Using statistical inference for designing termination conditions ensuring convergence of Evolutionary Algorithms

David Roche[1], Debora Gil [2] and Jesus Giraldo[1]
[1]Laboratory of Systems Pharmacology and Bioinformatics, UAB, Bellaterra,Spain
[2] Computer Science Department and Computer Vision Center (CVC), UAB, Bellaterra,Spain
{david.roche,jesus.giraldo}@uab.es,debora@cvc.uab.es

Abstract

A main challenge in Evolutionary Algorithms (EAs) is determining a termination condition ensuring stabilization close to the optimum in real-world applications. Although for known test functions distribution-based quantities are good candidates (as far as suitable parameters are used), in real-world problems an open question still remains unsolved. How can we estimate an upper-bound for the termination condition value ensuring a given accuracy for the (unknown) EA solution?

We claim that the termination problem would be fully solved if we defined a quantity (depending only on the EA output) behaving like the solution accuracy. The open question would be, then, satisfactorily answered if we had a model relating both quantities, since accuracy could be predicted from the alternative quantity. We present a statistical inference framework addressing two topics: checking the correlation between the two quantities and defining a regression model for predicting (at a given confidence level) accuracy values from the EA output.

Introduction

Evolutionary Algorithms (EAs) are a class of stochastic optimization methods that simulate the process of natural evolution. EAs maintain a population of possible solutions that evolve according to rules of selection and other operators, such as recombination and mutation. Several evolutionary methodologies have been proposed for solving real world optimization problems: genetic algorithms (Holland, tion), evolutionary strategies (Schwefel, 1995) and differential evolution (Storn and Price, 1997) among others. By their ability to optimizing non-analytic multi-modal functions, EAs have been successfully applied to a wide range of real life problems, such as parameter estimation (Ravikumar and Panigrahi, 2010), pattern and text recognition (Rizki et al., 2002) and image processing (Cagnoni et al., 2008).

As any iterative technique, EA requires a stop criterion. Unlike optimization methods evolving a single initial value (which rely on real analysis theory), by their stochastic nature, there is not a solid mathematical theory ensuring convergence of evolutionary methodologies in general (Safe et al., 2004; Bäck et al., 1997).

The simplest (an most extended (Safe et al., 2004; Price et al., 2005; Tagetiren and Suganthan, 2006)) stopping criterion consists in reaching a number of iterations or function evaluations. This stopping criterion is not useful by itself (the number of iterations that guarantee convergence significantly varies across problems (Safe et al., 2004)), though it can be necessary when used in addition with alternative criteria to ensure that the algorithm stops (Zielinski and Laur, 2010).

Existing approaches defining general alternative termination conditions address two issues. First, the definition of a quantity reflecting the amount of change between consecutive iterations and, second, the condition that such quantity should fulfill. The quantities reported in the literature (Zielinski and Laur, 2010; Safe et al., 2004) measure either the rate of change in the objective function (improvement-based) or the distribution of the evolving population (distribution-based). Concerning the termination condition, two different conditions are considered. The first condition terminates EA if the measure of the amount of change is below a given threshold. The second one terminates EA in the case that such measure is below a threshold for a number of generations. Improvement-based criteria may lead to early termination (possibly far from the optimum) due to the stochastic nature of EA (Zielinski and Laur, 2010). Meanwhile, distribution-based quantities compare to the accuracy of the solution (distance to the optimum) in terms of number of function executions, as far as suitable parameters (threshold and number of generations) for the termination condition are set (Zielinski and Laur, 2010).

A main limitation for application to real-world problems is that the parameters of the termination condition strongly depend on the topology of the objective function (Zielinski and Laur, 2010). Another concern is that current approaches constrain to statistically comparing the number of iterations reached by the termination condition to the number of iterations required to achieved a given distance to the optimum (Zielinski and Laur, 2010). Although experiments re-

port promising results, the statistical tools used so far can not answer two main (still open) questions. How can we define a termination condition?. Given a confidence level, how can we estimate an upper-bound for the number of iterations required to ensure convergence?.

We propose posing the termination problem in statistical inference terms. From the perspective of statistical inference, the termination problem consists in designing a quantity (depending only on the EA output) that correlates to the accuracy of the solution, so that they can be swapped. This paper introduces a general inference model for predicting the accuracy of the EA solution from the EA current state. We show that a linear regression model in logarithmic scale accurately relates accuracy and distribution-based quantities. We use the inference model to compare several types of distribution-based quantities reported in the literature (Zielinski and Laur, 2010). Our experiments indicate that the maximum distance to the best individual is the best choice in terms of computational efficiency and capability of predicting EA accuracy.

