, we show that when the 2-quantile of the noisy fitness increases with the true fitness, median sampling can be better than mean sampling; otherwise, it may fail and mean sampling can be better. The results may guide us to employ median sampling properly in practical applications.

**Keywords** evolutionary algorithms, noisy optimization, median sampling, computational complexity, runtime analysis

**Citation** Bian C, Qian C, Yu Y, et al. On the robustness of median sampling in noisy evolutionary optimization. *Sci China Inf Sci*, 2021, 64(5): 150103, <https://doi.org/10.1007/s11432-020-3114-y>

## 1 Introduction

As a kind of general-purpose optimization algorithms, evolutionary algorithms (EAs) [1] have wide applications in practical optimization problems [2, 3]. During the optimization procedure, the obtained objective (i.e., fitness) value is usually inaccurate because of noise [4]. For example, in machine learning, the estimated performance of a prediction model usually deviates from the true performance because the model is evaluated on a limited amount of data; in aerodynamic design, the computational fluid dynamics (CFDs) simulation is needed to evaluate the performance of a given structure, which is usually computationally expensive and approximated, leading to noisy fitness. The existence of noise may mislead the search direction and deteriorate the efficiency of EAs. Therefore, it is important to handle noise in fitness evaluation during evolutionary optimization.

The sampling strategy independently evaluates the fitness  $m$  times, where  $m$  is the sample size, and then the mean of these samples is used to estimate the exact fitness. Sampling is very popular to tackle noise, because it has a  $m$ -fold reduction in the variance of the noisy evaluation. Meanwhile, it also has an  $m$ -fold increase in the computation time, thus some variants are proposed, including adaptive sampling [5, 6] and sequential sampling [7, 8], which decide the value of  $m$  dynamically in each generation. However, there has been a great lack of the theoretical understanding for sampling.

Runtime analysis, an important theoretical aspect for EAs, has achieved a lot of progresses [9–14] recently. However, they mainly consider exact environments, and the results on noisy evolutionary

\* Corresponding author (email: qianc@lamda.nju.edu.cn)

optimization is rare. Noise increases the randomness in the optimization procedure, making the analysis more difficult. As a representative evolutionary algorithm, (1+1)-EA maintains one solution in the population, and generates a new solution in each iteration by mutating the parent solution. It was first studied on two frequently-used pseudo-Boolean problems, OneMax (OM) and LeadingOnes (LO). The goal of OM is maximizing the number of 1s in a solution, while the goal of LO is maximizing the number of continuous 1s from the first bit in a solution. Runtime analysis for the two problems under various noise models [15–20] showed that only if the noise level is low, (1+1)-EA can quickly find the optimum. For instance, onebit noise is a frequently-used noise model in theoretical analysis. With probability  $p$ , it changes a uniformly selected bit in a solution before evaluation, leading to a random fitness value. For OM of size  $n$  under onebit noise, the expected runtime (ERT) of (1+1)-EA is superpolynomial if  $p = \omega(\log n/n)$ . There are also some studies concerning the effectiveness of various strategies to tackle noise, e.g., threshold selection [19, 21, 22], populations [16, 18, 20, 23, 24], and sampling [25–27]. For instance, if  $\mu = \Theta(\log n)$ , the ERT of  $(\mu + 1)$ -EA optimizing OM under onebit noise is polynomial for any  $p \in [0, 1]$  (note that  $p$  denotes the noise probability). Several studies also show the robustness of the compact genetic algorithm [28] and a simple ant colony optimization algorithm [29–32] against noise.

The above mentioned runtime analyses concerning sampling [25–27] revealed that the exponential runtime under high noise levels can be turned to be polynomial by sampling, and the sample size may be critical to the effectiveness of sampling. Moreover, Akimoto et al. [33] showed that optimization under unbiased noise can perform like exact optimization, if the sample size  $m$  is large enough. In these studies, the sampling strategy utilizes the mean of the samples as an approximation of the true fitness. Then a natural question is whether other information of the samples can be used to make EAs more robust against noise.

Note that mean is actually a measure of central tendency, and thus, it is straightforward to use another widely known measure median. Compared to mean, median has the advantage of being insensitive to outliers. For example, “breakdown point” [34, 35] is a commonly used indicator for insensitivity, which denotes the minimum ratio of variables that need to be contaminated to make the estimator become infinite (i.e., cause breakdown). The breakdown point of mean is close to 0 because a single bad observation can make the mean become infinite, whereas the breakdown point of median is 0.5 because median becomes infinite only if more than 50% of the variables become infinite. In fact, economists use the sample median frequently when reporting statistics concerning certain economic measures, e.g., household income [36].

In this paper, we introduce the sampling strategy using median (called median sampling) into EAs and theoretically examine its effectiveness. Instead of taking the mean, median sampling takes the median of the samples as an estimate for the fitness. In order to better distinguish the two sampling strategies, we call the original sampling strategy mean sampling in the following context. We will consider (1+1)-EA solving noisy OM, and derive the ERT for reaching the optimum (with respect to the exact objective). Our main results are as follows.

- For OM under onebit noise with any  $p \in [0, 1]$ , we prove that the ERT of (1+1)-EA is polynomial when median sampling with  $m = 2n^3 + 1$  is used. Previous analysis [17] has proved that the ERT of (1+1)-EA is polynomial only if  $p = O(\log n/n)$ . Thus, the result shows the robustness of median sampling against noise.

- For OM under segmented noise, we show that the ERT of (1+1)-EA using median sampling is polynomial, while the ERT of (1+1)-EA using mean sampling is exponential. The results show that median sampling can be a better choice, if the 2-quantile of the noisy fitness increases with the true fitness. Note that the noisy fitness is a random variable, and the 2-quantile of a random variable  $X$  is the value  $a$  satisfying  $P(X \leq a) \geq 0.5$  and  $P(X \geq a) \geq 0.5$  (i.e., the median of  $X$ ).

- For OM under partial noise, we show that (1+1)-EA employing median sampling fails, while (1+1)-EA employing mean sampling works. The results suggest that it would be better to choose other strategies if the 2-quantile of the noisy fitness does not increase with the true fitness.

Note that in parallel with our work, Doerr and Sutton [37] showed that median sampling can handle the negative impact of noise for an integer valued objective function  $f$ , if  $f^n(x)$  satisfies the  $\epsilon$ -concentrate condition, that is,  $P(f^n(x) - f(x) \geq 0.5) \leq 0.5 - \epsilon$  and  $P(f^n(x) - f(x) \leq -0.5) \leq 0.5 - \epsilon$ , where  $f^n(x)$  denotes the noisy objective value of  $x$ . They also considered two specific cases to show the superiority of median sampling over mean sampling. For OM under additive Cauchy noise with parameter  $\gamma \geq 0.5$ , they showed that the runtime of (1+1)-EA is superpolynomial w.h.p. (with high probability) if mean sampling is used, and the runtime is polynomial w.h.p. if median sampling is used. For LO under bitwise

noise  $(p, q)$  [27] satisfying  $p = 0.5 - \epsilon \wedge q = \Omega(1)$ , they showed a superpolynomial ERT for (1+1)-EA if mean sampling with  $m = O(n/\log^2 n)$  is used, and the runtime is polynomial w.h.p. if median sampling with  $m = O(\log n)$  is used.

The remaining paper is presented as follows. First, Section 2 presents preliminaries. Then, Section 3 analyzes the effectiveness of median sampling. Next, Sections 4 and 5 compare median sampling with mean sampling, and Section 6 provides some guidance for employing median sampling in practice. Finally, Section 7 makes a conclusion.

## 2 Preliminaries

We first present the OM problem as well as (1+1)-EA which will be considered in this paper. Next, we present the sampling strategy. The analysis tool is presented in the end.

### 2.1 OneMax problem

We consider the frequently-used pseudo-Boolean function OM. Its goal is maximizing the number of 1s (namely, the bits with value 1) in a solution. Note that  $11 \cdots 1$  (denoted by  $1^n$ ) is the unique optimal solution. The ERT of (1+1)-EA solving OM (without noise) is  $\Theta(n \log n)$  [38]. For notational convenience,  $|x|_0$  will be used to represent the number of 0s (namely the bits with value 0) of  $x$ .

