We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Efficient Estimation of Distribution Algorithms by using the Empirical Selection Distribution

S. Ivvan Valdez[1], Arturo Hernández[2] and Salvador Botello[2]
[1]Universidad del Papaloapan (UNPA),
[2]Centre for Research in Mathematics A.C. (CIMAT),
México

1. Introduction

Estimation of Distribution Algorithms (EDAs) (Mühlenbein et al., 1996; Mühlenbein & PaaB, 1996) are a promising area of research in evolutionary computation. EDAs propose to create models that can capture the dependencies among the decision variables. The widely known Genetic Algorithm could benefit from the available dependencies if the building blocks of the solution were correlated. However, it was proved that the building blocks of a genetic algorithm have a limited capacity for discovering and using complex relationships (correlations) among variables. EDAs instead, focus on learning probability distributions which serve as the vehicle to capture the data dependencies and the data structure as well.

In order to show how the proposed method unifies the theory for infinite sized population with the finite sized population case of practical EDAs, we explain them first. An EDA with infinite sized population would perform the steps shown in the algorithm in Table 1.

EDA with an infinite size population	
1	*t=0*
2	Initialize a probability model $p(x,t)$ (usually a uniform distribution).
3	Generate an infinite sized sample X_t from $p(x,t)$.
4	Evaluate X_t in the objective function(s) and constraint(s).
5	**Compute the selection distribution $p^S(x,t)$, and use it as the new search distribution. Then: $p(x,t+1) = p^S(x,t)$. (Selection and model rcomputation step)**
6	*t=t+1*
7	If the stop criterion is not reached go to Step 3

Table 1. The estimation of distribution algorithm with an infinite size population: the selection distribution is equal to the search distribution.

The selection method introduces the search bias necessary to improve the current best solution, and in fact, "pushes" the population towards the optimum. For this case where the population is infinite, imagine a probability value could be computed for every point in the search space. Therefore, the "exact" probability distribution of the selected individuals (which is also an infinite set), can be used to sample the new population. Such distribution, the **selection distribution**, is the probability of any point in the search space of being selected. Four selection distribution formulas for an infinite population are shown in Table

2, left column: Truncation, Boltzmann, Proportional, and Binary Tournament selections. In the same table but right column, we show the proposed empirical selection distributions studied in this paper. The first objective of this paper is to show that the proposed distributions converge to the theoretical ones as the population grows. The second objective is to apply the empirical distributions in practical EDAs, and to compare their performance through the analysis of several experiments. It has been proved that the ideal EDAs with any of the four selection distributions shown, converges to the optimum after a large number of generations (Zhang & Mühlenbein, 2004).

Selection Model	Empirical Selection Distribution
Truncation $$p^S(x,t) = \begin{cases} \dfrac{p(x,t)}{\alpha(t)} & \text{if } f(x) \geq \theta_t \\ 0 & \text{otherwise} \end{cases}$$	$$\hat{p}^S(x_i,t) = \begin{cases} \dfrac{1}{\left|X_t^S\right|} & \text{if } f(x_i) \geq \theta_t \\ 0 & \text{otherwise} \end{cases}$$
Boltzmann $$p^S(x,t) = \dfrac{e^{\beta(t)f(x)}p(x,t)}{Z(t)}$$	$$\hat{p}^S(x_i,t) = \dfrac{e^{\beta(t)f(x)}}{\sum\limits_{j=1}^{|X|} e^{\beta(t)f(x_j)}}$$
Proportional $$p^S(x,t) = \begin{cases} \dfrac{p(x,t)}{\alpha(t)} & \text{if } f(x) \geq \theta_t \\ 0 & \text{otherwise} \end{cases}$$	$$\hat{p}^S(x_i,t) = \dfrac{f(x_i)}{\sum\limits_{j=1}^{|X|} f(x_j)}$$
Binary Tournament $$p^S(x,t) = 2p(x,t) \int\limits_{f(y) \leq f(x)} p(y,t)dy$$	$$\hat{p}^S(x_i,t) = \dfrac{\sum\limits_{j=1}^{|X|} I(i,j)}{\sum\limits_{i=1}^{|X|}\sum\limits_{j=1}^{|X|} I(i,j)}.$$ Where $I(i,j) = 1$, if $f(x_j) < f(x_i)$, and 0 otherwise.

Table 2. Left: Selection models for the EDA with an infinite sized population. Right: the respective **empirical selection distribution**.

Let us now contrast the infinite sized population case with the standard practical EDA, thereby, with a finite sized population. The practical EDA is presented in the algorithm in Table 3. Notice the selection operator returns a selected set X_t^S whose size is a fraction of the total population. In the next step the standard EDA seeks to approximate the distribution of the selected set via a parametric distribution. The distribution parameters, and in some cases the data structure, are learned from the selected set. Such distribution, the **search distribution**, is the model (with a predefined structure), used for learning the underlying joint probability density of the selected set, $p(x,t+1) \approx p(X_t^S,t)$. Also, note that the next population is simulated from the search distribution. The search distribution and its learning algorithm are so important for an EDA that, in fact, gives name to the EDA version.

Examples of search distributions are: Bayesian networks (Pelikan et al.., 1999), Polytrees (Soto & Ochoa, 2000), and Dependency Trees in discrete spaces. Also, Gaussian univariate and multivariate models (Bosman & Thierens, 2000; Larrañaga et al., 1999) in the continuous case, among others (Larrañaga & Lozano, 2001; Lozano et al., 2006; Pelikan et al., 2006).

Standard EDA with a finite size population	
1	*t=0*
2	Initialize a probability model *p(x,t)* (**search distribution**).
3	Generate an infinite sized sample X_t from *p(x,t)*.
4	Evaluate X_t in the objective function(s) and constraint(s).
5	S_t←**SELECTION(X_t) (selection step)**
6	**Recompute the search distribution model** *p(x,t+1)*, **such that,** *p(x,t+1)≈p(S_t)* **(parametercomputation step).**
7	*t=t+1*
7	If the stop criterion is not reached go to Step 3

Table 3. The estimation of distribution algorithm with a finite size population: the search distribution is computed by learning parameters from the selected set.

The approach introduced in this work is shown in the algorithm in Table 4. No selection operator is used, our approach firstly calculates the proposed **empirical selection distribution** to approximate to the theoretical selection distribution. The empirical selection distribution, $\hat{p}^S(x_i,t)$, is the exact selection distribution when the population is thought as a model of the whole search space. Then, the search distribution model is created directly using the information from the empirical selection distribution, and used to simulate the new population. Remember, the empirical selection distribution equations are presented in Table 2, right column. The calculation of the empirical selection distribution is easy (as we shall explain later), however, its main advantage is to carry all the information needed to build the best approximating search distribution.

EDA with the Empirical Selection Distribution	
1	*t=0*
2	Initialize a probability model *p(x,t)* (**search distribution**).
3	Generate an infinite sized sample X_t from *p(x,t)*.
4	Evaluate X_t in the objective function(s) and constraint(s).
5	**Compute the empirical selection distribution** $\hat{p}^S(x,t)$ **(selection step)**
6	**Recompute the search distribution model** *p(x,t+1)*, **with** $\hat{p}^S(x,t)$ **(parametercomputation step).**
7	*t=t+1*
7	If the stop criterion is not reached go to Step 3

Table 4. The estimation of distribution algorithm with a finite size population: the search distribution is computed by learning parameters from the selected set.

The chapter is presented as follows: Section 2 briefly reviews the most common selection methods used in EDAs. Section 3 discusses about the convergence of the empirical selection

distribution to the exact selection distribution. Section 4 introduces the general method to approximate the search distribution to the selection model. Section 5 is a comparison with related work. Section 6 presents well known EDAs, which have been modified to apply the proposed method. A set of experiments is presented in Section 7 and discussion about the performance of different selection methods. Finally, Section 8 shows the perspectives of future work and concludes.

2. Selection methods

The main goal of a selection operator is to bias the population towards promising regions of the search space. Truncation, Boltzmann, Proportional and Tournament selection operators discussed in this paper are introduced through an example. Assume the following objective function and an infinite sized population whose elements have as fitness the function value at $(x,y) = \exp(|x|-2)^2 + 4\cos(20x) + \exp(|y|-2)^2 + 4$. The plot of this function is shown in Figure 1(a). Most of the EDAs draw the initial population from a uniform distribution, which is shown in Figure 1(b). Hence, during the first iteration any point has the same sampling probability. Assuming an infinite sized population, the rest of the figures show the selection distribution function.