Inference Model

The distance to the (known) function minimum is our gold-standard reference convergence criterion, given that is directly associated to the algorithm accuracy. This criterion can only be computed if the optimum of the test function is known and, thus, is useless in real-world problems. We compute it as the maximum distance to the function minimum of a certain percentage p of the individuals (Zielinski and Laur, 2010) and note it by $RefCrit$. Regarding the alternative quantities, which we will note by $AltCrit$ in general, we have considered the following distribution-based quantities (Zielinski and Laur, 2010):

1. **Maximum Distance (MxD).** It is given by the maximum distance of the population to the best individual.

2. **Population Variability (Std).** It is the maximum standard deviation of the population.

Both quantities can be computed using all individuals or considering only a percentage p of the individuals. The latter is computationally faster and will be indicated by the sufix **Quick**.

Our final goal is to control (predict) the values taken by $RefCrit$ from the values taken by the alternative measure $AltCrit$. In inference statistics, this is achieved by relating both quantities using a regression model.

Regression Model

Given a sampling of two random variables (x and y), the linear regression of y (response variable) over x (explicative variable) is formulated as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

for x_i, y_i the sampling of x and y and ε_i a random error satisfying:

Model Assumptions

1. Linearity: $E(\varepsilon_i) = 0$
2. Homocedasticity: $VAR(\varepsilon_i) = \sigma^2, \forall i$
3. Uncorrelation: $COV(\varepsilon_i \varepsilon_j) = 0, , \forall i, j$
4. Gaussianity: $\varepsilon_i \sim N(0, \sigma^2)$, for $N(0, \sigma^2)$ a normal distribution.

The parameters of the regression model (1) are the regression coefficients $\beta = (\beta_0, \beta_1)$ and the error variance σ^2. The regression coefficients describe the way the two variables relate, while the variance indicates the accuracy of the model and, thus, measures to what extent x can predict y.

Given that, in our case, the inference is over $RefCrit$, our model is:

$$RefCrit_i = \beta_0 + \beta_1 AltCrit_i + \varepsilon_i \tag{2}$$

for $RefCrit_i$, $AltCrit_i$ the values of $RefCrit$ and $AltCrit$ obtained at the i-th iteration.

For a sample of length N, the regression coefficients, $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)$, are computed by Least Squares Estimation (LSE) as:

$$\widehat{\beta} = (X^T X)^{-1} X^T Y \tag{3}$$

for $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$, $Y = (y_1, \ldots, y_N)$ and ^{T} denoting the transpose of a matrix. The differences between estimated responses, $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$, and observed responses y_i:

$$e_i = y_i - \widehat{y}_i$$

are called residuals. Their square sum provides an estimation of the error variance:

$$S_R = \widehat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \tag{4}$$

The four model conditions endow desirable properties to the LSE of the regression coefficients (Ashish, 1990). By the Gauss-Markov theorem under the first three assumptions, the LSE are best linear unbiased estimators and assure that predictions made by least squares fitted equations are good. By adding the fourth assumption (error gaussianity), the LSE has minimum variance among all unbiased estimators (not just linear) and allows the use of parametric tests, such as the Student's t-test for testing hypothesis on parameter values. The central limit theorem (asymptotically) guarantees this last property for large samples. Therefore, given that we have as much samples as EA iterations, in our case, the gaussianity is not a critical issue.

