Selection in the Presence of Noise

Jürgen Branke and Christian Schmidt

Institute AIFB, University of Karlsruhe, 76128 Karlsruhe, Germany
{branke|csc}@aifb.uni-karlsruhe.de

Abstract. For noisy optimization problems, there is generally a trade-off between the effort spent to reduce the noise (in order to allow the optimization algorithm to run properly), and the number of solutions evaluated during optimization. However, for stochastic search algorithms like evolutionary optimization, noise is not always a bad thing. On the contrary, in many cases, noise has a very similar effect to the randomness which is purposefully and deliberately introduced e.g. during selection. Using the example of stochastic tournament selection, we show that the noise inherent in the optimization problem should be taken into account by the selection operator, and that one should not reduce noise further than necessary.

Keywords: Noise, tournament selection, stochastic fitness

1 Introduction

Many real-world optimization problems are noisy, i.e. a solution's quality (and thus the fitness function) is a random variable. Examples include all applications where the fitness is determined by a stochastic computer simulation, or where fitness is measured physically and prone to measuring error. Researchers have long argued that evolutionary algorithms (EAs) should be relatively robust against noise (see e.g. [FG88]), and recently a number of publications have appeared which support that claim at least partially [MG96,AB00a,AB00b,AB03].

For most noisy optimization problems, the uncertainty in fitness evaluation can be reduced by sampling an individual's fitness several times and using the average as estimate for the true mean fitness. Sampling n times reduces a random variable's standard deviation by a factor of $\sqrt{n}$, but on the other hand increases the computation time by a factor of n. Thus, there is a generally perceived trade-off: either one can use relatively exact estimations but only evaluate a small number of individuals (because a single estimation requires many evaluations), or one can let the algorithm work with relatively crude fitness estimations, but allow for more evaluations (as each estimation requires less effort). Generally, noise is considered harmful, as it may mislead the optimization algorithm. The main issue is probably the selection step: If due to the noise, a bad individual is evaluated better than it actually is, and/or a good individual is evaluated worse than its true fitness, the EA may wrongly select the worse individual although (according to the algorithmic design) it should have selected the better individual. Clearly, if such errors happen too frequently, optimization stagnates.

However, noise is not always a bad thing, on the contrary. EAs are randomized search algorithms, which use deliberate randomness to purposefully introduce errors into the selection process, primarily in order to get out of local minima. Therefore, in this paper we argue that it should be possible to accept the noise inherent in the optimization problem and to use it to (at least partially) replace the randomness in the optimization algorithm. As a result, it is possible to get the optimization algorithm to behave closer to its behavior on deterministic problems, even without excessive sampling. Furthermore, we will demonstrate that, depending on the fitness values and variances, noise affects some tournaments much stronger than others. As a consequence, we suggest a simple but effective resampling strategy to adapt the sample size to the specific tournament, allowing us to again get closer to the algorithm's behavior in a deterministic setting, while drastically reducing the number of samples required.

The paper is structured as follows: In Section 2, we survey some related work on EAs applied to noisy optimization problems, followed by a brief description of stochastic tournament selection. Section 4 demonstrates the effect noise has on tournament selection, and describes two ways to integrate a possible sampling error into the selection procedure. The idea of adapting not only the selection probability but also the sample size is discussed in Section 5. The paper concludes with a summary and some ideas for future work.

2   Related Work

The application of EAs in noisy environments has been the focus of many research papers. There are several papers that have looked at the trade-off between population size and sample size to estimate an individual's fitness, with sometimes conflicting results. Fitzpatrick and Grefenstette [FG88] conclude that for the genetic algorithm studied, it is better to increase the population size than the sample size. On the other hand, Beyer [Bey93] shows that for a $(1, \lambda)$ evolution strategy on a simple sphere, one should increase the sample size rather than λ. Hammel and Bäck [HB94] confirm these results and empirically show that it also doesn't help to increase the parent population size μ. Finally, Arnold and Beyer [AB00a,AB00b] show analytically that for the simple sphere, increasing the parent population size μ is helpful in combination with intermediate multirecombination. Miller [Mil97,MG96] has developed some simplified theoretical models which allow to simultaneously optimize the population size and the sample size. A good overview of theoretical work on EAs applied to noisy optimization problems can be found in [Bey00] or [Arn02].