**Definition 1.** The goal of the OM problem with size  $n$  is finding a binary string  $x^*$  to maximize  $f(x) = \sum_{i=1}^n x_i$ .

### 2.2 (1+1) evolutionary algorithm

(1+1)-EA reflects the general structure of EAs, and is widely analyzed to theoretically understand the behavior of EAs. Different from exact optimization, only a noisy fitness value  $f^n(x)$  can be obtained in noisy environments, and the value is a random variable because the noise may disturb the solution or the objective value randomly. For example, there are two kinds of widely used noise models: posterior and prior. The posterior noise comes from the variation on the fitness of a solution, e.g.,  $f^n(x) = f(x) + \delta$ , where  $\delta$  is randomly drawn from some distribution. The prior noise comes from the variation on a solution, i.e.,  $f^n(x) = f(x')$ , where  $x'$  is generated from  $x$  by random perturbations. Therefore, lines 5–7 in Algorithm 1 changes from the true fitness “ $f(\cdot)$ ” to the noisy fitness “ $f^n(\cdot)$ ”. In the optimization process, reevaluation strategy, which evaluates both the offspring and parent solutions in each generation, is used as in [17, 18, 29]. In the optimization procedure of an EA, fitness evaluations are the most time-consuming part, thus we will simply define its runtime as the number of objective evaluations. Termination condition is the finding of the optimum with respect to the exact objective [17, 18, 33]. In this work, we consider maximizing  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ .

---

#### Algorithm 1 (1+1)-EA

---

```

1: Let  $x$  be a uniformly chosen solution in  $\{0, 1\}^n$ ;
2: while the stopping criterion does not satisfy do
3:    $z \leftarrow$  copy  $x$ ;
4:   Change each bit of  $z$  with a probability of  $1/n$  independently;
5:   if  $f^n(z) \geq f^n(x)$  then
6:      $x \leftarrow z$ ;
7:   end if
8: end while

```

---

### 2.3 Median sampling

Mean sampling has often been used in noisy evolutionary optimization to tackle noise [5, 7]. As described in Definition 2, it uses the mean of  $m$  independent evaluations to approximate the true fitness  $f(x)$ , where  $m$  is called the sample size. By mean sampling, the output  $\bar{f}(x)$  is close to the mathematical expectation of  $f^n(x)$ . As described in Definition 3, median sampling takes the median of  $m$  independent evaluations to approximate the true fitness  $f(x)$ . By median sampling, the output  $\hat{f}(x)$  is close to the 2-quantile of  $f^n(x)$ , namely  $P(f^n(x) \leq \hat{f}(x)) \approx 1/2$ .

**Definition 2** (Mean sampling). The objective value of  $x$  is evaluated independently  $m$  times, then

$$\bar{f}(x) = \sum_{i=1}^m \frac{f_i^n(x)}{m}$$

is output, where  $f_1^n(x), f_2^n(x), \dots, f_m^n(x)$  denote  $m$  noisy fitness values.

**Definition 3** (Median sampling). The objective value of  $x$  is evaluated independently  $m$  times, then

$$\hat{f}(x) = \begin{cases} f_{i_{(m+1)/2}}^n(x), & \text{if } m \text{ is odd,} \\ (f_{i_{m/2}}^n(x) + f_{i_{m/2+1}}^n(x))/2, & \text{if } m \text{ is even} \end{cases}$$

is output, where  $f_{i_1}^n(x) \leq f_{i_2}^n(x) \leq \dots \leq f_{i_m}^n(x)$  denote the ordered noisy fitness values.

When mean (or median) sampling is used, lines 5–7 in Algorithm 1 becomes “ $\bar{f}(\cdot)$ ” (or “ $\hat{f}(\cdot)$ ”). For both of the sampling strategies,  $m = 1$  means that sampling is not used.

### 2.4 Analysis tool

It is straightforward to model the evolutionary optimization procedure as a Markov chain  $\{\xi_t\}_{t=0}^{+\infty}$ , because the subsequent procedure only depends on the current state. For (1+1)-EA optimizing OM, we can simply set the chain’s state space as  $\{0, 1\}^n$  and the optimal state as  $1^n$  (namely  $\xi_t \in \mathcal{X} = \{0, 1\}^n$  and  $\mathcal{X}^* = \{1^n\}$ ). The first hitting time (FHT) of  $\{\xi_t\}_{t=0}^{+\infty}$  is  $\tau = \min\{t \mid \xi_t \in \mathcal{X}^*, t \geq 0\}$ . If the chain’s initial state is  $\xi_0 = x$ , then its expected FHT (EFHT) is denoted by  $E(\tau \mid \xi_0 = x) = \sum_{t=0}^{+\infty} t \cdot P(\tau = t \mid \xi_0 = x)$ . If  $\xi_0$  obeys a distribution  $\pi_0$  (denoted by  $\xi_0 \sim \pi_0$ ), then its EFHT is defined as  $E(\tau \mid \xi_0 \sim \pi_0) = \sum_{x \in \mathcal{X}} \pi_0(x) E(\tau \mid \xi_0 = x)$ . For (1+1)-EA, the initial solution is evaluated once, then in each iteration, the parent solution and the offspring solution both need to be evaluated. Note that the initial solution is generated randomly from  $\{0, 1\}^n$ , thus the ERT of (1+1)-EA is  $1 + 2 \cdot E(\tau \mid \xi_0 \sim \pi_0)$ , where  $\pi_0$  denotes the uniform distribution. For (1+1)-EA using sampling, the ERT becomes  $m + 2m \cdot E(\tau \mid \xi_0 \sim \pi_0)$ , because it needs to perform  $m$  independent evaluations for each solution.

As presented in Theorem 1, the additive drift theorem aims to derive upper bounds of EFHT. To use the approach, first we need to design a function  $V(\cdot)$  as a measurement for the difference between a state and the optimal state, and  $V(\cdot)$  should satisfy  $V(x) = 0$  for any optimal  $x$  and  $V(x) > 0$  otherwise. Next, we need deriving a lower bound  $c$  for  $E(V(\xi_t) - V(\xi_{t+1}) \mid \xi_t)$ , i.e., the progress towards  $\mathcal{X}^*$  in each generation. Finally, we can upper bound EFHT through dividing  $V(\xi_0)$  by  $c$ . When the context is clear,  $V(\xi_t)/V(\xi_{t+1})$  will be briefly denoted by  $V_t/V_{t+1}$ .

**Theorem 1** (Additive drift [39]). Given  $\{\xi_t\}_{t=0}^{+\infty}$  and  $V(\cdot)$ , if  $\exists c > 0$  such that  $\forall t \geq 0$  and  $\forall \xi_t$  with  $V_t > 0$ ,

$$E(V_t - V_{t+1} \mid \xi_t) \geq c,$$

then  $E(\tau \mid \xi_0) \leq V(\xi_0)/c$ .

## 3 The robustness of median sampling against onebit noise

Onebit noise is commonly used in theoretical analyses [17, 18, 26, 27]. With probability  $p$ , it changes a uniformly selected bit in  $x$  before  $x$  is evaluated. For OM under such noise model, the ERT of (1+1)-EA is superpolynomial for  $p = \omega(\log n/n)$  [17]; the ERT is polynomial  $\forall p \in [0, 1]$  if using mean sampling with  $m = 4n^3$  [27]. Theorem 2 shows that the ERT is polynomial if using median sampling with  $m = 2n^3 + 1$ , which illustrates that median sampling can efficiently tackle noise.

**Definition 4** (Onebit noise). Suppose  $f^n(\cdot)/f(\cdot)$  denotes the noisy/true objective function. Then

$$P(f^n(x) = f(x)) = 1 - p, \quad P(f^n(x) = f(y)) = p,$$

where  $p \in [0, 1]$  and  $y$  is derived by changing a uniformly selected bit in  $x$ .

