
A Comparison of Selection Schemes Used in Evolutionary Algorithms

Tobias Blickle

Computer Engineering and
Networks Lab (TIK)
Swiss Federal Institute
of Technology
Gloriastrasse 35
CH-8092 Zurich, Switzerland
blickle@tik.ee.ethz.ch

Lothar Thiele

Computer Engineering and
Networks Lab (TIK)
Swiss Federal Institute
of Technology
Gloriastrasse 35
CH-8092 Zurich, Switzerland
thiele@tik.ee.ethz.ch

Abstract

Evolutionary algorithms are a common probabilistic optimization method based on the model of natural evolution. One important operator in these algorithms is the selection scheme, for which in this paper a new description model, based on fitness distributions, is introduced. With this, a mathematical analysis of tournament selection, truncation selection, ranking selection, and exponential ranking selection is carried out that allows an exact prediction of the fitness values after selection. The correspondence of binary tournament selection and ranking selection in the expected fitness distribution is proved. Furthermore, several properties of selection schemes are derived (selection intensity, selection variance, loss of diversity), and the three selection schemes are compared using these properties.

Keywords

Selection, genetic algorithms, evolutionary algorithms, diversity, selection intensity, tournament selection, truncation selection, linear ranking.

1. Introduction

Evolutionary algorithms (EAs) are probabilistic search algorithms characterized by the fact that a number N of potential solutions (called *individuals* $\mathcal{J}_i \in \mathbf{J}$, where \mathbf{J} represents the space of all possible individuals) of the optimization problem simultaneously sample the search space. This *population* $\tilde{\mathcal{J}} = \{\mathcal{J}_1, \mathcal{J}_2, \dots, \mathcal{J}_N\} \in \mathbf{J}^N$ is modified according to the natural evolutionary process: after initialization, selection $\omega : \mathbf{J}^N \mapsto \mathbf{J}^N$ and variation $\Xi : \mathbf{J}^N \mapsto \mathbf{J}^N$ are executed in a loop until some termination criterion is reached. Each run of the loop is called a *generation*, and $\tilde{\mathcal{J}}$ denotes the population at generation τ .

The selection operator is intended to improve the average quality of the population by giving individuals of higher quality a higher probability to be copied into the next generation. Selection thereby focuses the search on promising regions in the search space. The quality of an individual is measured by a fitness function $f : \mathbf{J} \mapsto \mathbf{R}$. The assumption thereby is that better individuals are more likely to produce better offspring, that is, there is a correlation between parental fitness and offspring fitness. In population genetics this correlation is named *heritability*. Without heritability, selection of better individuals makes no sense.

A nice feature of the selection mechanism is its independence of the representation of the individual, as only the fitness values of the individuals are taken into account. This simplifies the analysis of the selection methods and allows a comparison that can be used in all flavors of EAs.

Many selection mechanisms are known in EAs. Partly they are introduced according to a mathematical analysis (e.g., fitness proportional selection) partly based on similarities in nature (e.g., tournament selection). The principle of selection is very general, and different methodologies of selection schemes can be observed.

The balance between exploitation and exploration can be adjusted either by the selection pressure of the selection operator or by the recombination operator (e.g., by the probability of crossover). Because this balance is critical for the behavior of the EA, it is of great interest to know the properties of the selection and variation operators to understand their influence on the behavior of the EA.

In this paper a description of selection schemes is used based on the fitness distribution of the population before and after selection, as introduced in Blickle and Thiele (1995). This framework is used to analyze common selection schemes. In particular, tournament selection, truncation selection, and linear and exponential ranking selection will be considered. A large part of the analysis is also applicable to steady-state selection schemes because the selection probabilities are the same as in the generational case.

The methodology will be described next in Section 2. In Section 3 the most important results from Blickle and Thiele (1995) concerning tournament selection are collected. In the subsequent sections an analysis of truncation selection and linear and exponential ranking is given. Finally, the selection schemes are compared in Section 7.

2. Description of Selection Schemes

2.1 Related Work

Work has been done to classify the different selection schemes, such as *fitness proportional selection*, *ranking selection*, and *tournament selection*. Because the focus of this work is on a theoretical investigation, empirical comparisons are neglected.