- The **Truncation selection** is shown in Figure 1(c). This selection is widely used by many EDAs. Here the select set would be the best population above a function threshold of $\theta = 30$. The probability density of the truncation selection is shown in Figure 1(c). Observe the flat area; it means the selection probability for the population above the threshold value is the same. The truncation selection hides the roughness of the objective function above the threshold value by assigning to such area the same selection probability.

- The **Boltzmann selection** is shown in Figure 1(d). This operator exponentially favors the most promising regions (the zones with high function value), as shown in Figure 1(d). Most of the probability mass is condensed on a single peak which corresponds to the function optimum. Since the remaining region is quite flat, the Boltzmann selection will deliver a selected set clustered on the peak region.

- The **Proportional selection** is shown in Figure 1(e). This operator, proposed by John Holland for the standard genetic algorithm, selects points with some probability directly proportional to its objective value. The resulting probabilistic model of this method, shown in Figure 1(e), is very similar to the objective function (although in a different scale).

- The **Tournament** selection is shown in Figure 1(f). The tournament selection picks the best point found in a randomly chosen subset of the population. The usual size of the subset is 2. Figure 1(f) shows that the resulting probabilistic model acquires the roughness of the objective function. A larger subset would increase the selection pressure on the winning individual, therefore, flatting the density function shown.

Zhang and Mühlenbein have shown the selection models just illustrated can drive the population to the function optimum (Zhang & Mühlenbein, 2004). The main factor that makes convergence possible is the bias the selection operator introduces into the population. In the following section we provide simple proofs of the convergence of the empirical selection distribution to the selection distribution.
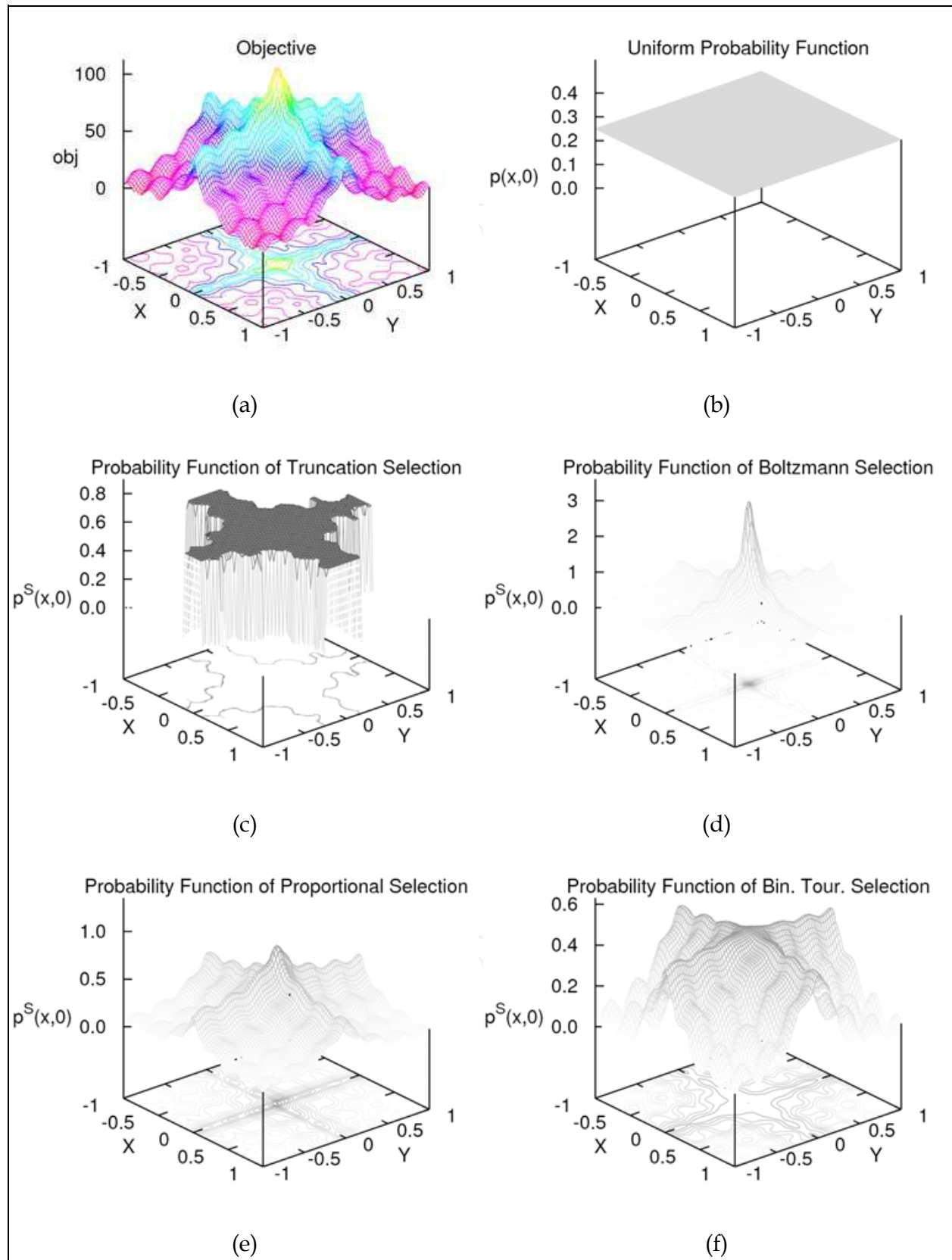
Fig. 1. a) Objective function example, b) uniform distribution, c) to f) probability densities of the most widely used selection methods.

3. Convergence of the empirical selection distribution to the selection distribution

The empirical selection distribution equations for the four selection methods are shown in Table 2. In this section we provide simple proofs on the convergence of the empirical selection distribution to the selection distribution with large population size. In fact, the larger the population size, the better the approximation of the empirical selection distribution to the selection distribution.

3.1 Discrete variables

Assume an EDA with discrete variables and a search distribution denoted by $p(x,t)$ (the continuous case will be further tackled).

Truncation selection. Say the search space is described by m binary variables, $\{y_1, y_2, ..., y_m\}$. There are $n = 2^m$ combinations, and denote by x_i, $i = \{1, 2, ..., n\}$ to each combination. Then, for a large sample X with size $|X| >> n$, every combination x_i receives a frequency $freq_i$ of instances. The empirical selection distribution at generation t is given by:

$$\hat{p}^S(x_i, t, \theta_t) = freq_i \cdot \frac{1}{|S_t|} \tag{1}$$

Being $|S_t|$ the number of instances of X such that $f(x) > \theta_t$. Now recall that for a infinite sample $freq_i / |X| = p(x_i, t)$, then for $\alpha(t) = |S_t| / |X|$ the empirical selection distribution is given by:

$$\hat{p}^S(x_i, t, \theta_t) = \frac{p(x_i, t)}{\alpha(t)}, \tag{2}$$

which is exactly the expression given in Table 2 for the infinite sized population. The $\alpha(t)$ value is the proportion of truncated solutions.

Boltzmann selection. The empirical selection model is given by:

$$\hat{p}^S(x_i, t) = \frac{e^{\beta(t)f(x_i)}}{\sum\limits_{j=1}^{|X|} e^{\beta(t)f(x_j)}}, \tag{3}$$

then, in an analogous way to the truncation method:

$$\hat{p}^S(x_i, t) = \frac{freq_i \cdot e^{\beta(t)f(x_i)}}{\sum\limits_{j=1}^{|X|} freq_k \cdot e^{\beta(t)f(x_k)}}, \tag{4}$$

Substituting $freq_i = |X| \cdot p(x_i, t)$, and $freq_i = |X| \cdot p(x_i, t)$, $Z(t) = \sum\limits_{k=1}^{n} p(x_k, t) \cdot e^{\beta(t)f(x_k)}$, then:

$$\hat{p}^S(x_i,t) = \frac{p(x_i,t) \cdot e^{\beta(t)f(x_i)}}{Z(t)} , \qquad (5)$$

So, we have the exact Boltzmann selection distribution.

Proportional selection. The empirical selection model is given as follows:

$$\hat{p}^S(x_i,t) = \frac{f(x_i)}{\displaystyle\sum_{k=1}^{|X|} f(x_k)} , \qquad (6)$$

then:

$$\hat{p}^S(x_i,t) = \frac{freq_i f(x_i)}{\displaystyle\sum_{k=1}^{|X|} freq_k f(x_k)} , \qquad (7)$$

Substituting, $freq_i = |X| \cdot p(x_i,t)$ then:

$$\hat{p}^S(x_i,t) = \frac{p(x_i,t)f(x_i)}{\displaystyle\sum_{k=1}^{|X|} p(x_k,t)f(x_k)} , \qquad (8)$$

and $E(t) = \displaystyle\sum_{k=1}^{n} p(x_k,t) \cdot f(x_k)$, then the empirical selection for such large sample X is given by:

$$\hat{p}^S(x_i,t) = \frac{p(x_i,t)f(x_i)}{E(t)} , \qquad (9)$$

which is exactly the expression in Table 2.