To prove Theorem 2, we present Lemma 1 to analyze  $\hat{f}(x)$  under onebit noise by taking a sample size of  $2n^3 + 1$ . It intuitively means  $\hat{f}(x)$  is close to the 2-quantile of  $f^n(x)$  w.h.p.

**Lemma 1.** Under onebit noise, if median sampling with  $m = 2n^3 + 1$  is used, then

- (i) If  $p \cdot \frac{n-|x|_0}{n} \geq \frac{1}{2} + \frac{\Omega(1)}{n}$ , then  $P(\hat{f}(x) = n - |x|_0 - 1) \geq 1 - e^{-\Omega(n)}$ ;
- (ii) If  $p \cdot \frac{|x|_0}{n} \geq \frac{1}{2} + \frac{\Omega(1)}{n}$ , then  $P(\hat{f}(x) = n - |x|_0 + 1) \geq 1 - e^{-\Omega(n)}$ ;
- (iii) If  $p \cdot \frac{|x|_0}{n} \leq \frac{1}{2} - \frac{\Omega(1)}{n}$ , then  $P(\hat{f}(x) \leq n - |x|_0) \geq 1 - e^{-\Omega(n)}$ ; furthermore, if  $p \cdot \frac{n-|x|_0}{n} \leq \frac{1}{2} - \frac{\Omega(1)}{n}$  also holds, then  $P(\hat{f}(x) = n - |x|_0) \geq 1 - e^{-\Omega(n)}$ .

*Proof.* First we consider (i). Suppose  $|x|_0 = i$  and  $(n - i)p/n \geq 1/2 + c/n$  for a constant  $c$ . Suppose  $s$  denotes the number of noisy evaluations satisfying  $f^n(x) = n - 1 - i$  in  $m$  independent noisy evaluations. Observe that in each evaluation,  $P(f^n(x) = n - 1 - i) = (n - i)p/n$ , thus  $E(s) = m(n - i)p/n \geq m(1/2 + c/n) \geq m/2 + cn^2$ . Then we get

$$\begin{aligned} P(s \leq m/2) &= P(s \leq m/2 + cn^2 - cn^2) \leq P(s \leq E(s) - cn^2) \\ &\leq P(|s - E(s)| \geq cn^2) \leq 2e^{-2c^2n^4/m} = e^{-\Omega(n)}, \end{aligned} \tag{1}$$

where the last inequality is derived according to Hoeffding’s inequality. Therefore,

$$P(s > m/2) \geq 1 - e^{-\Omega(n)}. \tag{2}$$

By the definition of median sampling,  $P(\hat{f}(x) = n - |x|_0 - 1) \geq 1 - e^{-\Omega(n)}$ , thus the claim holds. We can similarly prove (ii).

Now we consider (iii). Under onebit noise,  $f^n(x)$  can take at most three values (i.e.,  $n - 1 - |x|_0$ ,  $n - |x|_0$ ,  $n + 1 - |x|_0$ ), thus  $\hat{f}(x)$  can only take one of the three values by the definition of median sampling and  $m = 2n^3 + 1$ . Note that  $P(f^n(x) \leq n - |x|_0) = 1 - P(f^n(x) = n + 1 - |x|_0) = 1 - |x|_0 p/n \geq 1/2 + \Omega(1)/n$ , then similar to case (i),  $P(\hat{f}(x) \leq n - |x|_0) \geq 1 - e^{-\Omega(n)}$ .

Then we consider the “furthermore” clause. By  $P(f^n(x) \geq n - |x|_0) = 1 - P(f^n(x) = n - 1 - |x|_0) = 1 - (n - |x|_0)p/n \geq 1/2 + \Omega(1)/n$ , we also derive  $P(\hat{f}(x) \geq n - |x|_0) \geq 1 - e^{-\Omega(n)}$ . Thus,  $P(\hat{f}(x) = n - |x|_0) = 1 - P(\hat{f}(x) > n - |x|_0) - P(\hat{f}(x) < n - |x|_0) \geq 1 - e^{-\Omega(n)}$ , i.e., the claim holds. Combining the above analysis, Lemma 1 holds.

**Theorem 2.** For OM under onebit noise, the ERT of (1+1)-EA employing median sampling with  $m = 2n^3 + 1$  is polynomial.

*Proof.* The main idea is applying Theorem 1. We consider three cases for  $p$  and in each case, we will design a distance function  $V(x)$  and we need to examine  $E(V_t - V_{t+1} | \xi_t = x)$  for  $x \neq 1^n$ . Suppose  $|x|_0 = i$ ,  $1 \leq i \leq n$ . For ease of notation, let  $P_{\text{mut}}(x, z) = P(z \text{ is mutated from } x)$ , and  $P_{\text{acc}}(x, z) = P(\hat{f}(z) \geq \hat{f}(x))$ . For ease of analysis, the drift is divided into  $E_1$  and  $E_2$ . That is,

$$E(V_t - V_{t+1} | \xi_t = x) = E_1 - E_2, \tag{3}$$

where

$$E_1 = \sum_{|z|_0 < i} P_{\text{mut}}(x, z) \cdot P_{\text{acc}}(x, z) \cdot (V(x) - V(z)), \tag{4}$$

$$E_2 = \sum_{|z|_0 > i} P_{\text{mut}}(x, z) \cdot P_{\text{acc}}(x, z) \cdot (V(z) - V(x)). \tag{5}$$

(1)  $p \leq n/(2(n + 1))$ .  $V(x)$  is designed to be  $|x|_0$ , namely the number of 0s in  $x$ . For  $E_1$ , we consider mutating only one zero bit in  $x$  (namely  $|z|_0 = i - 1$ ), and its probability is  $i/n \cdot (1 - 1/n)^{n-1} \geq i/(en)$ . Then  $z$  will replace  $x$  if  $\hat{f}(z) = n + 1 - i$  and  $\hat{f}(x) = n - i$ . Conditions of Lemma 1-(iii) hold because  $p \leq n/(2(n + 1)) = 1/2 - 1/(2(n + 1))$ , then

$$P_{\text{acc}}(x, z) \geq P(\hat{f}(z) = n + 1 - i, \hat{f}(x) = n - i) \geq 1 - e^{-\Omega(n)}. \tag{6}$$

Thus,

$$E_1 \geq \frac{i}{en} \cdot (1 - e^{-\Omega(n)}) \geq \frac{1}{3n}, \tag{7}$$

where the last inequality is by  $n$  is large enough. For  $E_2$ , we consider the increase of 0s. For  $z$  satisfying  $|z|_0 > i$ , accepting it implies  $\hat{f}(z) \neq n - |z|_0$  or  $\hat{f}(x) \neq n - i$ . Note that conditions of (iii) in Lemma 1 are satisfied, then we have

$$P_{\text{acc}}(x, z) \leq e^{-\Omega(n)}. \tag{8}$$

Thus,

$$E_2 \leq (n - i) \cdot e^{-\Omega(n)} \leq e^{\log n} \cdot e^{-\Omega(n)} = e^{-\Omega(n)}, \tag{9}$$

where the last equality is by  $n$  is large enough.

Subtract  $E_2$  from  $E_1$ , we get

$$E(V_t - V_{t+1} | \xi_t = x) \geq \Omega\left(\frac{1}{n}\right), \tag{10}$$

where the last equation derives from large enough  $n$ . Therefore, by Theorem 1,  $E(\tau|\xi_0) \leq n/\Omega(1/n) = O(n^2)$ , because  $V(x) \leq n$ . Note that each iteration needs  $2m = 4n^3 + 2$  fitness evaluations, we can derive a polynomial ERT.