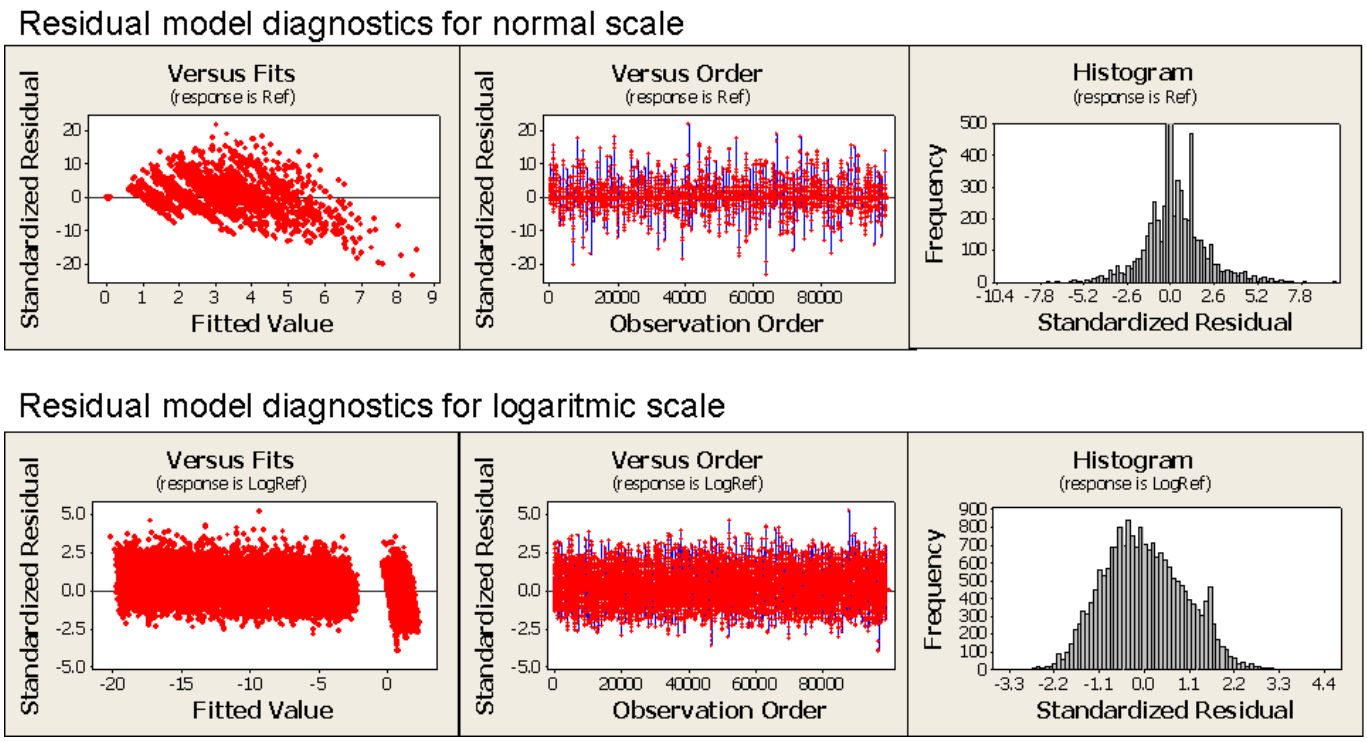

Figure 1: Residual diagnosis plots

The standarized residuals:

$$en_i = (e_i - \mu(e_i))/std(e_i)$$

,for μ the average and std the standard deviation, are used to verify the model assumptions. The plot of en_i over $\widehat{y}_i$ is called the versus fit plot and reflects linearity (in the measure that it is centered at zero) and homocedasticity (uniform deviation from zero). The plot of en vs the sorted explicative variable is called the versus order plot and serves to detect any correlation pattern. Finally, the histogram of the standarized residuals reflects Gaussianity (Newbold et al., 2007).

Figure 1 shows the residuals diagnosis plots. From left to right, we plot the versus fit plot, the versus order plot and the histogram for the standarized residuals en. The plots at normal scale in the first row show that linearity (versus fit plot is centered at zero) and uncorrelation (versus order plot presents no pattern) are fully satisfied. Meanwhile, we observe a clear heterocedasticity in the versus fit plot which presents an increasing deviation from zero. This heterocedasticity is due to a decrease in the population sparseness at advance stages of EA and also affects the Gaussianity assumption, as shown in the most-left histogram of the first row. A monotonous increase in σ^2 is usually solved by taking logarithms in both variables (Arnold, 1997). The residuals plots for the regression model in logarithmic scale (second row in fig. 1) indicate a good homocedasticity and Gaussianity for the standarized residuals.

From now on, the values of $RefCrit$ and $AltCrit$ will be assumed to be in logarithmic scale for the inference model:

$$log(RefCrit_i) = \beta_0 + \beta_1 log(AltCrit_i) + \varepsilon_i \quad (5)$$

We note that, by taking exponentials, the regression model in the original scale is polynomial with multiplicative errors:

$$RefCrit_i = e^{\beta_0} AltCrit_i^{\beta_1} e^{\varepsilon_i} \quad (6)$$

Model verification Previous to any kind of inference, it is mandatory to verify that the estimated parameters make sense. That is, whether it really exists a linear relation between x and y. By the Gauss-Markov theorem, such linear relation can be statistically checked using the following T-test (Newbold et al., 2007)

$$H_0 : \beta_1 = 0 \; H_1 : \beta_1 \neq 0$$

where a $p - value$ close to zero (below α) ensures the validity of the linear model with a confidence $(1 - \alpha)100\%$.

Prediction Model

In order to predict the values of $RefCrit$ from the values achieved by $AltCrit$, we use the regression prediction intervals (Newbold et al., 2007):

$$PI(x_0) = [L_{PI}(x_0), U_{PI}(x_0)]$$

since, for each $x = x_0$, they provide ranges for y at a given confidence level $1 - \alpha$. That is, given x_0, the values of the response y are within $L_{PI}(x_0) \leq y \leq U_{PI}(x_0)$ in $(1 - \alpha)100\%$ of the cases.