Binary tournament selection. For the tournament selection method the exact selection distribution for the discrete case is given by:

$$\hat{p}^S(x_i,t) = \frac{p(x_i,t) \displaystyle\sum_{f(y)<f(x_i)} p(y,t)}{C} , \qquad (10)$$

being C the normalization constant, the number 2 in Table 2 is also a constant, so it is absorbed by C. Then the empirical selection distribution is given by:

$$\hat{p}^S(x_i,t) = \frac{\displaystyle\sum_{j=1}^{|X|} I(i,j)}{\displaystyle\sum_{i=1}^{|X|} \sum_{j=1}^{|X|} I(i,j)} , \qquad (11)$$

Where $I(i,j) = 1$, if $f(x_j) < f(x_i)$ and 0 otherwise. The term $\displaystyle\sum_{i=1}^{|X|}\sum_{j=1}^{|X|} I(i,j)$ is a normalization constant. Then, considering that for any y value the number of instances in the large population X are $freq_y = |X| \cdot p(y,t)$, and the number of instances of a variable (vector) value x_i is $freq_i = |X| \cdot p(x_i,t)$, then substituting:

$$\hat{p}^S(x_i,t) = \frac{p(x_i,t)\displaystyle\sum_{f(y)<f(x_i)} p(y,t)}{C}, \tag{12}$$

which is the exact selection distribution for $C = \displaystyle\sum_{i=1}^{|X|} p(x_i,t) \cdot \sum_{f(y)<f(x_i)} p(y,t)$.

3.2 Continuous variables

For the continuous case, consider a univariate search space with domain in the interval *[a,b]*, then a set of points x_i for *i=1,2,...,|X|* define partitions. If we use these points as possible instances of a discrete variable, then, we have the equivalence with the discrete selections distribution previously shown. The partition size is $\Delta_i = x_{i+1} - x_i$, if $\Delta_i < \varepsilon$, the sums in Equation 8 can be written as Riemann sums. Even though this is not a proof for continuous cases, it is a strong argument to explain the convergence of the empirical selection distribution to the selection distribution. Figure 2 shows the similarities between the exact selection density function and the empirical selection distributions.

4. Computing the search distribution

Following the usual steps of an EDA, an arbitrary large sample should be obtained from the empirical selection distribution, and then used to learn the search distribution parameters. In our approach, sampling the empirical selection is avoided without diminishing the advantages of using all the information at hand. It is known that the relative frequency of a point x_i for an infinitely large sample is equal to the probability $\hat{p}^S(x_i,t)$. Thus, the sampling process can be avoided and $\hat{p}^S(x_i,t)$ can be used as the frequency of the point x_i.

For example, suppose that the search distribution is a Gaussian with mean μ and variance *v*. These parameters must be computed to get the best approximation of the Gaussian to the selection model. Assume (for a moment) that a sample $\hat{S}$ of size $|\hat{S}|$ is obtained from the empirical selection distribution. The only possible values sampled from the empirical selection distribution are those in the population. Hence, most of the points x_i would get more than one instance in the sample $\hat{S}$. Denote the number of instances of a point x_i as $freq_i$, the estimator of the mean is:

$$\mu = \frac{\displaystyle\sum_{i=1}^{|X|} freq_i \cdot x_i}{|\hat{S}|}, \tag{13}$$
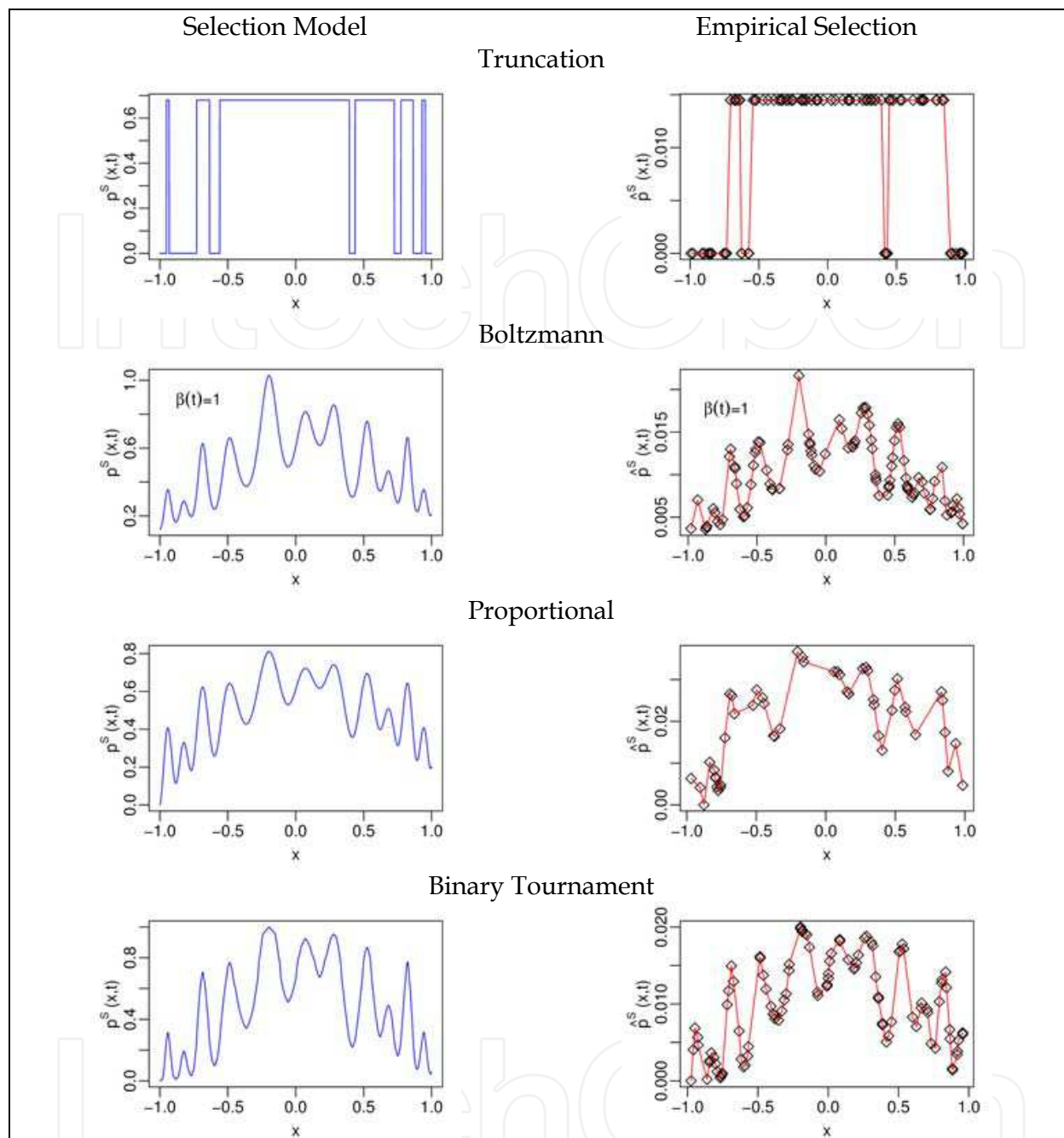
Fig. 2. Left: Selection model plot. Right: empirical selection distribution plot.