All papers mentioned so far assume that the sample size is fixed for all individuals. Aizawa and Wah [AW94] were probably the first to suggest that the sample size could be adapted during the run, and suggested two adaptation schemes: increasing with the generation number, and higher sample size for individuals with higher estimated variance. Albert and Goldberg [AG01] look at a slightly different problem, but also conclude that the sample size should increase over the run. For (μ, λ) or $(\mu + \lambda)$ selection, Stagge [Sta98] has suggested basing

the sample size on an individual's probability to be among the μ best (and thus to survive to the next generation).

Branke et al. [Bra98,BSS01] and Sano and Kita [SK00,SKKY00] propose taking the fitness estimations of neighboring individuals into account when estimating an individual's fitness. This improves the estimation without requiring additional samples.

Finally, another related subject is that of searching for robust solutions, where instead of a noisy fitness function the decision variables are perturbed (cf. [TG97, Bra98,Bra01]).

3   Stochastic Tournament Selection

Stochastic tournament selection (STS) [GD91] is a rather simple selection scheme where two individuals are randomly chosen from the population, and then the better is selected with probability $(1 - \gamma)$. If individuals are sorted from rank 1 (best) to rank m (worst), this results in a linearly decreasing selection probability for an individual on rank i, with the slope of the line being determined by the selection probability $(1 - \gamma)$.

4   Selection Based on a Fixed Sample Size

Selecting the better of two individuals with probability $(1 - \gamma)$ in a noisy environment can be achieved in two fundamental ways: The standard way would be to eliminate the noise as much as possible by using a large number of samples, and then selecting the better individual with probability $(1 - \gamma)$. The noise-adapted selection proposed here has a different philosophy: instead of eliminating the noise and then artificially introducing randomness, we propose accepting a higher level of noise, and only add a little bit of randomness to achieve the desired behavior.

In the following, we will start with the standard STS, demonstrate the consequences in a noisy environment, and then develop a simple and a more complex model to get closer to the ideal noise-adapted selection.

4.1   Basic Notations

Let us denote the two individuals to be compared as x and y. If the fitness is noisy, the fitness of individual x (y) is a random variable F_x (F_y) with $F_x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ $(F_y \sim \mathcal{N}(\mu_y, \sigma_y^2))$[1]. If $\mu_x > \mu_y$, we would like to select individual x with probability $(1-\gamma)$ and vice versa. However, μ_x and μ_y are unknown, we can only estimate them by sampling each individual's fitness a number of n times

---

[1] Note that it will be sufficient to assume that the average difference obtained from sampling the individuals' fitnesses n times is normally distributed. This is certainly valid if each individual's fitness is normally distributed, but also independent of the actual fitness distributions for large enough n (central limit theorem).

and using the averages $\bar{f}_x$ and $\bar{f}_y$ as estimators for the fitnesses, and the sample variances s_x^2 and s_y^2 as estimators for the true variances.

If the actual fitness difference between the two individuals is denoted as $\delta = \mu_x - \mu_y$, the observed fitness difference $D = \bar{f}_x - \bar{f}_y$ is again a random variable $D \sim \mathcal{N}(\delta, \sigma_d^2)$. The variance of D depends on the number of samples drawn from each individual, n, and can be calculated as $\sigma_d^2 = (\sigma_x^2 + \sigma_y^2)/n$. A specific realization of the observed fitness difference is named d. Furthermore, we will need a standardized observed fitness which we define as $d^* = d/\sqrt{s_d^2}$ where $s_d^2 = (s_x^2 + s_y^2)/n$ is the unbiased estimated standard deviation of the fitness difference. The corresponding true counterpart is $\delta^* = \delta/\sigma_d$. Note that non-linear transformations of unbiased estimators are no longer unbiased, therefore d^* is a biased estimator for δ^*.