Given $x_0 = AltCrit_0$, the confidence interval at a confidence level (1-α) predicting $RefCrit$ is given by:

$$PI(x_0) = [L_{PI}(x_0), U_{PI}(x_0)] = \\ = [\widehat{y_0} + t_{\alpha/2} S_R \sqrt{1 + h_0}, \widehat{y_0} - t_{\alpha/2} S_R \sqrt{1 + h_0}] \quad (7)$$

for $t_{\alpha/2}$ the value of a T-Student distribution with $N - 2$ degrees of freedom having a cumulative probability equal to $\alpha/2$ and:

$$h_0 = (1 \quad x_0)(X^T X)^{-1} \begin{pmatrix} 1 \\ x_0 \end{pmatrix} = a_0 + a_1 x_0 + a2 x_0^2$$

The exponential of PI already provides (with confidence $1-\alpha$) an upper bound for the accuracy of EA solution given EA current state. In order to obtain the upper bound for $AltCrit$ ensuring a given accuracy $U_{PI}(x_0)$, it suffices to find the value x_0 that solves:

$$\widehat{y_0} + t_{\alpha/2} S_R \sqrt{1 + h_0} = U_{PI}(x_0) \quad (8)$$

Using the expressions for $\widehat{y_0}$ and h_0 in (8) and solving for x_0, we obtain:

$$x_0 = \frac{2b_0 b_1 - t_{\alpha/2}^2 S_R^2 a_1 - 2b_1 U_{PI}(x_0) + \sqrt{D}}{2(t_{\alpha/2}^2 S_R^2 a_2 - b_1^2)} \quad (9)$$

where the discriminant is given by:

$$D = (t_{\alpha/2}^2 S_R^2 a_1 - 2b_0 b_1 + 2b_1 U_{PI}(x_0))^2 - \\ - 4(t_{\alpha/2}^2 S_R^2 a_2 - b_1^2)(t_{\alpha/2}^2 S_R^2 (a_0 + 1) - \\ - b_0^2 + 2U_{PI}(x_0) b_0 - U_{PI}(x_0)^2)$$

By taking exponentials from (9) we get the upper bound for $AltCrit$.

Experimental Settings

In this study we have compared the predictive capability of the following distribution-based measures given at the beginning of the previous Section: **MxD**, **MxDQuick** and **StdQuick**. We have used $p = 30\%$ of the population for computing **Quick** scores. We have considered seven well-known test functions (Digalakis and Margaritis, 2002) having a minimum at zero:

1. Esphere:

$$f_1(x) = \sum_{i=1}^{n} x_i^2$$

2. Rosenbrock:

$$f_2(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$$

3. Rastrigin:

$$f_3(x) = \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i + 10)]$$

4. Ackley:

$$f_5(x) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{2}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{2}\sum_{i=1}^{n} cos(2\pi x_i)}$$

5. GoldsteinPrice:

$$f_6(x) = (1 + (x_1 + x_2 + 1)^2 . \\ (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1 x_2 + 3x_2^2)) . \\ (30 + (2x_1 - 3x_2)^2 . \\ (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1 x_2 + 27x_2^2))$$

6. Easom:

$$f_7(x) = -\cos x_1 . \cos x_2 . \\ \exp(-((x_1 - \pi)^2 + (x_1 - \pi)^2))$$

We have used a Differential Evolution (DE) technique for the minimization task. Differential evolution is a real parameter encoding evolutionary algorithm for global optimization over continuous spaces (Storn and Price, 1997; Das and Konar, 2005). In this paper, we use the 3-parameter DE1 scheme (Storn and Price, 1997) for solving DE. For a real search space of dimension D, the population is randomly initialized with ND vectors (for ND the first algorithm parameter). Each vector v in the population is evolved by mutation and recombination operators. Given a mutation rate $F \in [0, 2]$ (second parameter of the algorithm), the mutation operator produces a new vector vm by adding a vector difference of two randomly chosen population vectors $v1$ and $v2$ to another randomly chosen vector $v3$:

$$vm = v1 + F(v2 - v3)$$

For the recombination step, a new vector vf is created from the mutation vector by means of a combination rate CR (third parameter of the algorithm) as follows:

$$vf_i = \begin{cases} vm_i & \text{if} \quad r_i < CR \text{ or } i = k \\ v_i & \text{otherwise} \end{cases}$$

for vf_i the i-th component of vf and $r_i \in [0, 1]$ a random number and k a random number uniformly distributed in $[1, D]$. Finally a selection operator is applied. The vector vf and the initial vector v are compared and the vector that better fits the objective function is selected and remains in the next population. This process is iteratively repeated until a stopping criterion is reached. Following the literature (Das and Konar, 2005), we have chosen the following values for DE parameters: D=2, ND=20, F=0.9, CR=0.5. For