Let us denote $\hat{p}_i^S = \hat{p}^S(x_i, t)$. We know that when $|\hat{S}| \to \infty$, then, $\frac{freq_i}{|\hat{S}|} = \hat{p}_i^S$. Substituting this

term in Equation 13, the mean is simply computed as follows:

$$\mu = \sum_{i=1}^{|X|} \hat{p}_i^S \cdot x_i \,, \tag{14}$$

where $\hat{p}_i^S$ is the probability of a point computed with the empirical selection distribution. Therefore, sampling the empirical distribution is not necessary. Also important, note that

the computation of the probabilities $\hat{p}_i^S$ is independent of the search distribution parameters. This allows us to easily adapt an EDA to any of the four selection models. These models cover most EDA implementations and other selection methods can be also expressed as distributions. In addition, the computational cost of $\hat{p}_i^S$ in most of the cases is lower or equal to that of applying a selection operator. For example, the Proportional and Boltzmann selections must calculate the probabilities in Table 2 in order to obtain a sample which becomes the selected set. This sample is not necessary for the empirical selection distribution based EDA (ES-EDA). For the truncation method in the standard EDA as well as the ES-EDA, it is only necessary to know the individuals whose objective function is greater than θ_t, thus, the computational cost is the same. The binary tournament selection requires comparisons of the order of the selected set size, while the empirical selection distribution requires comparisons of order $|X|(|X|-1)/2$, hence, it is the unique case in which an ES-EDA has a greater cost than the standard EDA. The parameter computation in the ES-EDA in general increases its computational cost only by multiplying each individual value x_i by its corresponding frequency $\hat{p}_i^S$, which is a minimal cost.

5. Related work

In order to show that this work is an extension of previous but less general approaches, we will show the equivalence of the resulting formulae when applying frequencies from the empirical selection distribution. Yunpeng et al. (Yunpeng et al., 2006) approximate the Boltzmann distribution with a Gaussian model by minimizing the Kullback-Leibeler divergence. The computation of the Gaussian distribution parameters is presented in Equations 15 and 16. Note that this approach which requires costly analytical work, gives exactly the same formulae as the Boltzmann empirical selection distribution presented in Table 2 with the UMDA$_c$ and the EMNA$_{global}$.

$$\mu = \frac{\sum_{i=1}^{|S|} e^{\Delta \beta f(x_i)} \cdot x_i}{\sum_{i=1}^{|S|} e^{\Delta \beta f(x_i)}}, \tag{15}$$

$$\Gamma = \frac{\sum_{i=1}^{|S|} \left[e^{\Delta \beta f(x_i)} \cdot (x_i - \mu)(x_i - \mu)^T \right]}{\sum_{i=1}^{|S|} e^{\Delta \beta f(x_i)}}, \tag{16}$$

6. Modifying successful EDAs with the empirical selection distribution

This section presents several successful EDAs which have been modified to accept the relative frequencies given by the empirical selection distribution. Continuous and discrete EDAs with univariate and multivariate models are presented. Table 5 presents the notation used.

$p(x,t)$	Search distribution in generation t.
n	Number of variables.
X_t	Population.
$\|X\|$	Population size.
$\hat{p}_j^S$	Frequency of x_i, according to $\hat{p}^S(x_i,t)$ in Table 2.
F_t	Population objective values.
X_{best}	An optimum approximation.
$I(x,k)$	Indicator, 1 if $x_i = k$, and 0 otherwise.

Table 5. Notation used for the empirical selection based EDAs.

A practical EDA based on the empirical selection distribution is presented in Table 6. Since the parameter computation in line 8 is the only section that should be changed according to the particular search distribution, a pseudo-code is separately presented for each algorithm. The selection procedure in line 6 must be understood as the computation of the empirical selection distribution (according to Equations in Table 2).

Practical EDA based on the Empirical Selection Distribution	
1	Initialize the search distribution model $p(x,0)$
2	Sampling(X_t ,$p(x,0)$)
3	Evaluation(F_t , X_t)
4	**While** the stopping criterion is not met **do**
5	Selection($\hat{p}^S(x,t)$, X_t, F_t)
6	Parameter_Computation($p(x,t)$, $\hat{p}^S(x,t)$, X_t, F_t)
7	Sampling(X_{t+1}, $p(x,t)$)
8	Evaluation(F_t, X_{t+1})
9	Elitism(X_{best}, F_{t+1}, X_{t+1})
10	**End While**
Ensure: An optimum approximation X_{best}.	

Table 6. Pseudo-code a practical EDA based on the empirical selection distribution.

```
1    For (i=1 to n) {
2        For (k=1 to mᵢ) {
3
```
$$b_{i,k} = \sum_{j=1}^{|X|} I(x_j,k) \cdot \hat{p}_j^S$$
```
4        }
5    }
```

Table 7. Pseudo-code for the ES-UMDA parameter computation.

6.1 UMDA

The Univariate Marginal Distribution Algorithm (UMDA) was introduced by Mühlenbein and PaaB (Mühlenbein & PaaB, 1996). It is based on the estimation of univariate marginal probabilities. This model considers that all variables are statistically independent. It uses the simplest model for discrete distributions. Each variable x_i has attached a probability

vector $b_{i,k}$, that is, the probability of x_i taking the value k, is: $b_{i,k} = p(x_i = k)$. Note that in the original UMDA the computation of the parameter $b_{i,k}$ is basically done by counting bits of a variable of the selected set.

The UMDA is quite simple to adapt to the relative frequencies given by the empirical selection distribution. The pseudo-code of the ES-UMDA (ES=Empirical Selection) parameter computation is shown in Table 7.

6.2 UMDA$_c$

The Univariate Marginal Distribution Algorithm for continuous domains (UMDA$_c$) was introduced by Larrañaga et al. (Larrañaga et al., 1999). It uses a univariate model, in the specific case of UMDA^G_c, there are n univariate Gaussian distributions (for a n-dimensional problem). Two parameters are needed for the Gaussian at each dimension i, the mean μ_i and the standard deviation σ_i. The computation of both parameters is simply done by weighting each point by its corresponding relative frequency (probability) as shown in Table 8.

1	For ($i=1$ to n) {
2	$$\mu = \sum_{j=1}^{|X|} \hat{p}_i^S \cdot x_{j,i}$$
3	$$\sigma_i^2 = \sum_{j=1}^{|X|} \hat{p}_i^S \cdot (x_{j,i} - \mu_i)^2$$
4	}

Table 8. Pseudo-code for the ES-UMDA$_c$ parameter computation.

6.3 K2-Bayesian-network based EDA

Bayesian networks have been successfully used in EDAs, for instance the Bayesian Optimization Algorithm (BOA) introduced by Pelikan et al. (Pelikan et al., 1999). A BOA-like algorithm based on the K2 algorithm (Cooper & Herskovits, 1992) is presented. The parameter computation has been modified to use the empirical selection distribution. The K2 is a greedy heuristic search method, for maximizing the probability $P(B_S,D)$ of the structure B_S and the data D. For maximizing $P(B_S,D)$ the K2 maximizes $g(i,\pi_i)$, which is a measure related with the probability of x_i given a set of parents π_i.

$$g(i,\pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)}{(N_{ij} + r_i - 1)} \prod_{k=1}^{r_i} N_{ijk}! , \tag{17}$$

where x_i has r_i possible discrete values. Each variable x_i in B_s has a set of parents, which are represented by a list of variables π_i. N_{ijk} is the number of cases in D in which the variable x_i has the value v_{ik}, and π_i is instantiated as w_{ij}. w_{ij} denotes the j-th unique instantiation of π_i relative to D, and q_i is the number of such unique instantiations of π_i. $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. For a deeper explanation of the K2 algorithm the reader is directed to (Cooper & Herskovits, 1992).

Learning a Bayesian network with the K2 is a counting process of point instances. Then, in order to compute Equation 17 an integer frequency for each point is needed. For obtaining an integer frequency, let us define a sample size for the selection method, say $|\hat{S}|$. When using the empirical selection distribution we have a relative frequency $\hat{p}_l^S$ associated with a point x_l, if that point is such a case for N_{ijk}, then, instead of summing a single case we will sum integer($|\hat{S}| \cdot \hat{p}_l^S$). As $|\hat{S}|$ grows, the approximation of the K2 Bayesian network to the empirical selection distribution will be more accurate, but the computational cost of Equation 17 increases as well. Once the K2 Bayesian network structure has been learned, the frequencies $|\hat{S}| \cdot \hat{p}_l^S$ must be used to compute the conditional probabilities.

6.4 EMNA$_{global}$

The EMNA$_{global}$ was introduced by Larrañaga et al. (Larrañaga et al., 2001). It is based on the estimation of a multivariate normal density function. This model can represent linear dependencies between normal variables in the covariance matrix. The pseudo-code of the ES-EMNA$_{global}$ parameter computation is shown in Table 9. Note that it is quite easy to insert the empirical selection distribution in univariate as well as multivariate algorithms, also it is inserted in discrete and continuous domains with a minimum analytical and computational effort.

$$
\begin{array}{ll}
1 & \text{For } (i=1 \text{ to } n) \ \{ \\[2mm]
2 & \quad \mu = \sum_{j=1}^{|X|} \hat{p}_i^S \cdot x_{j,i} \\[2mm]
3 & \quad \text{For } (k=1 \text{ to } i) \ \{ \\[2mm]
  & \qquad \sigma_{i,k} = \sum_{j=1}^{|X|} \hat{p}_i^S \cdot (x_{j,i} - \mu_i)(x_{j,k} - \mu_i) \\[2mm]
  & \qquad \sigma_{k,i} = \sigma_{i,k} \\[2mm]
  & \quad \} \\[2mm]
4 & \}
\end{array}
$$

Table 9. Pseudo-code for the ES-EMNA$_{global}$ parameter computation.