(2)  $n/(2(n + 1)) < p < n/(n + 7)$ . The proof procedure is similar to case (1), but the  $V(x)$  is more complicated because the effect of the noise on a solution  $x$  may vary as  $|z|_0$  changes. The distance function is as follows:

$$V(x) = \begin{cases} i, & \text{if } i > \frac{n}{2p} + 3 \text{ or } n - \frac{n}{2p} + 3 < i < \frac{n}{2p} - 3 \text{ or } i < \max\left\{1, n - \frac{n}{2p} - 3\right\}, \\ \frac{n}{2p}, & \text{if } \frac{n}{2p} - 3 \leq i \leq \frac{n}{2p} + 3, \\ n - \frac{n}{2p} + 2, & \text{if } \max\left\{1, n - \frac{n}{2p} - 3\right\} \leq i \leq n - \frac{n}{2p} + 3. \end{cases}$$

We consider five cases for  $i$ .

(2a)  $i > n/(2p) + 3$ . For  $E_1$ , we also consider mutating only one zero bit in  $x$ , then  $P_{\text{mut}}(x, z) \geq i/(en)$ . Note that  $|z|_0 > n/(2p) + 2$ , thus  $pi/n > p|z|_0/n \geq 1/2 + 2p/n$ . By (ii) in Lemma 1,

$$P_{\text{acc}}(x, z) \geq P(\hat{f}(z) = n + 2 - i, \hat{f}(x) = n + 1 - i) \geq 1 - e^{-\Omega(n)}. \tag{11}$$

If  $|z|_0 > n/(2p) + 3$ , then  $V(x) - V(z) = 1$ ; else  $V(x) - V(z) = i - n/(2p) > 3$ . Thus,  $V(x) - V(z) \geq 1$  and  $E_1 = \Omega(1/n)$ .

Now we consider  $E_2$ . For  $z$  satisfying  $|z|_0 > i$ ,  $p|z|_0/n > pi/n > 1/2 + 3p/n \geq 1/2 + 3/(2(n + 1))$ . Thus, by (ii) in Lemma 1,

$$P(\hat{f}(z) = n + 1 - |z|_0, \hat{f}(x) = n + 1 - i) \geq 1 - e^{-\Omega(n)}, \tag{12}$$

implying  $P_{\text{acc}}(x, z) \leq e^{-\Omega(n)}$ . Accordingly,  $E_2 \leq e^{-\Omega(n)}$ .

(2b)  $n/(2p) - 3 \leq i \leq n/(2p) + 3$ . Note that  $i \geq n/(2p) - 3 \geq (n + 7)/2 - 3 = \Omega(n)$ . First we consider the positive drift  $E_1$ . By  $n/(2p) - 3 - (n - n/(2p) + 3) = n/p - n - 6 > 1$ , there always exists some  $z$  such that  $n - n/(2p) + 3 < |z|_0 = \lceil n/(2p) - 3 \rceil - 1$  and such  $z$  can be mutated from  $x$  by flipping at most seven 0s. Thus,

$$P_{\text{mut}}(x, z) \geq \binom{i}{7} \left(1 - \frac{1}{n}\right)^{n-7} \left(\frac{1}{n}\right)^7 = \Omega(1).$$

Note that  $p(n - |z|_0)/n < 1/2 - 3p/n$  and  $p|z|_0/n < 1/2 - 3p/n$ , thus  $P(\hat{f}(z) = n - i) \geq 1 - e^{-\Omega(n)}$  by (iii) in Lemma 1. Therefore,  $z$  will replace  $x$  with probability  $1 - e^{-\Omega(n)}$ . Moreover,

$$V(x) - V(z) \geq n/(2p) - |z|_0 > n/(2p) - (n/(2p) - 3) = 3,$$

we have  $E_1 = \Omega(1)$ .

For  $E_2$ , we consider  $z$  with  $|z|_0 > i$ . If  $i > n/(2p) + 1$ , Eq. (12) holds and we have  $P_{\text{acc}}(x, z) \leq e^{-\Omega(n)}$ , then we get

$$E_2 \leq (n - i) \cdot e^{-\Omega(n)} \leq e^{\log n} \cdot e^{-\Omega(n)} = e^{-\Omega(n)},$$

where the last equality is by  $n$  is large enough.

If  $i \leq n/(2p) + 1$ , then any  $z$  satisfying  $|z|_0 \geq i + 3$  will never be accepted under onebit noise. For  $z$  satisfying  $|z|_0 \leq i + 2$ , we have  $|z|_0 \leq n/(2p) + 3$ , thus  $V(z) = n/(2p) = V(x)$ . Then  $E_2 = 0$ . Combining the two cases for  $z$ , we get  $E_2 \leq e^{-\Omega(n)}$ .

(2c)  $n - n/(2p) + 3 < i < n/(2p) - 3$ . First we examine  $\hat{f}(x)$ . Note that  $p(n - i)/n < 1/2 - 3p/n$  and  $pi/n < 1/2 - 3p/n$ , thus  $P(\hat{f}(x) = n - i) \geq 1 - e^{-\Omega(n)}$  by (iii) in Lemma 1.

For  $E_1$ , we consider mutating only one zero bit in  $x$ , namely  $|z|_0 = i - 1 > n - n/(2p) + 2$ . Note that  $p(n - |z|_0)/n < 1/2 - 2p/n$  and  $p|z|_0/n < 1/2 - 3p/n$ , we can derive Eq. (6) and  $E_1 = \Omega(1/n)$ .

For  $E_2$ , we consider two cases for  $z$  satisfying  $|z|_0 > i$ . If  $|z|_0 \geq i + 2$ , accepting  $z$  implies  $\hat{f}(x) = n - i - 1$ . Thus,  $P_{\text{acc}}(x, z) \leq e^{-\Omega(n)}$ . If  $|z|_0 = i + 1$ , then  $p(n - |z|_0)/n < 1/2 - 4p/n$  and  $p|z|_0/n < 1/2 - 2p/n$ , thus  $P(\hat{f}(z) = n - |z|_0) \geq 1 - e^{-\Omega(n)}$ . Then Eq. (8) still holds, i.e.,  $P_{\text{acc}}(x, z) \leq e^{-\Omega(n)}$ . Combining the two cases,  $E_2 \leq e^{-\Omega(n)}$ .

(2d)  $\max\{1, n - n/(2p) - 3\} \leq i \leq n - n/(2p) + 3$ . First we consider the positive drift  $E_1$ . Note that

$$n - \frac{n}{2p} + 3 - \max\left\{1, n - \frac{n}{2p} - 3\right\} \leq n - \frac{n}{2p} + 3 - \left(n - \frac{n}{2p} - 3\right) \leq 6,$$

$x$  can generate an offspring  $z$  with  $|z|_0 < \max\{1, n - n/(2p) - 3\}$  by flipping at most seven bits, whose probability is at least  $\Omega(1/n^7)$ . Then we examine  $P_{\text{acc}}(x, z)$ . Because

$$\frac{pi}{n} \leq p \left(n - \frac{n}{2p} + 3\right) \frac{1}{n} = p \frac{n+3}{n} - \frac{1}{2} \leq \frac{n+3}{n+7} - \frac{1}{2} = \frac{1}{2} - \frac{4}{n+7},$$

we derive  $P(\hat{f}(x) \leq n - i) \geq 1 - e^{-\Omega(n)}$  by (iii) in Lemma 1. Note that  $\hat{f}(z) \geq n - i$  always holds under onebit noise, we get  $P_{\text{acc}}(x, z) \geq 1 - e^{-\Omega(n)}$ . If  $1 \leq n - n/(2p) - 3$ , we get

$$V(x) - V(z) \geq n - \frac{n}{2p} + 2 - \left(n - \frac{n}{2p} - 3\right) \geq 5;$$

else  $|z|_0 = 0$  and

$$V(x) - V(z) \geq n - \frac{n}{2p} + 2 > 1.$$

Thus,  $E_1 = \Omega(1/n^7)$ .