Goldberg and Deb (1991) introduced the term *takeover time*, which is the number of generations needed for a single best individual to fill up the whole generation if no crossover or mutation is used. Hence, this analysis is restricted to the behavior of the best individual. Bäck (1994a,b) has analyzed the most prominent selection schemes used in EAs with respect to their takeover time.

Mühlenbein and Schlierkamp-Voosen (1993) used the *selection intensity* in the so-called *breeder genetic algorithm* (BGA) to measure the progress of the population, where selection intensity describes the change in the average fitness of the population due to selection. They derive the selection intensity for proportional selection and truncation selection. Here this work will be extended to tournament selection and linear and exponential ranking selection.

An analysis based on the behavior of the best individual or on the average population fitness describes only one aspect of a selection method. Here, a selection scheme is described by its interaction on the distribution of fitness values. From this description several properties can be derived (e.g., the behavior of the best or average individual).

The characterization of the population by its fitness distribution has also been used by other researchers, but in a more informal way. Mühlenbein and Schlierkamp-Voosen (1993) used the fitness distribution to calculate some properties of truncation selection. Shapiro,

Prügel-Bennett, and Rattray (1994) used a statistical mechanics approach to describe the dynamics of a genetic algorithm that also makes use of fitness distributions.

The consequent use of fitness distributions has many advantages:

- Fitness distributions are a powerful framework, giving a unified view of the selection schemes and allowing several, until now, independently and isolated obtained aspects of these selection schemes (e.g., selection intensity or takeover time) to be derived with a single methodology.
- Several new properties of selection schemes can be derived for the first time, such as
 - the expected fitness distribution of tournament selection (Theorem 3.1), truncation selection (Theorem 4.1), linear ranking selection (Theorem 5.1), and exponential ranking selection (Theorem 6.1);
 - a new important characteristic of selection schemes, called *loss of diversity* (Definition 2.8);
 - an important new property of tournament selection (the concatenation of tournament selection—Theorem 3.3);
 - new approximations for properties of tournament selection (selection intensity and the selection variance); and
 - new properties of exponential ranking selection (selection intensity, selection variance, and loss of diversity).
- The convergence of an EA optimizing the OneMax function can be predicted for several selection schemes.
- The unified description enables a new kind of comparison of selection schemes. In Section 7 for the first time several properties can be compared simultaneously.

These results are presented in the next sections. First, all necessary definitions are introduced.

2.2 Fitness Distribution

For selection only the fitness values of the individuals are taken into account. Hence, the state of the population is completely described by the fitness values of all individuals. In the following analysis, larger fitness values are “better” fitness values (i.e., fitness maximization is assumed).

There exists only a finite number of different fitness values f_1, \dots, f_n , ($n \leq N$), and the state of the population can also be described by the values $s(f_i)$ that represent the number of occurrences of the fitness value f_i in the population.

DEFINITION 2.1 (Fitness Distribution): *The function $s : \mathbf{R} \mapsto \mathbf{Z}_0^+$ assigns to each fitness value $f \in \mathbf{R}$ the number of individuals in a population $\vec{f} \in \mathbf{J}^N$ carrying this fitness value, where s is called the fitness distribution of a population \vec{f} .*

It is possible to describe a selection method as a function that transforms a fitness distribution into another fitness distribution.

DEFINITION 2.2 (Selection Method): A selection method Ω is a function that transforms a fitness distribution s into a new fitness distribution s' :

$$s' = \Omega(s, \text{par_list}) \quad (1)$$

where par_list is an optional parameter list of the selection method.

Because the selection methods are probabilistic, the next definition is introduced.

DEFINITION 2.3: $\Omega^*(s, \text{par_list})$ denotes the expected fitness distribution after applying the selection method Ω to the fitness distribution s . The notation $s^* = \Omega^*(s, \text{par_list})$ will be used as an abbreviation.

It is interesting to note that it is also possible to calculate the variance of the resulting distribution.