While γ is the *desired* selection probability for the *truely* worse individual, we denote with β the *implemented* probability for choosing the worse individual based on the estimated standardized fitness difference d^*, and $\xi(\delta^*, \beta)$ the actual selection probability for the *better* individual given a true standardized fitness difference of δ^*.

4.2   Standard Stochastic Tournament Selection

The simplest (and standard) way to apply STS would be to ignore the uncertainty in evaluation by making the following assumption:

Assumption: The observed fitness difference is equal to the actual fitness difference, i.e. $d = \delta$.

As a consequence, individual x is selected with probability $(1 - \beta) = (1 - \gamma)$ if $d \geq 0$ and with probability $\beta = \gamma$ if $d < 0$.

However, there can be two sources of error: Either we observe a fitness difference $d > 0$ when actually $\delta < 0$, or vice versa. The corresponding error probability α can be calculated as

$$
\alpha = \begin{cases} P(D > 0) = 1 - \Phi\left(\frac{-\delta}{\sigma_d}\right) = \Phi\left(\frac{\delta}{\sigma_d}\right) & : \quad \delta \leq 0 \\ P(D < 0) = \Phi\left(\frac{-\delta}{\sigma_d}\right) & : \quad \delta > 0 \end{cases}
$$

$$
= \Phi\left(\frac{-|\delta|}{\sigma_d}\right) = \Phi\left(-|\delta^*|\right) \tag{1}
$$

with Φ being the cumulative distribution function for a standard gaussian.

The overall selection probability for individual x can then be calculated as

$$
\xi = P(D > 0)(1 - \beta) + P(D < 0)\beta
$$

$$
= (1 - \alpha)(1 - \beta) + \alpha\beta \tag{2}
$$

Example: To visualize the effect of the error probability on the actual selection probability ξ, let us consider an example with $\sigma_x^2 = \sigma_y^2 = 10, n = 20$ and $\gamma = 0.2$. The actual selection probability for individual x depending on δ^* can be determined by a Monte Carlo simulation. We did this in the following way: For

a given δ^*, we generated 100,000 realizations of d^* according to $d^* = \frac{\bar{f}_x - \bar{f}_y}{\sqrt{(s_x^2 + s_y^2)/n}}$ based on $F_x \sim \mathcal{N}(0, \sigma_x^2)$, $F_y \sim \mathcal{N}(-\delta^* \sigma_d, \sigma_y^2)$. For each observed d^*, we select x with probability $(1 - \beta)$ if $d^* > 0$ and with probability β otherwise. The actual selection probability $\xi(\delta^*, \beta)$ is then the fraction of times x has been selected.
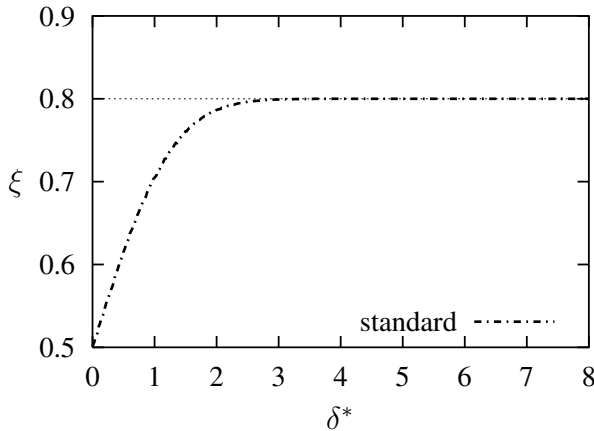

Fig. 1. True selection probability of individual x depending on the actual standardized fitness difference δ^*. The dotted line represents the desired selection probability $(1 - \gamma)$.