For  $E_2$ , it is only necessary to take  $z$  satisfying  $|z|_0 \leq i + 2$  into account because  $z$  with  $|z|_0 \geq i + 3$  will be rejected. If  $i \geq n - n/(2p) + 1$ , we have  $n - n/(2p) + 1 \leq i < |z|_0 \leq n - n/(2p) + 5$ . Thus,

$$p \cdot \frac{n - |z|_0}{n} < p \cdot \frac{n - i}{n} \leq \frac{1}{2} - \frac{p}{n},$$

and

$$p \cdot \frac{i}{n} < p \cdot \frac{|z|_0}{n} \leq \frac{(n+5)p}{n} - \frac{1}{2} < \frac{n+5}{n+7} - \frac{1}{2} = \frac{1}{2} - \frac{2}{n+7},$$

which implies that Eq. (8) holds by (iii) in Lemma 1. If  $i < n - n/(2p) + 1$ , we have  $|z|_0 < n - n/(2p) + 3$ . By  $V(x) = V(z) = n - n/(2p) + 2$ , we have  $E_2 = 0$ . Combining the two cases,  $E_2 \leq e^{-\Omega(n)}$ .

(2e)  $i < \max\{1, n - n/(2p) - 3\}$ . If  $1 \geq n - n/(2p) - 3$ , then  $i = 0$ , thus we only need to consider that  $1 < n - n/(2p) - 3$ , namely  $i < n - n/(2p) - 3$ .

For  $E_1$ , we consider mutating only one zero bit in  $x$ , i.e.,  $|z|_0 = i - 1$ . Similar to the above analysis,  $P_{\text{mut}}(x, z) \geq 1/(en)$ . Note that  $p(n - i)/n > 1/2 + 3p/n$ , we derive  $P(\hat{f}(x) = n - i - 1) \geq 1 - e^{-\Omega(n)}$  by (i) in Lemma 1. Thus,  $P_{\text{acc}}(x, z) \geq 1 - e^{-\Omega(n)}$ . Note that  $V(x) - V(z) = i - |z|_0 = 1$ , thus  $E_1 = \Omega(1/n)$ .

For  $E_2$ , it is only necessary to take  $z$  with  $|z|_0 \leq i + 2 < n - n/(2p) - 1$  into account. Note that  $p(n - |z|_0)/n > 1/2 + p/n$ , thus by Lemma 1,  $P(\hat{f}(z) = n - |z|_0 - 1) = 1 - e^{-\Omega(n)}$ , implying that  $P_{\text{acc}}(x, z) \leq e^{-\Omega(n)}$ . Then we have  $E_2 \leq e^{-\Omega(n)}$ .

Combining the five cases, we have  $E_1 \geq \Omega(1/n^7)$  and  $E_2 \leq e^{-\Omega(n)}$ . By subtracting  $E_2$  from  $E_1$ , Eq. (10) becomes

$$E(V_t - V_{t+1} | \xi_t = x) = \Omega\left(\frac{1}{n^7}\right), \tag{13}$$

and we can also derive a polynomial ERT.

(3)  $p \geq n/(n + 7)$ . The effect of the noise changes when the level of the noise changes. Accordingly, we need to design a new distance function:

$$V(x) = \begin{cases} i & \text{if } i > \frac{n}{2p} + 3 \text{ or } i < n - \frac{n}{2p} - 3, \\ \frac{n}{2} & \text{if } n - \frac{n}{2p} - 3 \leq i \leq \frac{n}{2p} + 3. \end{cases}$$

Next we consider three cases for  $i$ .

(3a)  $i > n/(2p) + 3$ . The proof procedure is the same as case (2a), except that “ $V(x) - V(z) = i - n/(2p) > 3$ ” changes to  $V(x) - V(z) = i - n/2 > n/(2p) + 3 - n/2 \geq 3$ . We derive  $E_1 = \Omega(1/n)$  and  $E_2 \leq e^{-\Omega(n)}$ .

(3b)  $n - n/(2p) - 3 \leq i \leq n/(2p) + 3$ . Note that  $i \geq n - n/(2p) - 3 = \Omega(n)$ . First we consider the positive drift  $E_1$ . There exists some  $z$  with  $|z|_0 = \lceil n - n/(2p) - 3 \rceil - 1$  and such  $z$  can be mutated from  $x$  by flipping at most  $n/(2p) + 3 - (n - n/(2p) - 3) + 1 = n/p - n + 6 + 1 \leq 14$  0s. Thus,

$$P_{\text{mut}}(x, z) \geq \binom{i}{14} \left(1 - \frac{1}{n}\right)^{n-14} \left(\frac{1}{n}\right)^{14} = \Omega(1).$$

If  $i \leq n - n/(2p) - 1$ , then  $p(n - i)/n \geq 1/2 + p/n$ . Thus,  $P_{\text{acc}}(x, z) \geq P(\hat{f}(x) = n - i - 1) \geq 1 - e^{-\Omega(n)}$  by (i) in Lemma 1. If  $i > n - n/(2p) - 1$ , it can be verified that  $z$  will always be accepted under onebit noise. Note that

$$V(x) - V(z) \geq n/2 - |z|_0 \geq n/2 - (n - n/(2p) - 3) \geq 3.$$

Thus, we have  $E_1 = \Omega(1)$ .

For  $E_2$ , the proof procedure is the same as that of case (2b), except that “ $V(z) = n/(2p) = V(x)$ ” changes to  $V(z) = n/2 = V(x)$ . Thus, we get  $E_2 \leq e^{-\Omega(n)}$ .

(3c)  $i < n - n/(2p) - 3$ . The analysis for  $E_1$  and  $E_2$  is the same as that of case (2e), then we have  $E_1 = \Omega(1/n)$  and  $E_2 \leq e^{-\Omega(n)}$ .

Combining the three cases, we have  $E_1 = \Omega(1/n)$  and  $E_2 \leq e^{-\Omega(n)}$ . Subtract  $E_2$  from  $E_1$ , we get

$$E(V_t - V_{t+1} | \xi_t = x) = \Omega\left(\frac{1}{n}\right), \tag{14}$$

and we can also derive a polynomial ERT.

By the above proof, we can give an intuitive explanation for the effectiveness of median sampling. For  $x$  and  $z$  which satisfy  $f(x) > f(z)$ , when the 2-quantile of  $f^n(x)$  is larger than that of  $f^n(z)$ ,  $x$  will be estimated better than  $z$  by median sampling w.h.p., implying a correct comparison.

### 4 Cases where median sampling is better than mean sampling

For OM under segmented noise (Definition 5), we show that (1+1)-EA equipped with median sampling can do better than (1+1)-EA using mean sampling. The segmented noise is from [25], but we make a little modification to simplify the analysis. As presented in Definition 5, the noisy evaluation of a solution  $x$  can be divided into three segments. The objective evaluation is accurate in the first segment, but inaccurate in other segments because of noise. We show that for OM under segmented noise, the ERT of (1+1)-EA using mean sampling is exponential (i.e., Theorem 3), and the ERT of (1+1)-EA employing median sampling with  $m = 2n^3 + 1$  is polynomial (i.e., Theorem 4). The analyses show that median sampling can be better if the 2-quantile increases with the true fitness.

**Definition 5.**  $\forall x \in \{0, 1\}^n$ , its noisy objective  $f^n(\cdot)$  is defined as follows:

- (1) if  $|x|_0 > \frac{n}{50}$ ,  $f^n(x) = n - |x|_0$ ;
- (2) if  $\frac{n}{100} < |x|_0 \leq \frac{n}{50}$ ,

$$P(f^n(x) = n - |x|_0) = \frac{1}{2} + \frac{1}{n}, \quad P(f^n(x) = 3n + |x|_0) = \frac{1}{2} - \frac{1}{n};$$

- (3) if  $|x|_0 \leq \frac{n}{100}$ ,

$$P(f^n(x) = 4n(n - |x|_0)) = 1 - \frac{1}{n}, \quad P(f^n(x) = (2n + |x|_0)^3) = \frac{1}{n},$$

where  $n/100 \in \mathbb{N}^+$ .