7. Experiments and performance analysis

This section provides a set of experiments to address the performance and advantages of the empirical selection distribution. Two kinds of experiments are presented:

1. Graphically we show that the search distribution based on the empirical selection distribution constitutes a robust and better approximation to the selection model.
2. Using the EMNA$_{global}$ (Larrañaga et al., 2001) and the BMDA (Pelikan & Mühlenbein, 1999), we show the impact of the empirical selection distribution on the EDA performance.

7.1 Graphical comparison

For better plots, the unidimensional objective function shown in Figure 3 is used. This analysis contrasts the three concepts discussed in the introduction: the ideal EDA, the

standard EDA, and the EDA with the empirical selection distribution, combined with the four discussed selection methods.

Even though the exact selection distribution can be computed for the ideal EDA, to perform the correct comparison we use a predefined Gaussian model for the three approaches. We show the Gaussian approximation with a very large population (considered as infinite), contrasted with the approximation computed by using the selected set from a standard sized population, and the approximation computed by using the empirical selection distribution. The experiments are designed as follows:

- **The ideal EDA**. A very large population (10^4 individuals) equally spaced is used, then, a larger selected set (10^5) is extracted using the selection method. Using the huge selected set (10^5 individuals) the parameters of the search distribution are computed. This approximation will be called the exact approximation. A special case is the truncation method which uses the same population size, but a smaller selected set is obtained by truncating at $\theta = 0.9$.

- **The standard EDA**. A relatively small population (30 individuals) is randomly drawn, the selection method is applied delivering the selected set used to compute the search distribution. Most of the selection methods add randomness to the procedure (except the truncation method which delivers a deterministic selected set), as a consequence the search distributions could differ after applying the selection method to the same population several times. Thus, we present the different search models computed when selecting 10 selected sets of 15 individuals from the same population.

- **The EDA with the empirical selection distribution**. Using the same population of 30 individuals used by the standard EDA, the empirical selection distribution is used to compute the search distribution parameters. Notice that the empirical selection distribution is unique for a given population as well as the search model delivered by this method.
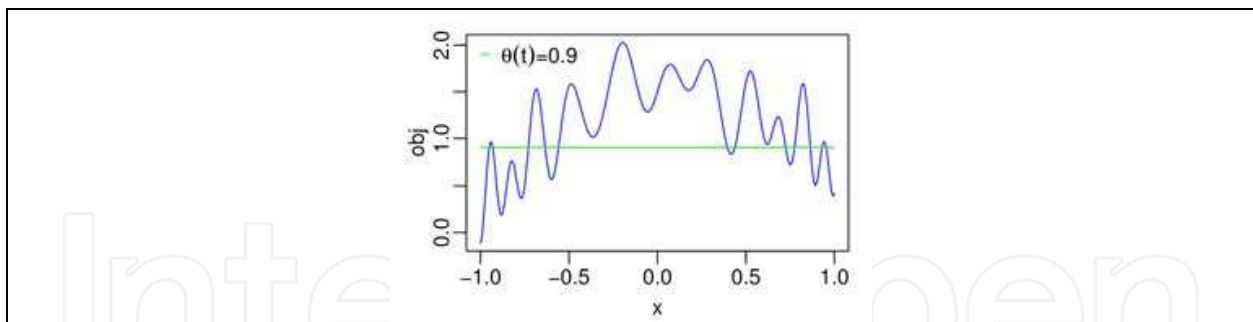

Fig. 3. Unidimensional objective function used for the graphical experiments.

Truncation selection. The truncation method shown in Figure 4(a) is basically the same for the standard EDA and the empirical selection distribution, because the points used are the same in both approximations. Thus, in this case the empirical selection performs at least as well as the standard selection method.

Boltzmann selection. As shown in Figure 4(b), the empirical selection method delivers a very good approximation to the exact model. The search models computed by the standard EDA approximation are a belt of Gaussians. The empirical selection approximation is at the middle of this belt, it is less affected by a single point or small set of points (robustness). It is possible that the randomness of the standard selection method guides the search to a better optimum approximation. In general, it is not the expected behaviour, given its difference

with the exact model, the small size of the selected set (usually 50% of the population) and the consequent loss of information which favors the tendency of being biased to sub-optimal regions. Also, it is expected that the behaviour of the Boltzmann selection varies according to the β value. The empirical selection computes the same search model from the same population, thus, a more stable behaviour is expected in contrast with the standard EDA. This could be especially useful when designing annealing schedules, because with the empirical selection distribution a similar performance under similar conditions is expected.
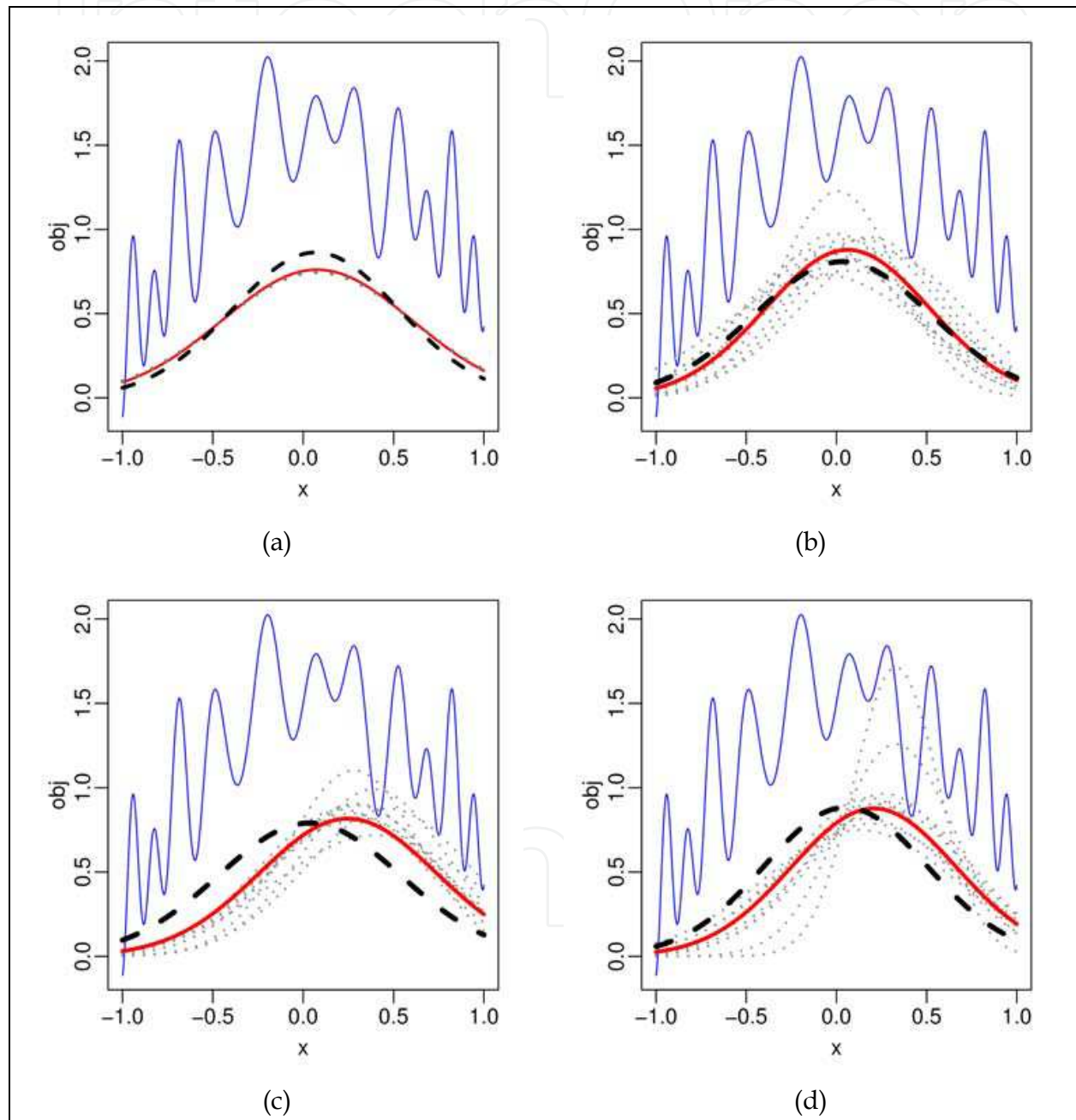

Fig. 4. Selection methods: (a) truncation, (b) Boltzmann, (c) proportional and (d) tournament. Search distributions for: 1) The exact approximation with dashed line, 2) the standard EDA approximation with doted lines, and 3) the empirical selection approximation with solid line. The objective function (solid line), and the population used for practica EDAs (circles).