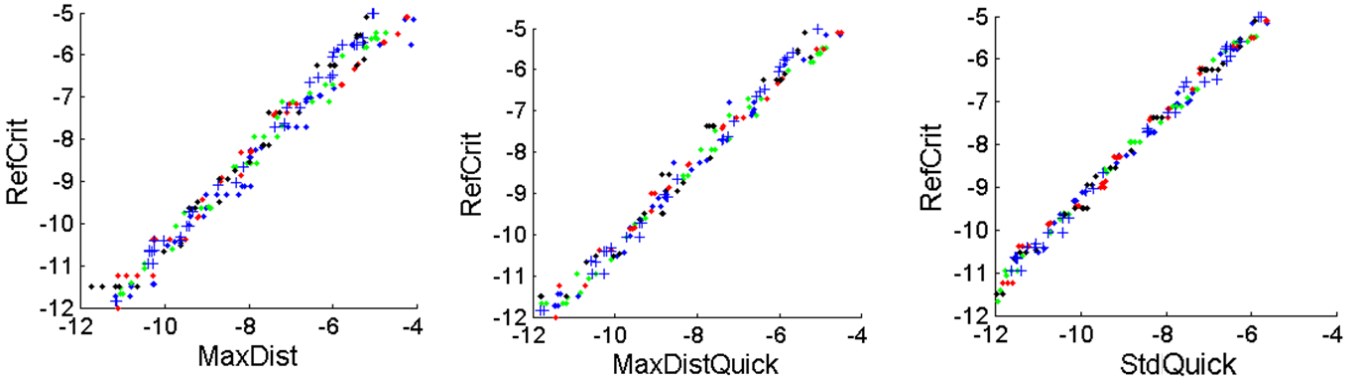

Figure 2: Scattered plots for Rastringin test function and 10 different runs of DE.

each test function, we have executed 100 trails of the algorithm during 10.000 iterations each one.

For each test function and alternative quantity, two different experiments have been carried out:

1. **Model Assessment.** The suitability and accuracy of the linear model in logarithmic scale has been assessed by the T-test on the regression coefficients, as well as, the analysis of the residuals variance (S_R).

2. **Model Prediction.** In order to assess the prediction capabilities of each model two different experiments have been addressed. On one hand, we have explored the relation between $RefCrit$ and $AltCrit$ by analyzing the confidence intervals of the regression coefficients. On the other hand, we have compared the prediction intervals across the three distribution-based quantities.

Experiments and Results

Model Assessment

Figure 2 shows scattered plots associated to the regression model for the Rastringin test function. The y axis represents $RefCrit$ values and the x axis each of the alternative quantities (from left to right **MxD**, **MxDQuick** and **StdQuick**. Each plot shows 10 different runs marked with distinct colors and markers. For all alternative quantities, we observe a uniform behavior across DE executions, which present the same linear pattern with a small variation.

Table 1 reports the estimation of the model parameters (the regression coefficients $\widehat{\beta}_0$, $\widehat{\beta}_1$ and the residual variance S_R) and the the p-value of the model verification T-test. We report values for each test function (rows) and alternative quantity (columns). For all cases, there is a clear linear relation between accuracy and the alternative quantities (with p close to the working precision). Besides the goodness-of-fit is excellent, given that S_R is extremely small compared to the variable ranges (see fig. 2).

Concerning the relation between the two variables, it is worth noticing two aspects. Firstly, we observe that the estimated slope $\widehat{\beta}_1$ is close to 1 for all cases. This implies that the relation in logarithmic scale is a translation of the identity and the regression model in the original scale is also linear. Secondly, the constant coefficients $\widehat{\beta}_0$ are sorted as follows:

$$\widehat{\beta}_0(\mathbf{MxD}) \leq \widehat{\beta}_0(\mathbf{MxDQuick}) \leq 0 \leq \widehat{\beta}_0(\mathbf{StdQuick})$$

The above commented points indicate that there might be the following tendency:

$$\mathbf{StdQuick} \leq RefCrit \leq \mathbf{MxDQuick} \leq \mathbf{MxD}$$

This already suggests that the value of maximum distances itself might guarantee an upper bound for the EA accuracy. In order to really confirm such hypothesis, we should analyze the prediction intervals.

Model Prediction

Figure 3 shows the prediction intervals for the 6 test functions. Each plot shows the prediction interval for all alternative quantities, as well as, the identity line (solid line) for a better visual comparison between $AltCrit$ prediction and $RefCrit$ values. The alternative quantity can substitute $RefCrit$ in the measure that the identity line is within the range given by the prediction interval. This is the case for quantities based on maximum distances. In the case of **StdQuick** the predicted values are above the reference identity line. This implies that **StdQuick** and $RefCrit$ can not be directly swapped and, thus, we need the upper bound given in (9) for predicting $RefCrit$ values.