Figure 1 depicts the resulting true selection probability of individual x depending on the actual standardized fitness difference δ^*. The dotted line corresponds to the desired behavior in the deterministic case, the bold line labeled "standard" is the actual selection probability due to the noise. As can be seen, the actual selection probability for the better individual largely depends on the ratio δ^* of the fitness difference δ and the amount of noise measured as σ_d. While it corresponds to the desired selection probability of $(1 - \gamma)$ for $\delta^* > 3$, it approaches 0.5 for $\delta^* \to 0$. The latter fact is unavoidable, since for $\delta^* \to 0$ it becomes basically impossible to determine the better of the two individuals. The interesting question is how quickly ξ approaches $1 - \gamma$, and whether this behavior can be improved. Note that we only show the curves for $\delta^* \geq 0$ (assuming without loss of generality that $\mu_x > \mu_y$). For $\delta^* < 0$ the curve would be symmetric to $(0, 0.5)$.

In previous papers, it has been noted that the effect of noise on EAs is similar to a smaller selection pressure (e.g. [Mil97]). Figure 1 demonstrates that this is not entirely true for STS. A lower selection pressure in form of a higher γ would change the level of the dotted line, but it would still be horizontal, i.e. the selection probability for the better individual would be independent of the actual fitness difference. With noise, only the tournaments between individuals

of similar fitness are affected. Hence, a dependence on the actual fitness values is introduced which somehow contradicts the idea of rank-based selection.

4.3 A Simple Correction

If we know that our conclusion about which of the two individuals has a better fitness is prone to some error, it seems straightforward to take this error probability into account when deciding which individual to select. Instead of always selecting the better individual with probability $(1-\gamma)$, we could try to replace γ by a function $\beta(d^*)$ which depends on the standardized observed difference d^*.

Let us make the following assumption:

Assumption: It is possible to accurately estimate the error probability α.

Then, since we would like to have an overall true selection probability of $(1-\gamma)$, an appropriate β-function could be derived as

$$(1-\alpha)(1-\beta) + \alpha\beta \stackrel{!}{=} (1-\gamma) \tag{3}$$
$$1 - \beta - \alpha + \alpha\beta + \alpha\beta = (1-\gamma)$$
$$\beta(-1 + 2\alpha) = (1-\gamma) - 1 + \alpha$$
$$\beta = \frac{\gamma - \alpha}{1 - 2\alpha}. \tag{4}$$

β is a probability and can not be smaller than 0, i.e. the above equation assumes $\alpha \leq \gamma < 0.5$. For $\alpha > \gamma$ we set $\beta = 0$.

Unfortunately, α can not be calculated using Equation 1, because we don't know either δ nor σ_d.
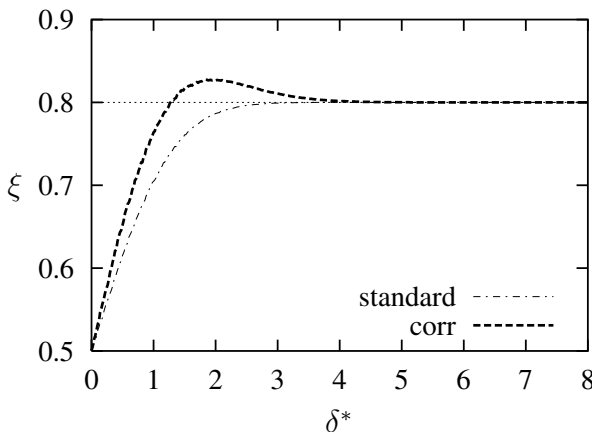

Fig. 2. True selection probability of individual x depending on the actual standardized fitness difference δ^*. The dotted line represents the desired selection probability $(1-\gamma)$.