Theorem 3 shows that mean sampling fails under segmented noise and the reason is similar to that found in [25]. Consider  $x$  and  $z$  satisfying  $|z|_0 = |x|_0 + 1$ . In segment (2), a small sample cannot eliminate the impact of noise, and  $P(\hat{f}(x) \leq \hat{f}(z))$  is still very large. In segment (3), the expected gap between



$f^n(z)$  and  $f^n(x)$  is positive. Therefore, a larger sample size will enlarge  $P(\bar{f}(x) \leq \bar{f}(z))$  and performs worse; moreover, no medium sample size makes a good tradeoff. Therefore, mean sampling fails. Its rigorous proof can be derived directly from Theorem 5.2 in [25], because the change of noise does not affect the proof.

**Theorem 3.** For OM under segmented noise, the ERT of (1+1)-EA employing mean sampling is exponential.

To prove Theorem 4, Lemma 2 is used. This lemma can upper bound the runtime, when the true better solution has a large probability to be recognized as better. Note that  $x^j$  denotes some solution with  $j$  0s, and  $F(\cdot)$  denotes the estimated fitness of a solution.

**Lemma 2** ([18]). The EFHT of (1+1)-EA solving noisy OM is polynomial if

$$\forall 0 < i \leq j : P(F(x^j) \geq F(x^{i-1})) \leq \frac{\log n}{15n}. \tag{15}$$

We also present Lemma 3 to analyze  $\hat{f}(x)$  under segmented noise by taking a sample size of  $2n^3 + 1$ .

**Lemma 3.** Under segmented noise, if median sampling with  $m = 2n^3 + 1$  is used, then  $P(\hat{f}(x) = n - |x|_0) = 1 - e^{-\Omega(n)}$  if  $\frac{n}{100} < |x|_0 \leq \frac{n}{50}$ ;  $P(\hat{f}(x) = 4n(n - |x|_0)) = 1 - e^{-\Omega(n)}$  if  $|x|_0 \leq \frac{n}{100}$ .

*Proof.* The main procedure is analogous to Lemma 1. If  $\frac{n}{100} < |x|_0 \leq \frac{n}{50}$ , suppose there are  $s$  noisy evaluations where  $f^n(x) = n - |x|_0$  in  $m$  independent noisy evaluations. Then Eq. (2) also holds and  $P(\hat{f}(x) = n - |x|_0) = 1 - e^{-\Omega(n)}$ . If  $|x|_0 \leq \frac{n}{100}$ , we similarly have  $P(\hat{f}(x) = 4n(n - |x|_0)) = 1 - e^{-\Omega(n)}$ . Thus, Lemma 3 holds.

**Theorem 4.** For OM under segmented noise, the ERT of (1+1)-EA employing median sampling with  $m = 2n^3 + 1$  is polynomial.

*Proof.* The main idea is applying Lemma 2. Given  $0 < i \leq j$ , let  $g = \hat{f}(x^j) - \hat{f}(x^{i-1})$ . To analyze  $P(g \geq 0)$ , we consider four cases for  $i$ .

(1)  $i > \frac{n}{50}$ . Note that  $f^n(x^j) = f(x^j)$  and  $f^n(x^{i-1})$  is larger by considering  $i - 1 > \frac{n}{50}$  and  $i - 1 \leq \frac{n}{50}$ , respectively. Therefore, we get  $P(g \geq 0) = 0$ .

(2)  $\frac{n}{100} + 1 < i \leq \frac{n}{50}$ . If  $j > \frac{n}{50}$ , we have  $P(g \geq 0) = 0$ , because  $f^n(x^j) = n - j$  and  $f^n(x^{i-1}) \geq n - i + 1 > n - j$ . If  $j \leq \frac{n}{50}$ , by Lemma 3, we have

$$P(g < 0) \geq P(\hat{f}(x^j) = n - j, \hat{f}(x^{i-1}) = n - i + 1) = 1 - e^{-\Omega(n)}.$$

(3)  $i \leq \frac{n}{100} + 1$ . The analysis is analogous to case (2). If  $j > \frac{n}{100}$ , then  $P(g \geq 0) = 0$ . If  $j \leq \frac{n}{100}$ , then

$$P(g < 0) \geq P(\hat{f}(x^j) = 4n(n - j), \hat{f}(x^{i-1}) = 4n(n - i + 1)) = 1 - e^{-\Omega(n)}.$$

Combining the three cases, we have shown  $\forall 0 < i \leq j : P(g \geq 0) \leq \log n / (15n)$  for sufficiently large  $n$ . Then, by Lemma 2, the EFHT is polynomial. In each iteration, the algorithm needs  $2m = 4n^3 + 2$  evaluations, thus the ERT is polynomial.

## 5 Cases where mean sampling is better than median sampling

For OM under partial noise (Definition 6), we show that (1+1)-EA using median sampling is sometimes worse than (1+1)-EA using mean sampling. For partial noise presented in Definition 6, a false objective value is returned when  $|x|_0 < n/2$ . We prove that for OM under partial noise, the ERT of (1+1)-EA employing mean sampling with  $m = n^3$  is polynomial (i.e., Theorem 5), and the ERT of (1+1)-EA employing median sampling is exponential (i.e., Theorem 6). The analyses suggest that median sampling may fail if the 2-quantile of the noisy fitness does not increase with the true objective value, and it is better to choose other strategies.

**Definition 6.**  $\forall x \in \{0, 1\}^n$ , its noisy objective  $f^n(\cdot)$  is defined as follows:

- (1) if  $|x|_0 \geq \frac{n}{2}$ ,  $f^n(x) = n - |x|_0$ ;
- (2) if  $|x|_0 < \frac{n}{2}$ ,

$$P(f^n(x) = |x|_0/2) = 2/3, \quad P(f^n(x) = 2(n - |x|_0)) = 1/3.$$

**Theorem 5.** For OM under partial noise, the ERT of (1+1)-EA employing mean sampling with  $m = n^3$  is polynomial.

*Proof.* The main idea is applying Lemma 2. Given  $0 < i \leq j$ , let  $g = \bar{f}(x^{i-1}) - \bar{f}(x^j)$ . To analyze  $P(g \geq 0)$ , we classify  $i$  into two cases.

(1)  $i \geq \frac{n}{2} + 1$ . We have  $\bar{f}(x^{i-1}) = n - i + 1$  and  $\bar{f}(x^j) = n - j$ , thus  $P(g \leq 0) = 0$ .

(2)  $i < \frac{n}{2} + 1$ . First we need to derive  $\mu := E(g)$ . Note that  $E(\bar{f}(x^{i-1})) = E(f^n(x^{i-1})) = (i - 1)/2 \cdot 2/3 + 2(n - i + 1) \cdot 1/3 = (2n - i + 1)/3$ . We classify  $j$  into two cases. (a) If  $j \geq \frac{n}{2}$ , then  $E(\bar{f}(x^j)) = n - j$ , thus

$$E(g) = \frac{2n}{3} - \frac{i}{3} + \frac{1}{3} - n + j \geq -\frac{n}{3} + \frac{1}{3} + \frac{2j}{3} \geq \frac{1}{3}.$$

(b) If  $j < \frac{n}{2}$ , then  $E(\bar{f}(x^j)) = (2n - j)/3$  and  $E(g) = (j - i + 1)/3 \geq 1/3$ . Thus, we have  $\mu \geq 1/3$ . Then we have

$$P(g \leq 0) \leq P(|g - \mu| \geq \mu) \leq 2e^{-2m\mu^2/(2n)^2} \leq 2e^{-m/(18n^2)} = 2e^{-n/18},$$

where the second inequality holds by  $|f^n(x^{i-1}) - f^n(x^j)| \leq 2n$  and Hoeffding's inequality. Similar to the discussion at the end of Theorem 4, the ERT is polynomial.

From the proof, we can derive an intuitively explanation for the effectiveness of mean sampling. For  $x$  and  $z$  satisfying  $f(x) > f(z)$  (i.e.,  $z$  is worse), the expectation of  $f^n(x)$  is larger than  $f^n(z)$ . Then, there is a small enough probability to accept  $z$  if using mean sampling. Thus, the search direction of (1+1)-EA will not be misled and the optimal solution can be quickly found.