Table 2 reports the upper bounds for each alternative quantity ensuring a given accuracy for $RefCrit$. For each test function (rows), we report values for two accuracies 10^{-6} and 10^{-9}. As suggested by the plots in fig. 3, for Rastrigin, Ackley and Easom test functions, the upper bound for **MxD** is almost equal to the accuracies 10^{-6} and 10^{-9}. This

	MxD				MxDQuick				StdQuick			
	p	b_1	b_0	S_R	p	b_1	b_0	S_R	p	b_1	b_0	S_R
Esfere	$\leq 10^{-32}$	1.000	-0.487	0.12	$\leq 10^{-32}$	1.000	-0.183	0.05	$\leq 10^{-32}$	1.000	0.685	0.03
Rosenbrock	$\leq 10^{-32}$	1.002	-0.478	0.16	$\leq 10^{-32}$	1.002	-0.122	0.09	$\leq 10^{-32}$	1.003	0.761	0.05
Rastrigin	$\leq 10^{-32}$	0.996	-0.557	0.07	$\leq 10^{-32}$	1.006	-0.207	0.06	$\leq 10^{-32}$	1.007	0.686	0.03
Ackley	$\leq 10^{-32}$	1.000	-0.487	0.07	$\leq 10^{-32}$	1.004	-0.148	0.06	$\leq 10^{-32}$	1.004	0.725	0.03
GoldstenPrice	$\leq 10^{-32}$	0.998	-0.504	0.16	$\leq 10^{-32}$	1.001	-0.127	0.07	$\leq 10^{-32}$	1.002	0.804	0.05
Easom	$\leq 10^{-32}$	1.001	-0.484	0.05	$\leq 10^{-32}$	1.011	-0.143	0.04	$\leq 10^{-32}$	1.011	0.766	0.03