Proportional selection. The proportional selection does not use extra parameters, therefore, no expensive parameter tuning is required. At the same time, however, there is no way to control the high selection pressure exerted over the fittest individuals that usually lead the population to premature convergence. Thus, a search distribution which better represents the selection method could be useful, also a more predictable search model is useful when tuning other controls such as the population size, because the expected behaviour could be better inferred. Figure 4(c), shows the search densities (points) computed when applying the selection method 10 times on the same population. Notice that these models are easily biased to different regions, most of them suboptimal.

Binary tournament selection. The binary tournament selection seems to be the most robust selection method according to Figure 4(d). The plot shows that this selection delivers Gaussians more similar than those delivered by other selection methods. This selection can be used as a reference for the expected performance of an EDA, considering the robustness, and the good performance of this selection according to the experiments in the next section. The parameter tuning for other selection methods can be done trying to obtain at least the same performance as the binary tournament selection. It is parameter free, and it is less sensible to large objective values in the population. As shown in Figure 4(d) the empirical selection distribution computes an accurate approximation when compared with the exact method.

7.2 Performance comparison

It has been previously shown that the estimation of the search distribution according to the empirical selection distribution accurately approximates the exact search distribution. Other claim is that the empirical selection distribution preserves some building blocks that can not be preserved by the standard EDA, due to the fact that the last one only uses a part of the population, while the empirical selection based EDA use the whole population in most of the cases. To support these arguments, we present experiments using the bivariate marginal distribution algorithm (BMDA) (Pelikan & Mühlenbein, 1999), and the $EMNA_{global}$ (Larrañaga et al., 2001), as well as the counterparts based on the empirical selection distribution: the ES-BMDA and ES-$EMNA_{global}$.

7.3 Test 1. The $EMNA_{global}$

We present a comparison using the $EMNA_{global}$ and three problems widely used to test EDAs performance (Larrañaga et al., 2001): Sphere, Griewangk and Rosenbrock. In order to maximize the functions and convert them to have positive objective values, an objective function $g(x)$ is adjusted as: $f(x)=-g(x)+1.8e7$, and $f(x)$ is used as the fitness function.

Experiments description: 30 independent runs are performed with 50 variables, the domain for all variables is [-600,600], the population size is 2000, selecting 1000 individuals for the standard EMNA. Truncation is 50% of the population, and elitism according to the general EDA in Table 6. The annealing schedule for the Boltzmann selection is fixed by initializing $\beta_0 = 0$, $\Delta\beta = 100/n$, and $\beta_t = \beta_{t-1} + \Delta\beta$. For the Rosenbrock function $\Delta\beta = 50/n$. For the Boltzmann and proportional selections the objective function is normalized by applying $f(x)=g(x)-g^{min}/(g^{max}-g^{min}+1)$, in order to avoid numerical problems.

Stopping criterion: 300 000 function evaluations.

Results: Table 10 compares the results obtained by the standard EMNA (denoted by St.), and the Empirical Selection based EMNA (denoted by ES). We report the mean and standard deviation of the objective function as well as the result of a hypothesis test based

n		Truncation	Boltzmann	Proportional	Binary Tournament
			Sphere		
10	St	5.048e-9 (1.158e-9)	2.168e-6 (7.745e-7)	7.703e-2 (2.666e-2)	6.219e-9 (1.234e-9)
	ES	5.159e-9 (1.105e-9)	2.370e-6 (5.901e-7)	6.738e-2 (2.347e-2)	5.762e-9 (1.395e-9)
	Hyp. Test	N	N	N	N
20	St	3.767e-8 (7.306e-9)	9.924e-1 (1.356)	7.365 (2.314)	6.607e-8 (1.166e-8)
	ES	3.790e-8 (6.003e-9)	1.092 (2.298e+0)	6.661 (1.071)	7.628e-8 (1.074e-8)
	Hyp. Test	N	N	N	St
30	St	1.093e-7 (1.123e-8)	2.725e+1 (1.804e+1)	2.018e+2 (1.524e+2)	1.660e-5 (1.914e-6)
	ES	1.091e-7 (7.991e-9)	1.726e+1 (1.503e+1)	5.549e+1 (1.005e+1)	3.520e-5 (3.874e-6)
	Hyp. Test	N	ES	ES	St
40	St	2.213e-7 (1.673e-8)	1.060e+2 (4.624e+1)	1.283e+3 (6.554e+2)	1.889e-3 (2.983e-3)
	ES	2.239e-7 (1.682e-8)	6.394e+1 (4.757e+1)	2.142e+2 (2.060e+1)	1.094e-3 (1.099e-4)
	Hyp. Test	N	ES	ES	ES
50	St	1.999e-6 (2.266e-7)	4.289e+2 (1.779e+2)	4.735e+3 (1.379e+3)	3.543 (5.801)
	ES	1.712e-6 (2.082e-7)	1.826e+2 (9.569e+1)	5.357e+2 (6.471e+1)	1.111e-2 (1.032e-3)
	Hyp. Test	ES	ES	ES	ES
			Griewnagk		
10	St	9.356e-10 (2.077e-10)	0.36098 (0.18955)	0.3194 (0.05671)	1.062e-9 (3.608e-10)
	ES	9.231e-10 (2.251e-10)	0.17173 (0.09324)	0.3343 (0.06324)	1.114e-9 (2.647e-10)
	Hyp. Test	N	ES	N	N
20	St	4.921e-9 (5.351e-10)	0.19738 (0.19323)	0.6325 (0.07413)	1.003e-8 (1.788e-9)
	ES	4.968e-9 (6.329e-10)	0.06382 (0.08591)	0.6553 (0.04763)	1.207e-8 (1.526e-9)
	Hyp. Test	N	ES	ES	St
30	St	1.044e-8 (1.007e-9)	0.13667 (0.11265)	1.0302 (0.06677)	1.125e-6 (1.846e-7)
	ES	1.004e-8 (9.885e-10)	0.03527 (0.05150)	0.9466 (0.03131)	2.223e-6 (3.029e-7
	Hyp. Test	N	ES	ES	St
40	St	1.817e-8 (1.357e-9)	0.26945 (0.12897)	1.3946 (0.19001)	6.056e-4 (2.088e-3)
	ES	1.809e-8 (1.694e-9)	0.05683 (0.05088)	1.0543 (0.00679)	4.868e-5 (5.506e-6)
	Hyp. Test	N	ES	ES	ES
50	St	7.568e-8 (1.352e-8)	0.54303 (0.24312)	2.3571 (0.46186)	1.036e-1 (1.265e-1)
	ES	6.852e-8 (1.364e-8)	0.17259 (0.07979)	1.1365 (0.02057)	3.899e-4 (3.021e-5)
	Hyp. Test	ES	ES	ES	ES
			Rosenbrock		
10	St	9.292 (7.714)	6.795 (0.5562)	1.125e+05 (65566)	7.647 (0.2958)
	ES	8.929 (3.113)	6.862 (0.4907)	1.133e+05 (56613)	7.880 (0.7487)
	Hyp. Test	N	N	N	St
20	St	20.131 (5.753)	19.404 (2.4900)	1.313e+07 (7154997)	17.915 (0.9468)
	ES	19.339 (4.762)	18.386 (1.1359)	9.138e+06 (2823051)	17.657 (0.8191)
	Hyp. Test	N	ES	ES	N
30	St	28.865 (1.218)	35.728 (7.9373)	1.656e+08 (154793875)	31.681 (8.1603)
	ES	30.695 (6.149)	35.246 (5.2851)	5.800e+07 (16388688)	27.852 (0.6024)
	Hyp. Test	N	ES	ES	ES
40	St	42.413 (4.465)	71.699 (18.6654)	5.469e+08 (269436266)	84.128 (52.2806)
	ES	41.709 (3.324)	53.984 (8.8081)	1.845e+08 (53383769)	40.544 (3.1687)
	Hyp. Test	N	ES	ES	ES
50	St	65.604 (13.422)	1109.840 (3713.7494)	1.183e+09 (641262869)	2594.724 (3675.9199)
	ES	60.987 (9.248)	93.034 (16.7544)	3.695e+08 (82888904)	61.753 (4.9202)
	Hyp. Test	N	ES	ES	ES