It seems straightforward then to estimate δ by the observed difference d, and σ_d^2 by the observed variance s_d^2. Then, α is estimated as $\hat{\alpha} = \Phi(-|d|/s_d) = \Phi(-|d^*|)$, which is only a biased estimator due to the non-linear transformations. Nevertheless, this may serve as a reasonable first approximation of an optimal β-function. Figure 3 visualizes this β-function (labeled as "corr"). As can be seen, the probability to select the worse individuals decreases when the standardized difference d^* becomes small, and is 0 for $|d^*| < -\Phi^{-1}(\gamma)$ (i.e. the observed better individual is always selected if the observed standardized fitness difference d^* is small). Assuming the same parameters as in the example above, the resulting true selection probabilities $\xi(\delta^*, \beta(.))$ are depicted in Figure 2 (labeled as "corr"). The true selection probability approaches the desired selection probability faster than with the standard approach, but then it overshoots before it converges towards $(1 - \gamma)$. Nevertheless, the approximation is already much better than the standard approach (assuming a uniform distribution of δ^*).

4.4   Bootstrapping

The β-function proposed above can be further improved by bootstrapping [Efr90]. This method compares the observed selection probabilities p given the current β function with the desired selection probabilities, and then reduces β where the selection probability is too low, and increases β where the selection probability is too high. The observed selection probabilities $\xi(\delta^*, \beta(.))$ have to be estimated by Monte Carlo simulation, generating realisations of d^* and then selecting according to $\beta(d^*)$. Unfortunately, the distribution of d^* depends on the variance σ_d^2 of the observed fitness difference which is unknown. Therefore, in this approach we make the following simplifying assumption:

Assumption: The estimated variance of the difference corresponds to the true variance of the difference, i.e. $s_d^2 = \sigma_d^2$.

From that it follows that d^* is normally distributed according to $\mathcal{N}(\delta^*, 1)$.

More specifically, our bootstrapping approach starts with an initial $\beta_0(z)$ which corresponds to the β function defined in the section above. Then, it iteratively adapts beta according to

$$\beta_{t+1}(z) = \beta_t(z) + \xi(z, \beta_t(.)) - (1 - \gamma). \tag{5}$$

This procedure can be iterated until one is satisfied with the outcome.

The resulting β-function is depicted in Figure 3. At first sight, the strong fluctuations seems surprising. However, a steeper ascent of the true selection probability can only be achieved by keeping $\beta(d^*) = 0$ for as long as possible. The resulting overshoot then has to be compensated by a very high β etc. such that in the end, an oscillating acceptance pattern emerges as optimal. The corresponding true selection probabilities $\xi(\delta^*)$ are shown in Figure 4. As can be seen, despite the oscillating β-function, this curve is very smooth, and much closer to the actually desired selection probability of γ resp. $(1 - \gamma)$ than either the standard approach of ignoring the noise, or the first approximation of an appropriate β-function presented in the previous section.
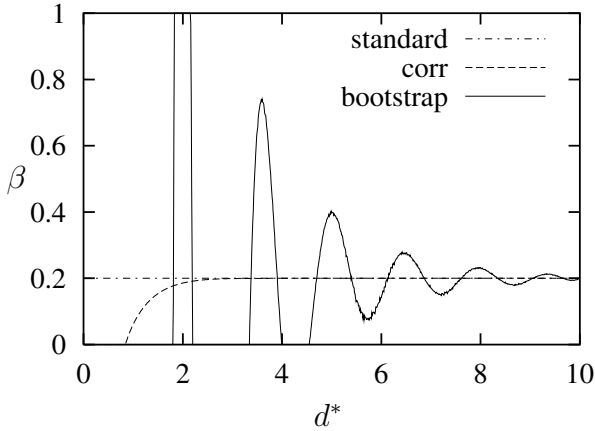
Fig. 3. The probability to select the worse individual (β-function), depending on the observed standardized fitness difference $d*$. Results of the different approaches.
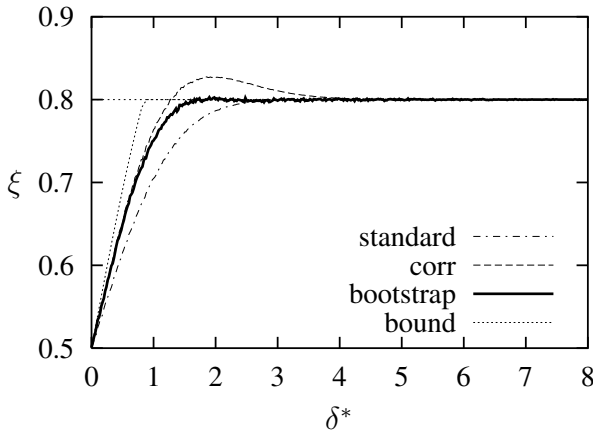