THEOREM 2.1: The variance in obtaining the fitness distribution s' is

$$\sigma_s^2 = s^* \left(1 - \frac{s^*}{N} \right) \quad (2)$$

PROOF: $s^*(f_i)$ denotes the expected number of individuals with fitness value f_i after selection. It is obtained by doing N experiments of the form “select an individual from the population using a certain selection mechanism.” Hence, the selection probability of an individual with fitness value f_i is given by $p_i = s^*(f_i)/N$. For each fitness value there exists a Bernoulli trial; that is, an individual with fitness f_i is selected. Since the variance of a Bernoulli experiment with N trials is given by $\sigma^2 = Np(1 - p)$, Equation 2 is obtained using $p = p_i$. \square

The index s in σ_s stands for “sampling,” since it is the mean variance due to the sampling of the finite population. The variance of Equation 2 is obtained by performing the selection method in N independent experiments. It is possible to reduce the variance almost completely by using more sophisticated sampling algorithms to select the individuals. The “stochastic universal sampling” (SUS) algorithm of Baker (1987), which is an optimal sampling algorithm, is introduced in Section 7, when the different selection schemes are compared.

DEFINITION 2.4 (Cumulative Fitness Distribution): Let n be the number of unique fitness values and $f_1 < \dots < f_{n-1} < f_n$ ($n \leq N$) the ordering of the fitness values, with f_1 denoting the worst fitness and f_n the best fitness occurring in the population. $S(f_i)$ denotes the number of individuals with fitness value f_i or worse and is called the cumulative fitness distribution, that is,

$$S(f_i) = \begin{cases} 0 & : i < 1 \\ \sum_{j=1}^{j=i} s(f_j) & : 1 \leq i \leq n \\ N & : i > n \end{cases} \quad (3)$$

EXAMPLE 2.1: As an example of a discrete fitness distribution, the initial fitness distribution of the 6-multiplexer problem is used (Kozá, 1992) with a population size of $N = 250$. Worse individuals have lower fitness values. Figure 1 shows the distribution $s(f)$.

Next, the distribution $s(f)$ will be described as a continuous distribution $\bar{s}(f)$. The range of the function $\bar{s}(f)$ is $f_0 < f \leq f_n$, using the same notation as in the discrete case.

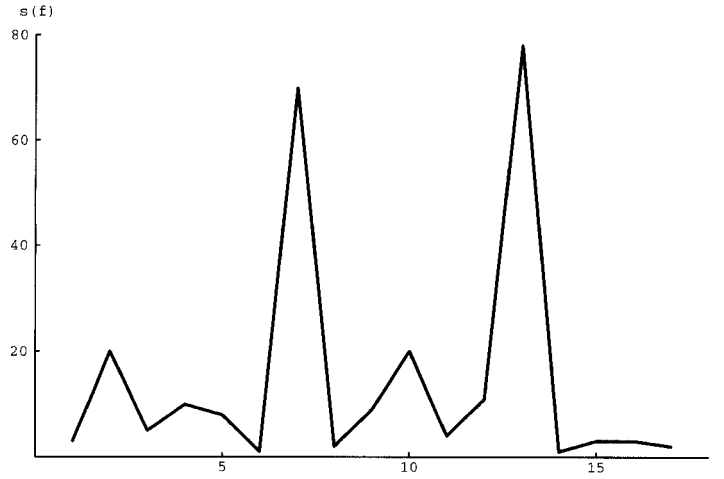


Figure 1. The fitness distribution $s(f)$ for the 6-multiplexer problem ($N = 250$).

All functions in the continuous case will be denoted with an overbar, for example, $\bar{s}(f)$ instead of $s(f)$. Sum signs are replaced by integral signs, for example,

$$\bar{S}(f) = \int_{f_0}^f \bar{s}(x) dx \quad (4)$$

denotes the continuous cumulative fitness distribution.

EXAMPLE 2.2: *As an example of a continuous fitness distribution, the Gaussian distribution $\mathcal{G}(\mu, \sigma)$ with*

$$\mathcal{G}(\mu, \sigma)(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (5)$$

is chosen.