Table 10. Performance comparison between the standard EMNA$_{global}$ (St) and the empirical selection distribution based EMNA$_{global}$ (ES).

on the Bootstrap methodology with a significance of 5% (Efron & Tibshirani, 1993). We test that the mean of the ES-based EMNA$_{global}$ is less than the mean of the standard EMNA$_{global}$, and vice versa. An **N** below the mean and standard deviation means that neither is better, **St** means that the standard EMNA$_{global}$ is the best, an ES that the empirical selection based EMNA$_{global}$ is the best. The conditions of the experiment are maintained while the dimensions are growing, the purpose is to show that the empirical selection in general performs better, and can use the information efficiently. Note that the standard EDA (St-EDA) and the Empirical Selection EDA (ES-EDA) are very similar in lower dimensions, but for higher dimensions the ES-EDA clearly outperforms the St-EDA. The explanation is that the population size is not growing with the dimensions, thus, for lower dimensions the selected set size is enough to adequately estimate the search distribution, but for higher dimensions the information is not sufficient, or it is not efficiently used. The runs where the St-EDA outperforms the ES-EDA (tournament 20 and 30 variables), can be explained by selection pressure issues. The St-EDA uses a subset of the best individuals in the population, no matter which selection method is used. On the other hand, the ES-EDA uses all the points, so for the same parameters the variance of the ES-EDA will be greater than the St-EDA, because the whole population covers a wider region than the selected set. So, the convergence of the St-EDA is faster, and the exploration is concentrated in a smaller region, resulting in a poor performance for higher dimensions as shown in Table 10. Notice that even the hypothesis test says that the St-EDA is better in 4 cases, the ES-EDA finds solutions very close to the optimum in all cases.

7.4 Test 2. The BMDA

The BMDA works in discrete domains, in this case a binary domain. It builds a graphical model with bivariate dependencies, the dependencies are considered according a χ^2 hypothesis test. The χ^2 computation uses empirical probabilities, thus, the ES-BMDA computes these empirical probabilities by summing the relative frequencies $\hat{p}_i^S$ given by the empirical selection distribution formulae. For example, if we need to compute the marginal probability of the variable x_j takes the value of 1, say $p(x_j = 1)$, if an individual i in the population is an instance of $x_j = 1$ then we sum $\hat{p}_i^S$, the standard procedure at the end divides the sum over the total number of individuals, when using the empirical selection this normalization is not necessary. This test is performed by using the deceptive function of order 3 shown in Equation 18, which was also used for the experiments in the original paper of the BMDA (Pelikan & Mühlenbein, 1999).

$$f_{3deceptive}(x, \pi_i) = \sum_{i=1}^{\frac{n}{3}} f_3(x_{\pi(3(i-1)+1)} + x_{\pi(3(i-1)+2)} + x_{\pi(3(i-1)+3)}) ,$$
(18)

where x is a bit string, π is any permutation of order n, and f_3 is defined in Equation 19.

$$f_3(u) = \begin{cases} 0.9 & \text{if } u = 0 \\ 0.8 & \text{if } u = 1 \\ 0 & \text{if } u = 2 \\ 1 & \text{otherwise} \end{cases} ,$$
(19)

For n = 30 we use π = { 6, 27, 18, 20, 28, 16, 1, 23, 24, 3, 2, 13, 8, 5, 17, 11, 29, 15, 30, 9, 25, 12, 19, 22, 4, 10, 14, 21, 26). For n = 60, π = { 49, 31, 40, 59, 5, 23, 57, 37, 47, 19, 27, 30, 8, 56, 3, 36, 45, 17, 41 , 33, 21, 53, 39, 51, 50, 29, 16, 10, 24, 55, 15, 32, 7, 13, 2, 52, 14, 60, 12, 22, 34, 25, 35, 1, 28, 18, 20, 9, 38, 26, 46, 58, 42, 43, 44, 48, 6, 4, 11}. The deceptive function correlates subsets of 3 variables. Pelikan and Mühlenbein (Pelikan & Mühlenbein, 1999) shown that the BMDA is capable of solving this problem that can not be solved by the simple GA. Our purpose is to show that the ES-BMDA is more efficient than the standard BMDA and can use the information in the whole population, thus it is capable of finding and using more correlations to approximate the optimum.

Experiments description: We reproduce the settings of the original BMDA paper (Pelikan & Mühlenbein, 1999). 30 independent runs are performed with 30 and 60 variables, the population size is 1300, selecting 1000 individuals for the standard EDA. 50% of the best individuals are preserved no matter which selection is used (as in the original paper) for the BMDA and the ES-BMDA. Truncation is 50% of the population, and elitism according to the general EDA in Table 6. The annealing schedule for the Boltzmann selection is fixed by initializing $\beta_0 = 0$, $\Delta\beta = 1$, and $\beta_t = \beta_{t-1} + \Delta\beta$. For the Boltzmann and proportional selections the objective function is normalized by applying $f(x)=g(x)-g^{min}/(g^{max}-g^{min}+1)$, in order to avoid numerical problems.

Stopping criterion: To reproduce the results of the original BMDA we use the ordering parameter as stopping criterion (Pelikan & Mühlenbein, 1999). The ordering parameter is defined in Equation 20, where p is the vector of univariate marginal frequencies $p_i(1)$. When $\chi(p) > 0.95$ we stop the algorithm that means that the univariate probabilities are almost 1 or 0.

$$\chi(p) = \sum_{i=1}^{n} \left(p_i(1) - \frac{1}{2} \right)^2 ,$$
(20)

Results: Table 11 compares the results obtained by the standard BMDA (denoted by St.), and the Empirical Selection based BMDA (denoted by ES). We report the mean and standard deviation of the evaluations and objective function, as well as the result of a hypothesis test based on the Bootstrap methodology with a significance of 5% (Efron & Tibshirani, 1993). We test that the mean of the ES-BMDA is less than the mean of the standard BMDA, and vice versa, for 30 and 60 variables. For the evaluations test we only use the runs in which the optimum is found. The evaluations comparison is not perform for 60 variables, because most of the runs did not find the optimum. The results bring out evidence about the arguments that the empirical selection based EDAs are more efficient, and that the use of the information in the whole population is an advantage to build optimal solutions, because as shown the ES-BMDA needs less evaluations than the St-BMDA, also, the ES-BMDA is more effective to find the optimum in the 60 dimension runs, so, with the same resources (population size) the ES-BMDA can deliver better results.

8. Perspectives, future work and conclusions

EDAs researchers have approximated the selection distribution since the first approaches (Mühlenbein, 1997; Bosman & Thierens, 2000). This chapter proposes a general method for this purpose. Most of the search distributions used by EDAs are parametric, and the

n		Truncation	Boltzmann	Proportional	Bin.ary Tournament
Deceptive Function of Order 3,Objective Function					
30	St	10 (0)	9.993333(0.02537081)	9.997 (0.01825742)	9.997 (0.01825742)
	ES	10 (0)	10 (0)	9.997 (0.01825742)	9.997 (0.01825742)
	Hyp. Test	N	N	N	N
60	St	19.79 (0.1213431)	19.73 (0.1368362)	19.60667 (0.1818171)	19.81 (0.1028893)
	ES	19.84 (0.1101723)	19.85(0.1252584)	19.80667(0.1080655)	19.85333(0.1166585)
	Hyp. Test	N	ES	ES	N
Evaluations					
30	St	14841.67(783.7887)	14300(482.8079)	24526.67(1497.223)	15925(609.5151)
	ES	14755(619.0023)	14148.33(441.2919)	22923.33(869.2777)	15578.33(578.4467)
	Hyp. Test	N	N	ES	ES
Number of times the optimum was reached (in 30 runs)					
30	St	30	28	29	29
	ES	30	30	29	29
60	St	3	2	0	3
	ES	6	9	3	7

Table 11. Performance comparison between the standard BMDA (St) and the empirical selection distribution based BMDA (ES).

parameters are learned from a sample by counting frequencies. Thus, in addition to the presented algorithms, any other search distribution which uses frequencies can be modified to use the empirical selection distribution. For instance bivariate models (Pelikan & Mühlenbein, 1999) and histograms (Tsutsui et al., 2001) are completely based on frequencies. Another important line of study are clustering based algorithms (Larrañaga & Lozano, 2001), for example the k-means algorithm is based on distances. When using the empirical selection distribution in clustering, instead of using a single point in the position x_i, we use its relative frequency $\hat{p}^S(x_i,t)$. This measurement will move the mean of the cluster to the regions with highest fitness values, helping to perform a better search.