Fig. 4. True selection probability of individual x depending on the actual standardized fitness difference δ^*. The line denoted by "bound" is an idealized curve which depicts a limit to how close one can get to the desired selection probability. The dotted line represents the desired selection probability $(1 - \gamma)$.

Even though the bootstrapping method yields a much better approximation to the desired selection probability than the other two approaches, it could perhaps be further improved by basing it not only on d^* but on all three observed variables, namely d, σ_x^2, and σ_y^2. However, we expect that the additional improvement would be rather small. Furthermore, there is a bound to how close one can get to the desired selection probability: the steepest possible ascent of the true selection probability is clearly obtained if the individual with the higher

observed fitness is *always* selected. However, as long as α exceeds γ, the resulting true selection probability would still be below the desired selection probability. The corresponding steepest ascent curve is also shown in Figure 4 and denoted as "bound".

Instead of trying to further improve the estimation, we will now turn to the idea of drawing additional samples if the probability for a selection error is high.

5   Resampling

From the above discussion, it is clear that the deviation from actual selection probability to desired selection probability is only severe for small values of δ/σ_d, i.e. if the individuals have similar fitness and/or the noise is large.

Therefore, we now attempt to counteract that problem by adapting the number of samples to the expected error probability, i.e. by drawing a large number of samples whenever we assume that the selection error would be high and vice versa. We propose to do that in the following way: Starting with a reduced number of 10 samples for every individual, we calculate d^*. If $|d^*| \geq \epsilon$ where ϵ is a constant, we stop and use d^* to decide which individual to select. Otherwise, we repeatedly draw another sample for each of the two individuals until either $|d^*| \geq \epsilon$ or the total number of samples exceeds a maximum number N. For our experiments, we set $N = 100$ and $\epsilon = 1.33$, which approximately yields an error probability of 1% if $\delta^* = 1$ assuming that d^* is normally distributed as $d^* \sim \mathcal{N}(\delta^*, 1)$, i.e. if $\delta^* = 1$, there is only a 1% chance that we will observe a distance $d < 0$.

For our standard example with $\sigma_x^2 = \sigma_y^2 = 10$ and $\gamma = 0.2$, the above sampling scheme results in an average number of samples depending on δ^* as depicted in Figure 5. For small standardized distances d^*, the average number of samples is quite high, but it drops quickly and approaches the lower limit of 20 for $\delta^* > 3$. Depending on the distribution of δ^* in a real EA, this sampling scheme is thus able to achive tremendous savings compared to the fixed sampling rate of 20 samples per individual (40 samples in total). Furthermore, the actual selection probabilities using this sampling scheme are much closer to the desired selection probability than if a fixed number of samples is used. The two sampling schemes in combination with standard STS are compared in Figure 6.