To prove Theorem 6, we use Lemma 4 [18], which intuitively means that if a true worse solution (i.e., a solution with more 0s) is estimated better than a true better solution with some probability, then we can derive the lower bound for the runtime.

**Lemma 4** ([18]). If there exists a real number  $l \leq n/4$  satisfying

$$\forall 0 < i \leq l : P(F(x^i) < F(x^{i-1})) \leq 1 - \frac{16i}{n}, \tag{16}$$

then w.h.p., the FHT of (1+1)-EA solving noisy OM is  $2^{\Omega(l)}$ .

**Theorem 6.** For OM under partial noise, the ERT of (1+1)-EA employing median sampling is exponential.

*Proof.* We use Lemma 4 for the proof. Given  $0 < i < \frac{n}{2}$ , let  $g = \hat{f}(x^i) - \hat{f}(x^{i-1})$ . First we show that  $P(\hat{f}(x^{i-1}) = (i - 1)/2) \geq 1/3$  for  $i < \frac{n}{2}$ . Suppose  $s$  denotes the number of noisy evaluations satisfying  $f^n(x^{i-1}) = (i - 1)/2$  in  $m$  independent noisy evaluations. We classify  $m$  into 2 cases.

(1)  $m$  is even. Let

$$A = \sum_{j=m/2+1}^m \binom{m}{j} \left(\frac{2}{3}\right)^j \left(\frac{1}{3}\right)^{m-j}, \quad B = \sum_{j=0}^{m/2-1} \binom{m}{j} \left(\frac{2}{3}\right)^j \left(\frac{1}{3}\right)^{m-j}, \quad C = \binom{m}{m/2} \left(\frac{2}{3}\right)^{m/2} \left(\frac{1}{3}\right)^{m/2}.$$

Note that the sum of the three items is 1, and  $A \geq B$ ,

$$A \geq \binom{m}{m/2+1} \left(\frac{2}{3}\right)^{m/2+1} \left(\frac{1}{3}\right)^{m/2-1} \geq C.$$

Thus,  $P(s \geq m/2 + 1) = A \geq 1/3$ . By definition of median sampling, we derive  $P(\hat{f}(x^{i-1}) = (i - 1)/2) \geq 1/3$ .

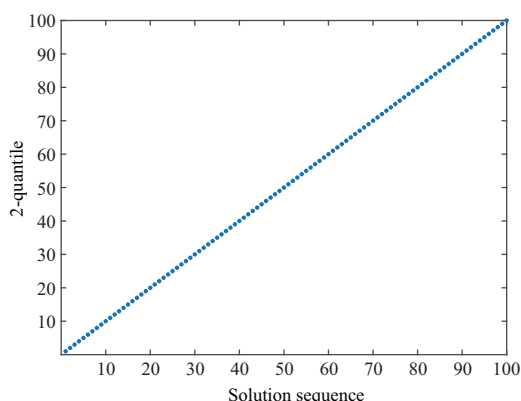
(2)  $m$  is odd. We have

$$P(s \geq (m + 1)/2) = \sum_{j=(m+1)/2}^m \binom{m}{j} \left(\frac{2}{3}\right)^j \left(\frac{1}{3}\right)^{m-j} \geq \sum_{j=0}^{(m-1)/2} \binom{m}{j} \left(\frac{2}{3}\right)^j \left(\frac{1}{3}\right)^{m-j}.$$

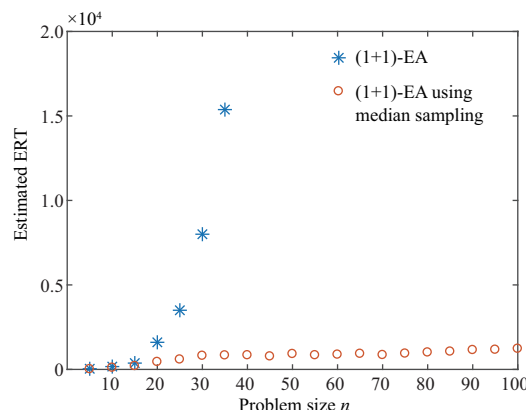
Thus,  $P(s \geq (m + 1)/2) \geq 1/2$ . By the definition of median sampling, we can derive that  $P(\hat{f}(x^{i-1}) = (i - 1)/2) \geq 1/2$ .

To make  $g \geq 0$ , it is sufficient that  $\hat{f}(x^{i-1}) = (i - 1)/2$  since it always holds that  $\hat{f}(x^i) \geq i/2$ . Thus,  $P(g \geq 0) \geq 1/3$  for  $i < \frac{n}{2}$ . Then, the condition of Lemma 4 holds by setting  $l = n/48$ . Thus, the EFHT is  $2^{\Omega(l)} = 2^{\Omega(n)}$ , i.e., exponential.

From the analysis, we can give an intuitive explanation for the failure of median sampling. Let us consider  $x$  and  $z$  satisfying  $|x|_0 = |z|_0 - 1$  (that is,  $z$  is worse), the 2-quantile of  $f^n(z)$  is larger than that of  $f^n(x)$ , and  $z$  will be estimated better than  $x$  by median sampling w.h.p., implying a wrong comparison.



**Figure 1** (Color online) 2-quantile of the noisy fitness of the sequence  $0^n, 10^{n-1}, \dots, 1^n$ . Note that  $i$  in  $x$ -axis corresponds to the  $(i + 1)$ -th solution  $1^i 0^{n-i}$  in the sequence.



**Figure 2** (Color online) Estimated ERT for (1+1)-EA with and without median sampling on OM under onebit noise.

## 6 Application illustration

In this section, we provide some guidance for employing median sampling in practice. The theoretical results have revealed that if the 2-quantile of the noisy fitness increases with the true fitness, we can use median sampling to tackle noise. Inspired by this finding, we may use the following three steps to check the effectiveness of median sampling in practice.

(1) Find a sequence of solutions with increasing true objective values. Note that the solution space can be very large, and we only need to find some representative solutions. The true fitness of a solution can be obtained by conducting evaluation accurately, instead of using an approximation. For example, a prediction model in machine learning can be evaluated using a large amount of data, and a structure in aerodynamic design can be evaluated by CFDs simulation. Note that the number of representative solutions is very limited, and the evaluation process can be easily parallelized, thus the computational cost is usually acceptable.

(2) Find an appropriate sample size  $m$ , such that the 2-quantile of the noisy fitness increases with the sequence. If such sample size does not exist or the sample size is too large, it would be better to choose other strategies.

(3) If finding such a sample size, evaluate each solution  $m$  times independently and output the median of the  $m$  objective values as the estimated fitness during the optimization procedure.

As an application illustration, we use (1+1)-EA to solve OM under onebit noise. It has been known that the ERT of (1+1)-EA solving OM under onebit noise is super-polynomial if the noise probability  $p = \omega(\log n/n)$ , thus we set  $p = \log^2 n/n$ . We set the problem size  $n = 100$  and use  $0^n, 10^{n-1}, \dots, 1^n$  as the sequence of solutions with increasing true fitness. We select the sample size  $m$  from 5, 10, 15,  $\dots$ , such that the 2-quantile of the noisy fitness increases with the sequence. Figure 1 shows that it holds when  $m = 15$ . Thus, using a sample size  $m = 15$  is probably enough to reduce the negative effect of noise.

To show the effectiveness of median sampling, we next compare the ERT of (1+1)-EA with and without median sampling for the problem size  $n \in \{5, 10, \dots, 100\}$ . For each  $n$ , we run (1+1)-EA 100 times independently. In each run, we record the number of fitness evaluations until an optimal solution with respect to the true fitness function is found for the first time. The total number of evaluations of the 100 runs are averaged as the estimation of the ERT. The results are shown in Figure 2. It can be observed that though using median sampling needs to evaluate a solution  $m$  times for estimating the fitness, the total number of evaluations required by (1+1)-EA to find the optimum is decreased drastically.