Table 1: Model fitting scores

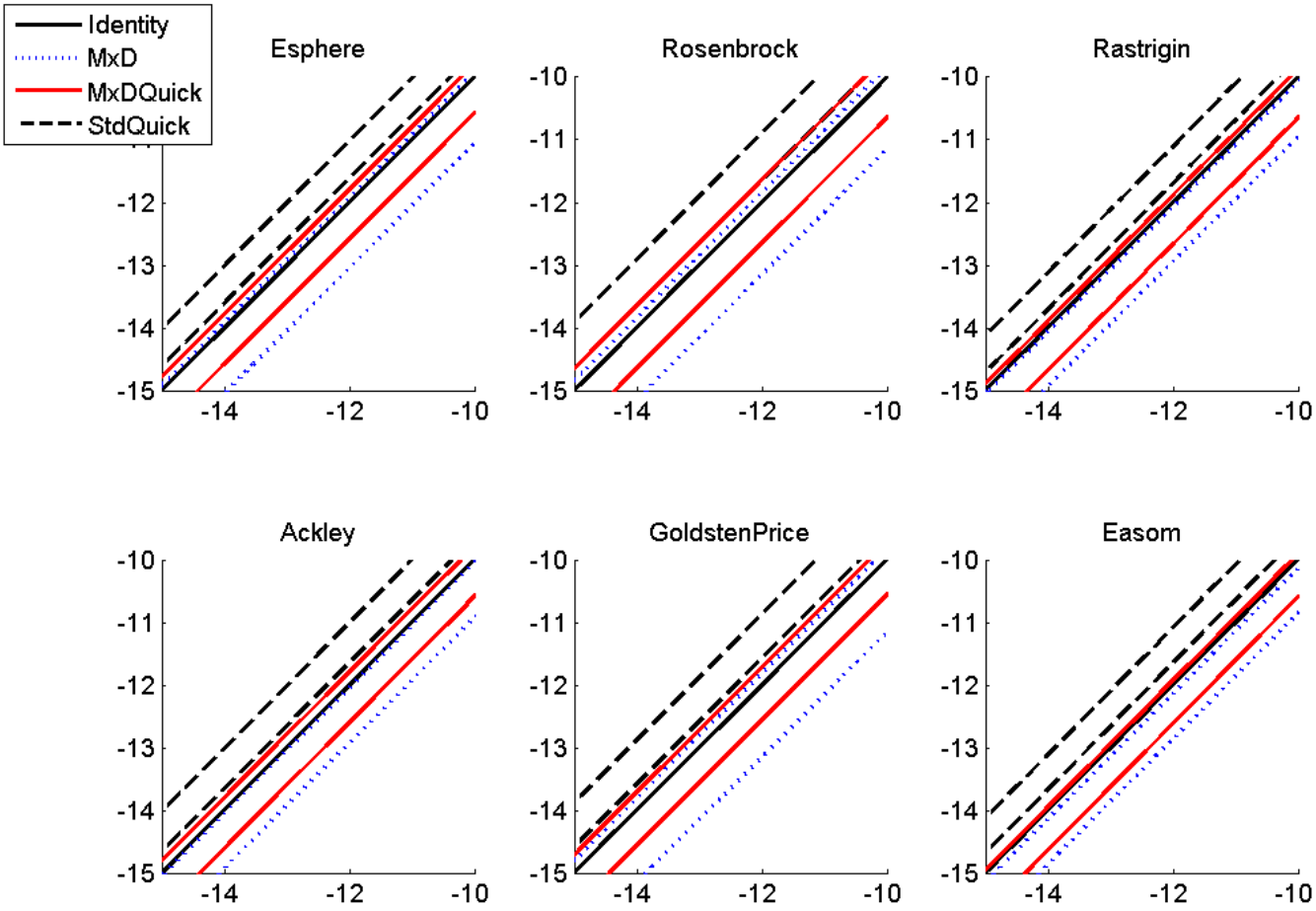

Figure 3: Prediction intervals

is also the case for Easom test function and **MxDQuick**. For the remaining cases, **MxD** and **MxDQuick** upper bounds are a little lower (though still comparable). We would like to note that this does not contradict the swapability of the two quantities. The upper bound condition requires that $RefCrit$ equals U_{PI}. This is a stronger condition than the swapability one, which just requires $RefCrit \leq U_{PI}$. Concerning **StdQuick**, its upper bounds are clearly lower (a 40% at most) than the two accuracies. This confirms that **StdQuick** and $RefCrit$ are not directly swapable.

Conclusions and Future work

In real-world problems (which have unknown optimums) it is mandatory to design a termination condition for EA ensuring stabilization close to the unknown optimum. As far as we know, this is the first work addressing EA termination condition in terms of statistical inference. In this context, we have explored to what extend a reference quantity

		MxD	MxDQuick	StdQuick
Esfere	10^{-6}	$8.3 \cdot 10^{-7}$	$7.5 \cdot 10^{-7}$	$3.5 \cdot 10^{-7}$
	10^{-9}	$8.3 \cdot 10^{-10}$	$7.5 \cdot 10^{-10}$	$3.5 \cdot 10^{-10}$
Rosenbrock	10^{-6}	$7.5 \cdot 10^{-7}$	$6.4 \cdot 10^{-7}$	$3.1 \cdot 10^{-7}$
	10^{-9}	$7.6 \cdot 10^{-10}$	$6.5 \cdot 10^{-10}$	$3.2 \cdot 10^{-10}$
Rastrigin	10^{-6}	$9.8 \cdot 10^{-7}$	$8.4 \cdot 10^{-7}$	$4.0 \cdot 10^{-7}$
	10^{-9}	$9.5 \cdot 10^{-10}$	$8.7 \cdot 10^{-10}$	$4.2 \cdot 10^{-10}$
Ackley	10^{-6}	$9.7 \cdot 10^{-7}$	$7.6 \cdot 10^{-7}$	$3.5 \cdot 10^{-7}$
	10^{-9}	$9.7 \cdot 10^{-10}$	$7.8 \cdot 10^{-10}$	$3.6 \cdot 10^{-10}$
GoldsteinPrice	10^{-6}	$7.3 \cdot 10^{-7}$	$7.0 \cdot 10^{-7}$	$3.0 \cdot 10^{-7}$
	10^{-9}	$7.2 \cdot 10^{-10}$	$7.0 \cdot 10^{-10}$	$3.1 \cdot 10^{-10}$
Easom	10^{-6}	$1.0 \cdot 10^{-6}$	$8.9 \cdot 10^{-7}$	$3.9 \cdot 10^{-7}$
	10^{-9}	$1.1 \cdot 10^{-9}$	$9.6 \cdot 10^{-10}$	$4.2 \cdot 10^{-10}$

Table 2: Upper-bound ensuring a given accuracy of EA

(not available in real-world problems) measuring EA accuracy ($RefCrit$) can be substituted by an alternative quantity ($AltCrit$) computed from EA population.

According to our experiments on several known test functions, there is a strong (almost ideal) linear relation between distribution-based quantities (**MxD**, **MxDQuick** and **StdQuick**) and the distance to the optimum. This allows analyzing the prediction capabilities of each distribution-based quantity by means of the regression prediction intervals. From our analysis, we conclude that quantities based on maximum distances (**MxD**, **MxDQuick**) have the highest concordance to EA accuracy and, thus, can substitute it as termination condition. Given that **MxDQuick** is computationally faster than **MxD**, it is the best candidate for terminating EA in real-world problems.

We consider that there are some issues that should be further developed. The test functions used are a small set of benchmarking data sets (we cover two out of the five categories described in (Hansen et al., 2010)) and only 2-D problems have been solved. Enlarging the test function data set including groups of functions with specific key features (Hansen et al., 2010) is work currently under development. However, the functions used include three properties (multimodality, global structure and scalability) reported in a recent study (Mersmann et al., 2010) to have a high influence in the performance of EA's. Regarding size, although it definitely influences convergence rate (more iterations of EA are required (Hansen et al., 2010)), this is independent of the relationship between *RefCrit* and *AltCrit*. Thus, size is not a limitation for the prediction model, which links convergence rate with population stability.