Important issues in EDAs such as diversity and premature convergence can be tackled using the empirical selection distribution. Studies on convergence phases (Grahl et al., 2007) have shown that the maximum likelihood estimated variance might not be the best to perform an optimum search. Thus, a possible line of work is: how to favor diversity using the empirical selection distribution? A possibility is by simply modifying the fitness function into the empirical selection distribution formulae. This line of work may be an alternative to recent proposals on variance scaling (Grahl et al., 2006; Bosman et al., 2007).

Multi-objective applications and how to insert the Pareto dominance in the selection distribution is another possible research line. With respect to this topic the Pareto ranking seems to be the most natural way of tackling this important issue.

Ever since the very first proposed EDAs (Baluja, 1994) to the most recent works (Pelikan et al., 2008), incremental learning has been applied to the learning distribution phase. Future work must contemplate how to insert the empirical selection distribution into incremental approaches, or how to use historical or previous empirical selection distributions.

The selection methods presented are just a sample of the possibilities, other methods such as combined truncation-proportional, truncation-Boltzmann, 3-tournament, etcetera must be explored.

Finally, the advantages of the presented method are summarized as follows:

- it is easy to implement.
- It has a wide range of applications.
- It has low computational as well as analytical cost.
- It avoids being wrongly biased by a single solution or a small set.
- It uses all the information from the population to accurately approximate the selection distribution.
- The perspectives, future use and applications are promising, and the possible lines of work are really extensive.

9. References

Baluja , S. (1994). Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning. *Technical report, Carnegie Mellon University*, Pittsburgh, PA, USA, 1994.

Bosman, P. A. N.; Grahl, J. & Rothlauf, F. (2007). SDR: A Better Trigger for Adaptive Variance Scaling in Normal EDAs. *In GECCO '07: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pp. 516-522, ISBN 978-1-59593-697-4, London, UK. July, 2007, ACM.

Bosman, P. A. N. & Thierens, D. (2000). Expanding from Discrete to Continuous Estimation of Distribution Algorithms: The IDEA. *In PPSN VI: Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*, pp. 767-776, ISBN 3-540-41056-2, Paris, France, 2000. Springer-Verlag.

Cooper, G. & Herskovits, E. (1992). A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, Vol. 9, No. 4, (october, 1992), pp. 309-347, ISSN:0885-6125.

Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, Monographs on Statistics and Applied Probability, ISBN-13: 9780412042317, New York, 1993.

Grahl, J.; Bosman, P.A. & Rothlauf, F. (2006). The Correlation-Triggered Adaptive Variance Scaling IDEA. *In GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pp. 397-404, ISBN 1-59593-010-8, Seattle, Washington, USA, July, 2006. ACM.

Grahl, J.; Bosman, P. A. N. & Minner, S. (2007). Convergence Phases, Variance Trajectories, and Runtime Analysis of Continuos EDAs. *In GECCO'07: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pp. 516-522, ISBN 978-1-59593-697-4, London, UK. July, 2007, ACM.

Larrañaga, P.; Etxeberria, R.; Lozano, J. & Peña, J. (1999). Optimization by Learning and Simulation of Bayesian and Gaussian Networks. *Technical Report EHU-KZAA-IK-4/99*, Department of Computer Science and Artificial Intelligence, University of the Basque Country, 1999.

Larrañaga, P. & Lozano, J. A. (2001). *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Kluwer Academic Publishers, ISBN 978-0-7923-7466-4, Norwell, MA, USA, 2001.

Lozano, J. A.; Larrañaga, P.; Inza, I. & Bengoetxea, E. (2006). Towards a New Evolutionary Computation, *Studies in Fuzziness and Soft Computing*. Vol. 192, ISBN 978-3-540-29006-3, Springer, 2006.

Müehlenbein, H.; Bendisch, J.; & Voight, H.M. (1996). From recombination of genes to the estimation of distributions II. Continuous parameters. *In PPSN IV: Proceedings of the 4th International Conference on Parallel Problem Solving from Nature*, pp. 188-197, ISBN 978-3-540-61723-5, Berlin, Germany, September, 1996, Springer-Verlag.

Mühlenbein, H. (1997). The Equation for Response to Selection and Its Use for Prediction. *Evolutionary Computation*, Vol. 5, No. 3, September, 1997,pp. 303-346, ISSN 1063-6560.

Mühlenbein, H. & PaaB, G. (1996). From Recombination of Genes to the Estimation of Distributions I. Binary Parameters. *In PPSN IV: Proceedings of the 4th International Conference on Parallel Problem Solving from Nature*, pp. 178-187, ISBN 978-3-540-61723-5, Berlin, Germany, September, 1996. Springer-Verlag.

Pelikan, M.; Goldberg, D. E.; & Cantú-Paz, E. (1999). BOA: The Bayesian Optimization Algorithm. In W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, editors, *Proceedings of the Genetic and Evolutionary Computation Conference GECCO-99*, Vol. I, pp. 525-532, Orlando, FL, USA, 1999. Morgan Kaufmann Publishers, San Fransisco, CA.

Pelikan, M. & Mühlenbein, H. (1999). The Bivariate Marginal Distribution Algorithm. *Advances in Soft Computing Engineering Design and Manufacturing*, pp. 521-535, 1999.

Pelikan, M. ; Sastry, K. & Cantu-Paz, E. (2006). Scalable Optimization via Probabilistic Modeling, *Studies in Computational Intelligence*. Vol. 33, Springer, ISBN: 978-3-540-34953-2.

Pelikan, M.; Sastry, K. & Goldberg, D. E. (2008). iBOA: The Incremental Bayesian Optimization Algorithm. *In GECCO '08: Proceedings of the 2008 Conference on Genetic and Evolutionary Computation*, pp. 455-462, ISBN 978-1-60558-131-6, Atlanta, USA, July, 2008, ACM.

Soto, M. & Ochoa, A. (2000). A Factorized Distribution Algorithm based on Polytrees. *In Proceedings of the 2000 Congress on Evolutionary Computation (CEC 2000)*, pp. 232-237, ISBN: 0-7803-6375-2, La Jolla, CA, USA, July, 2000, IEEE.

Tsutsui, S.; Pelikan, M.; & Goldberg, D. E. (2001). Evolutionary Algorithm using Marginal Histogram Models in Continuous Domain. *Technical Report 2001019, Illinois Genetic Algorithms Laboratory*, 2001.

Yunpeng, C.; Xiaomin, S. & Peifa, J. (2006). Probabilistic Modeling for Continuous EDA with Boltzmann selection and Kullback-Leibeler Divergence. *In GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*, pp. 389-396, ISBN 1-59593-010-8, Seattle, Washington, USA, July, 2006. ACM.

Zhang, Q. & Mühlenbein, H. (2004). On the Convergence of a Class of Estimation of Distribution Algorithms. *IEEE Transactions on Evolutionary Computation*, Vol. 8, No. 2, (April, 2004), pp. 127-136, ISSN: 1089-778X.

New Achievements in Evolutionary Computation

Edited by Peter Korosec

Evolutionary computation has been widely used in computer science for decades. Even though it started as far back as the 1960s with simulated evolution, the subject is still evolving. During this time, new metaheuristic optimization approaches, like evolutionary algorithms, genetic algorithms, swarm intelligence, etc., were being developed and new fields of usage in artificial intelligence, machine learning, combinatorial and numerical optimization, etc., were being explored. However, even with so much work done, novel research into new techniques and new areas of usage is far from over. This book presents some new theoretical as well as practical aspects of evolutionary computation. This book will be of great value to undergraduates, graduate students, researchers in computer science, and anyone else with an interest in learning about the latest developments in evolutionary computation.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

S. Ivvan Valdez, Arturo Hernández and Salvador Botello (2010). Efficient Estimation of Distribution Algorithms by Using the Empirical Selection Distribution, New Achievements in Evolutionary Computation, Peter Korosec (Ed.), ISBN: 978-953-307-053-7, InTech, Available from: http://www.intechopen.com/books/new-achievements-in-evolutionary-computation/efficient-estimation-of-distribution-algorithms-by-using-the-empirical-selection-distribution

INTECH
open science | open minds