The aspects of the fitness distribution that will be compared are now introduced. The definitions given all refer to continuous distributed fitness values. Although “discrete calculations” are more general and accurate, and the results of the continuous case can be derived from the discrete case, we restrict ourselves to the continuous case. The main reason for this restriction is that discrete calculations are complicated and messy, and the continuous case is simpler to work with. Nevertheless, the results for the continuous case are relevant for the exact discrete case. As an example, a result obtained with the continuous formulation is experimentally verified with discrete distributions in Section 7.

2.3 Average Fitness

DEFINITION 2.5 (Average Fitness): \bar{M} denotes the average fitness of the population before selection, and \bar{M}^* denotes the expected average fitness after selection:

$$\bar{M} = \frac{1}{N} \int_{f_0}^{f_n} \bar{s}(f) f df \quad (6)$$

$$\bar{M}^* = \frac{1}{N} \int_{f_0}^{f_n} \bar{s}^*(f) f df \quad (7)$$

2.4 Fitness Variance

DEFINITION 2.6 (Fitness Variance): *The fitness variance $\bar{\sigma}^2$ denotes the variance of the fitness distribution $\bar{s}(f)$ before selection, and $(\bar{\sigma}^*)^2$ denotes the variance of the fitness distribution $\bar{s}^*(f)$ after selection:*

$$\bar{\sigma}^2 = \frac{1}{N} \int_{f_0}^{f_c} \bar{s}(f) (f - \bar{M})^2 df = \frac{1}{N} \int_{f_0}^{f_c} f^2 \bar{s}(f) df - \bar{M}^2 \quad (8)$$

$$(\bar{\sigma}^*)^2 = \frac{1}{N} \int_{f_0}^{f_c} \bar{s}^*(f) (f - \bar{M}^*)^2 df = \frac{1}{N} \int_{f_0}^{f_c} f^2 \bar{s}^*(f) df - \bar{M}^{*2} \quad (9)$$

Note the difference of this variance in relation to the variance in obtaining a certain fitness distribution characterized by Theorem 2.1.

2.5 Reproduction Rate

DEFINITION 2.7 (Reproduction Rate): *The reproduction rate $\bar{R}(f)$ denotes the ratio of the number of individuals with a certain fitness value f after and before selection:*

$$\bar{R}(f) = \begin{cases} \frac{\bar{s}^*(f)}{\bar{s}(f)} : \bar{s}(f) > 0 \\ 0 : \bar{s}(f) = 0 \end{cases} \quad (10)$$

A reasonable selection method should favor good individuals by assigning them a reproduction rate $\bar{R}(f) > 1$ and punish bad individuals by a ratio $\bar{R}(f) < 1$.

2.6 Loss of Diversity

During every selection phase, bad individuals are lost and replaced by copies of better individuals. Thereby, a certain amount of “genetic material” is lost that was contained in the bad individuals. The number of individuals that are replaced corresponds to the strength of the “loss of diversity.” This leads to the following definition.

DEFINITION 2.8 (Loss of Diversity): *The loss of diversity D is the proportion of individuals of a population that are not selected during the selection phase.*

THEOREM 2.2: *If the reproduction rate $\bar{R}(f)$ increases strictly monotonically in f , the loss of diversity of a selection method is*

$$D = \frac{1}{N} (\bar{S}(f_z) - \bar{S}^*(f_z)) \quad (11)$$

where f_z denotes the fitness value such that $\bar{R}(f_z) = 1$.

PROOF: Since $\bar{R}(f)$ increases strictly monotonically in f , f_z is uniquely defined. For all fitness values $f \in (f_0, f_z]$, the reproduction rate is less than 1. Hence, the number of individuals that are not selected during selection is given by $\int_{f_0}^{f_z} (\bar{s}(x) - \bar{s}^*(x)) dx$. It follows that

$$\begin{aligned} D &= \frac{1}{N} \int_{f_0}^{f_z} (\bar{s}(x) - \bar{s}^*(x)) dx \\ &= \frac{1}{N} \left(\int_{f_0}^{f_z} \bar{s}(x) dx - \int_{f_0}^{f_z} \bar{s}^*(x) dx \right) \\ &= \frac{1}{N} (\bar{S}(f_z) - \bar{S}^*(f_z)) \quad \square \end{aligned}$$

A small loss of diversity should reduce the risk of premature convergence as more unique genetic material is preserved for the next generation.