In this study we have restricted to DE algorithm. We are currently enlarging EA methods in order to cover existing EA paradigms: genetic algorithms (Goldberg and Richardson, 1987), evolutionary strategies (Beyer and Schwefel, 2002), particle Swarm optimization (Barrera and Coello, 2009), among others. Nevertheless, we do not expect any significant changes in our conclusions since DE already presents the main features of EA (Ronkkonen, 2009).

Finally, in our experimental setting test functions have been studied separately. We consider that the influence of the test function should be taken into account, so that the inference can be done independently of the function features. This will be studied by using generalized regression models including random effects (Lee et al., 2006) modelling the impact of the test function group.

Acknowledgements. This work was supported by the Spanish projects PI071188, TIN2009-13618 and CONSOLIDER-INGENIO 2010 (CSD2007-00018). The 2nd author has been supported by The Ramon y Cajal Program.

We would like to thank Llorenç Badiella from the Statistics Service at the Universitat Autonoma de Barcelona for his advise.

References

Arnold, S. (1997). *The theory of linear models and multivariate observations*. Wiley.

Ashish, S.and Srivastava, M. (1990). *Regression analysis. Theory, methods and applications.* Springer-Verlag.

Bäck, T., Hammel, U., and Schwefel, H. (1997). Evolutionary computation: Comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, 1:3–17.

Barrera, J. and Coello, C. A. C. (2009). *A Review of Particle Swarm Optimization Methods used for Multimodal Optimization*, pages 9–37. Springer-Verlag.

Beyer, H.-G. and Schwefel, H.-P. (2002). Evolution strategies: A comprehensive introduction. *Journal Natural Computing*, 1(1):3–52.

Cagnoni, S., Lutton, E., and Olague, G. (2008). *Genetic and Evolutionary Computation for Image Processing and Analysis*. Hindawi Publishing Corporation.

Das, S. and Konar, A. (2005). An improved differential evolution scheme for noisy optimization problems. In *LNCS vol. 3776*, pages 417–421.

Digalakis, J. and Margaritis, K. (2002). An experimental study of benchmarking functions for genetic algorithms. *International Journal of Computer Mathematics*, 79(4):403–416.

Goldberg, D. and Richardson, J. (1987). Genetic algorithms with sharing for multimodal function optimization. In *Second International Conference on Genetic Algorithms and their Application*, pages 41–49.

Hansen, N., Auger, A., Ros, R., Finck, S., and Posík, P. (2010). Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009. In *GECCO 2010*.

Holland, J. (1975 (1192 edition)). *Adaptation in Natural and Artificial Systems*. MIT Press.

Lee, Y., Nelder, J., and Pawitan, Y. (2006). *Generalized linear models with random effects*. Chapman and Hall/CRC Boca Raton.

Mersmann, O., Preuss, M., and Trautmann, H. (2010). Benchmarking evolutionary algorithms: Towards exploratory landscape analysis. In *PPSN XI, Part I, LNCS 6238*, pages 73–82.

Newbold, P., Carlson, W., and Thorne, B. (2007). *Statistics for Business and Economics*. Pearson Education, 6th edition.

Price, K., Storm, R., and Lampinen, J. (2005). *Differential Evolution: A Practical Approach to Global Optimization*. Springer.

Ravikumar, V. and Panigrahi, B. (2010). Comparative study of evolutionary computing methods for parameter estimation of quality signals. *International Journal of Applied Evolutionary Computation*, 1 (2):28–59.

Rizki, M., Zmuda, M., and Tamburino, L. (2002). Evolving pattern recognition systems. *IEEE Transactions on Evolutionary Computation*, 6 (6):594–609.

Ronkkonen, J. (2009). *Continuous Multimodal Global Optimization with Diferential Evolution Based Methods*. PhD thesis, Lappeenranta University of Technology.

Safe, M., Carballido, J., and Ponzoni, I. (2004). On stopping criteria for genetic algorithms. In *17th BSAI, LNAI 3117*, pages 405–413.

Schwefel, H. (1995). *Evolution and Optimum Seeking*. Wiley, New York.

Storn, R. and Price, K. (1997). Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359.

Tagetiren, M. and Suganthan, P. (2006). A multi-populated differential evolution algorithm for solving contrained optimization problem. In *IEEE Congres on Evolutionary Computation*, pages 340–347.

Zielinski, K. and Laur, R. (2010). Stopping criteria for differential evolution in constrained single-objective optimization. In *Advances in Differential Evololution, vol. 143*, pages 111–138.