Just as for the fixed sample size, we can apply bootstrapping also to the adaptive sampling scheme. The resulting β-function and selection probabilities are depicted in Figures 7 and 8. The resulting beta-function is much smoother than the one obtained for the fixed sampling scheme. Also, although there is still a clear benefit of bootstrapping with respect to the deviation of ξ from the desired $(1-\gamma)$, the improvement over standard STS is significantly smaller than with a fixed sample size. This is probably because due to the smaller initial sample size in combination with the resampling scheme used, our assumption that D^* is normally distributed may be less appropriate.
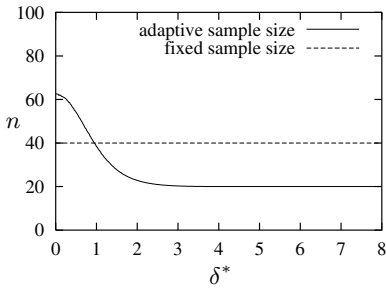
Fig. 5. Average sample size depending on the actual standardized fitness difference δ^*, with the fixed sampling scheme (dashed line) and the adaptive sampling scheme (solid line).
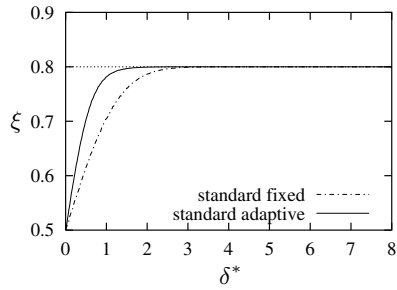

Fig. 6. Actual sampling probability depending on the actual standardized fitness difference δ^*, for the standard stochastic tournament selection with fixed and with adaptive sampling scheme.
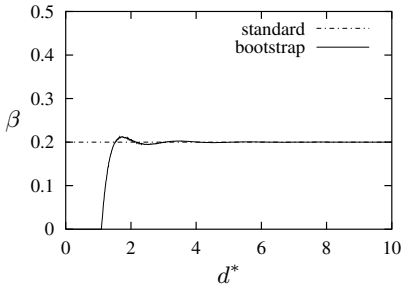

Fig. 7. β-function derived by bootstrapping for the case of an adaptive sample size.
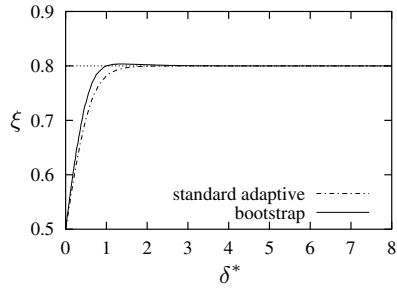

Fig. 8. Comparison of the actual sampling probability depending on the actual standardized fitness difference δ^* for the standard STS and the bootstrapping approach, when using the adaptive sampling scheme.

6   Conclusion

In this paper, we have argued that the error probability due to a noisy fitness function should be taken into account in the selection step. At the example of stochastic tournament selection, we have demonstrated that it is possible to obtain a much better match between actual and desired selection probability for an individual.

In a first step, we have derived two models which determine the selection probability for the better individual depending on the observed fitness difference. The simple model was based on some simplifying assumptions regarding

the distribution of the error probability; the second model was based on boot-strapping.

In a second step, we looked at a different sampling scheme, namely adapting the number of samples to the expected error probability. That way, a pair of similar individuals is sampled much more often than a pair of individuals with very different fitness values. This approach also greatly improves the accuracy of the actual selection probability. Additionally, depending on the distribution of fitness differences in an actual EA run, it will significantly reduce the number of samples required.

We are currently exploring a number of different extensions. For one, it should be relatively straightforward to extend our framework to other selection schemes and even to other heuristics like simulated annealing. Furthermore, we intend to improve the adaptive sampling scheme by using statistical test theory.

References

[AB00a]   D. V. Arnold and H.-G. Beyer. Efficiency and mutation strength adaptation of the $(\mu/\mu_i, \lambda)$-es in a noisy environment. In Schoenauer et al. [SDR$^+$00], pages 39–48.