## 7 Conclusion

In this paper, we introduce median sampling into EAs to handle noise and theoretically analyze the effectiveness of median sampling. We first consider one classical case, i.e., OM under onebit noise, and show that median sampling can reduce the ERT of (1+1)-EA from exponential to polynomial. Next,

by two illustrative examples, we show that when the 2-quantile of the noisy fitness increases with the true objective value, median sampling is better than the commonly used mean sampling; otherwise, it is worse. The results provide us with some guidance to employ median sampling in practice. In the future, it would be interesting to analyze the effect of median sampling on real-world noisy optimization problems.

**Acknowledgements** This work was supported by National Key Research and Development Program of China (Grant No. 2017YFB1003102), National Natural Science Foundation of China (Grant Nos. 62022039, 61672478, 61876077), and MOE University Scientific-Technological Innovation Plan Program. The authors would like to thank the anonymous reviewers for their helpful comments and suggestions to this work.

## References

- 1 Bäck T. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*. Oxford: Oxford University Press, 1996
- 2 Xu P, Liu X, Cao H F, et al. An efficient energy aware virtual network migration based on genetic algorithm. *Front Comput Sci*, 2019, 13: 440–442
- 3 Yuan Q, Tang H B, You W, et al. Virtual network function scheduling via multilayer encoding genetic algorithm with distributed bandwidth allocation. *Sci China Inf Sci*, 2018, 61: 092107
- 4 Jin Y C, Branke J. Evolutionary optimization in uncertain environments — a survey. *IEEE Trans Evol Comput*, 2005, 9: 303–317
- 5 Aizawa A N, Wah B W. Scheduling of genetic algorithms in a noisy environment. *Evolary Comput*, 1994, 2: 97–122
- 6 Stagge P. Averaging efficiently in the presence of noise. In: *Proceedings of the 5th International Conference on Parallel Problem Solving from Nature*, Amsterdam, 1998. 188–197
- 7 Branke J, Schmidt C. Selection in the presence of noise. In: *Proceedings of the 5th ACM Conference on Genetic and Evolutionary Computation*, 2003. 766–777
- 8 Branke J, Schmidt C. Sequential sampling in noisy environments. In: *Proceedings of the 8th International Conference on Parallel Problem Solving from Nature*, 2004. 202–211
- 9 Auger A, Doerr B. *Theory of Randomized Search Heuristics: Foundations and Recent Developments*. Singapore: World Scientific, 2011
- 10 Neumann F, Witt C. *Bioinspired Computation in Combinatorial Optimization: Algorithms and Their Computational Complexity*. Berlin: Springer, 2010
- 11 Zhang Y S, Huang H, Wu H Y, et al. Theoretical analysis of the convergence property of a basic pigeon-inspired optimizer in a continuous search space. *Sci China Inf Sci*, 2019, 62: 070207
- 12 Hwang H-K, Witt C. Sharp bounds on the runtime of the (1+1) EA via drift analysis and analytic combinatorial tools. In: *Proceedings of the 15th International Workshop on Foundations of Genetic Algorithms*, 2019
- 13 Huang H, Su J P, Zhang Y S, et al. An experimental method to estimate running time of evolutionary algorithms for continuous optimization. *IEEE Trans Evol Comput*, 2020, 24: 275–289
- 14 Zhang Y A, Qin X F, Ma Q L, et al. Markov chain analysis of evolutionary algorithms on OneMax function — from coupon collector’s problem to (1+1) EA. *Theory Comput Sci*, 2020, 820: 26–44
- 15 Bian C, Qian C, Tang K. Towards a running time analysis of the (1+1)-EA for OneMax and LeadingOnes under general bit-wise noise. In: *Proceedings of the 15th International Conference on Parallel Problem Solving from Nature*, 2018. 165–177
- 16 Dang-Nhu R, Dardinier T, Doerr B, et al. A new analysis method for evolutionary optimization of dynamic and noisy objective functions. In: *Proceedings of the 20th ACM Conference on Genetic and Evolutionary Computation*, 2018. 1467–1474
- 17 Droste S. Analysis of the (1+1) EA for a noisy OneMax. In: *Proceedings of the 6th ACM Conference on Genetic and Evolutionary Computation*, 2004. 1088–1099
- 18 Gießen C, Kötzing T. Robustness of populations in stochastic environments. *Algorithmica*, 2016, 75: 462–489
- 19 Qian C, Yu Y, Zhou Z H. Analyzing evolutionary optimization in noisy environments. *Evolary Comput*, 2018, 26: 1–41
- 20 Sudholt D. On the robustness of evolutionary algorithms to noise: refined results and an example where noise helps. In: *Proceedings of the 20th ACM Conference on Genetic and Evolutionary Computation*, 2018. 1523–1530
- 21 Qian C, Shi J-C, Yu Y, et al. Subset selection under noise. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017. 3563–3573
- 22 Qian C. Distributed pareto optimization for large-scale noisy subset selection. *IEEE Trans Evol Comput*, 2020, 24: 694–707
- 23 Dang D-C, Lehre P K. Efficient optimisation of noisy fitness functions with population-based evolutionary algorithms. In: *Proceedings of the 13th International Workshop on Foundations of Genetic Algorithms*, 2015. 62–68
- 24 Prügel-Bennett A, Rowe J, Shapiro J. Run-time analysis of population-based evolutionary algorithm in noisy environments. In: *Proceedings of the 13th International Workshop on Foundations of Genetic Algorithms*, 2015. 69–75
- 25 Qian C, Bian C, Yu Y, et al. Analysis of noisy evolutionary optimization when sampling fails. In: *Proceedings of the 20th ACM Conference on Genetic and Evolutionary Computation*, 2018. 1507–1514
- 26 Qian C, Yu Y, Tang K, et al. On the effectiveness of sampling for evolutionary optimization in noisy environments. *Evolary*

- Comput, 2018, 26: 237–267
- 27 Qian C, Bian C, Jiang W, et al. Running time analysis of the (1+1)-EA for OneMax and leadingones under bit-wise noise. *Algorithmica*, 2019, 81: 749–795
  - 28 Friedrich T, Kötzing T, Krejca M S, et al. The compact genetic algorithm is efficient under extreme gaussian noise. *IEEE Trans Evol Comput*, 2017, 21: 477–490
  - 29 Doerr B, Hota A, Kötzing T. Ants easily solve stochastic shortest path problems. In: *Proceedings of the 14th ACM Conference on Genetic and Evolutionary Computation*, 2012. 17–24
  - 30 Feldmann M, Kötzing T. Optimizing expected path lengths with ant colony optimization using fitness proportional update. In: *Proceedings of the 12th International Workshop on Foundations of Genetic Algorithms*, 2013. 65–74
  - 31 Friedrich T, Kötzing T, Krejca M S, et al. Robustness of ant colony optimization to noise. *Evolary Comput*, 2016, 24: 237–254
  - 32 Sudholt D, Thyssen C. A simple ant colony optimizer for stochastic shortest path problems. *Algorithmica*, 2012, 64: 643–672
  - 33 Akimoto Y, Astete-Morales S, Teytaud O. Analysis of runtime of optimization algorithms for noisy functions over discrete codomains. *Theory Comput Sci*, 2015, 605: 42–50
  - 34 Huber P, Ronchetti M. *Robust Statistics*. Hoboken: John Wiley & Sons, 2009
  - 35 Leys C, Ley C, Klein O, et al. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J Exp Social Psychol*, 2013, 49: 764–766
  - 36 DeNavas-Walt C, Proctor B, Smith J. *Income, poverty, and health insurance coverage in the United States: 2011*. U.S. Census Bureau, 2012
  - 37 Doerr B, Sutton A. When resampling to cope with noise, use median, not mean. In: *Proceedings of the 21st ACM Conference on Genetic and Evolutionary Computation*, 2019. 242–248
  - 38 Droste S, Jansen T, Wegener I. On the analysis of the (1+1) evolutionary algorithm. *Theory Comput Sci*, 2002, 276: 51–81
  - 39 He J, Yao X. Drift analysis and average time complexity of evolutionary algorithms. *Artif Intell*, 2001, 127: 57–85