[AB00b]   D. V. Arnold and H.-G. Beyer. Local performance of the $(\mu/\mu_i, \lambda)$-es in a noisy environment. In W. Martin and W. Spears, editors, *Foundations of Genetic Algorithms*, pages 127–142. Morgan Kaufmann, 2000.

[AB03]   D. V. Arnold and H.-G. Beyer. A comparison of evolution strategies with other direct search methods in the presence of noise. *Computational Optimization and Applications*, 24:135–159, 2003.

[AG01]   L. A. Albert and D. E. Goldberg. Efficient evaluation genetic algorithms under integrated fitness functions. Technical Report 2001024, Illinois Genetic Algorithms Laboratory, Urbana-Champaign, USA, 2001.

[Arn02]   D. V. Arnold. *Noisy Optimization with Evolution Strategies*. Kluwer, 2002.

[AW94]   A. N. Aizawa and B. W. Wah. Scheduling of genetic algorithms in a noisy environment. *Evolutionary Computation*, pages 97–122, 1994.

[Bey93]   H.-G. Beyer. Toward a theory of evolution strategies: Some asymptotical results from the $(1 \overset{+}{,} \lambda)$-theory. *Evolutionary Computation*, 1(2):165–188, 1993.

[Bey00]   H.-G. Beyer. Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice. *Computer methods in applied mechanics and engineering*, 186:239–267, 2000.

[Bra98]   J. Branke. Creating robust solutions by means of an evolutionary algorithm. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature*, volume 1498 of *LNCS*, pages 119–128. Springer, 1998.

[Bra01]   J. Branke. *Evolutionary Optimization in Dynamic Environments*. Kluwer, 2001.

[BSS01]   J. Branke, C. Schmidt, and H. Schmeck. Efficient fitness estimation in noisy environments. In L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. H . Garzon, and E. Burke, editors, *Genetic and Evolutionary Computation Conference*, pages 243–250. Morgan Kaufmann, 2001.

[Efr90]   B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans.* SIAM, 1990.

[FG88]   J. M. Fitzpatrick and J. J. Grefenstette. Genetic algorithms in noisy environments. *Machine Learning*, 3:101–120, 1988.

[GD91]   D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In G. Rawlins, editor, *Foundations of Genetic Algorithms*, San Mateo, CA, USA, 1991. Morgan Kaufmann.

[HB94]   U. Hammel and T. Bäck. Evolution strategies on noisy functions, how to improve convergence properties. In Y. Davidor, H. P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature*, volume 866 of *LNCS*. Springer, 1994.

[MG96]   B. L. Miller and D. E. Goldberg. Genetic algorithms, selection schemes, and the varying effects of noise. *Evolutionary Computation*, 4(2):113–131, 1996.

[Mil97]   Brad L. Miller. *Noise, Sampling, and Efficient Genetic Algorithms.* PhD thesis, Dept. of Computer Science, University of Illinois at Urbana-Champaign, 1997. available as TR 97001.

[SDR+00]   M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, editors. *Parallel Problem Solving from Nature*, volume 1917 of *LNCS*. Springer, 2000.

[SK00]   Y. Sano and H. Kita. Optimization of noisy fitness functions by means of genetic algorithms using history of search. In Schoenauer et al. [SDR+00], pages 571–580.

[SKKY00]   Y. Sano, H. Kita, I. Kamihira, and M. Yamaguchi. Online optimization of an engine controller by means of a genetic algorithm using history of search. In *Asia-Pacific Conference on Simulated Evolution and Learning.* Springer, 2000.

[Sta98]   P. Stagge. Averaging efficiently in the presence of noise. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, editors, *Parallel Problem Solving from Nature V*, volume 1498 of *LNCS*, pages 188–197. Springer, 1998.

[TG97]   S. Tsutsui and A. Ghosh. Genetic algorithms with a robust solution searching scheme. *IEEE Transactions on Evolutionary Computation*, 1(3):201–208, 1997.