In his dissertation, Baker (1989) introduced a similar measure, the “reproduction rate RR,” which gives the percentage of individuals that are selected to reproduce; hence, $RR = 100(1 - D)$. Baker used this measure as a dynamic convergence measure. By observing the RR during the run of a genetic algorithm, he tried to extract the state of convergence of the population. Note that in this paper, “reproduction rate” is used in the sense of Definition 2.7.

2.7 Selection Intensity

The term “selection intensity” or “selection pressure” is often used in different contexts and for different properties of a selection method. Goldberg and Deb (1991) and Bäck (1994b) use the “takeover time” to define the selection pressure. Whitley calls the parameter ϵ (see Section 5) of his ranking selection method “selection pressure.”

In this paper, the term “selection intensity” is used in the same way as in population genetics (Bulmer, 1980). Mühlenbein and Schlierkamp-Voosen (1993) have adopted the definition and applied it to genetic algorithms. Recently, more and more researchers are using this term to characterize selection schemes (Thierens & Goldberg 1994a,b; Bäck, 1995; Miller & Goldberg, 1995).

The change in the average fitness of the population due to selection is a reasonable measure for selection intensity. In population genetics the term *selection intensity* was introduced to obtain a normalized and dimensionless measure. The idea is to measure the progress due to selection by the so-called “selection differential,” that is, the difference between the average population fitness after and before selection. Dividing this selection differential by the mean variance of the population fitness leads to the desired dimensionless measure called *selection intensity*.

DEFINITION 2.9 (Selection Intensity): *The selection intensity of a selection method Ω for the fitness distribution $\bar{s}(f)$ is the quantity*

$$I = \frac{\bar{M}^* - \bar{M}}{\bar{\sigma}} \quad (12)$$

By Equation 12, the selection intensity depends on the fitness distribution of the initial population. Hence, different fitness distributions will in general lead to different selection intensities for the same selection method. For comparison, it is necessary to restrict oneself to a certain initial distribution. Using the normalized Gaussian distribution $\mathcal{G}(0, 1)$ as the initial fitness distribution leads to the following definition.

DEFINITION 2.10 (Standardized Selection Intensity): *The standardized selection intensity I_Ω is the expected average fitness value of the population after applying the selection method Ω to the normalized Gaussian distribution $\mathcal{G}(0, 1)(f) = (\sqrt{2\pi})^{-1} \exp(-f^2/2)$:*

$$I_\Omega = \int_{-\infty}^{\infty} f \bar{\Omega}^*(\mathcal{G}(0, 1))(f) df \quad (13)$$

The “effective” average fitness value of a Gaussian distribution with mean μ and variance σ^2 can easily be derived as $\bar{M}^* = \sigma I_\Omega + \mu$. Note that this definition of the standardized selection intensity can only be applied if the selection method is scale and translation invariant. This is the case for all selection schemes examined in this paper. Likewise, this definition has no equivalent in the case of discrete fitness distributions. If the selection intensity for a discrete distribution has to be calculated, one must refer to Definition 2.9. In the remainder

of this paper the term “selection intensity” is used as equivalent to “standardized selection intensity,” since the intention is to compare selection schemes.

2.8 Selection Variance

In addition to selection intensity the term “selection variance” is introduced. The definition is analogous to the definition of the selection intensity but describes the new variance of the fitness distribution after selection.

DEFINITION 2.11 (Selection Variance): *The selection variance Φ is the normalized expected variance of the fitness distribution of the population after applying the selection method Ω to the fitness distribution $\bar{s}(f)$, that is,*

$$\Phi = \frac{(\bar{\sigma}^*)^2}{\bar{\sigma}^2} \quad (14)$$

For comparison, the standardized selection variance is used.

DEFINITION 2.12 (Standardized Selection Variance): *The standardized selection variance Φ_Ω is the normalized expected variance of the fitness distribution of the population after applying the selection method Ω to the normalized Gaussian distribution $\mathcal{G}(0, 1)$:*

$$\Phi_\Omega = \int_{-\infty}^{\infty} (f - I_\Omega)^2 \bar{\Omega}^*(\mathcal{G}(0, 1))(f) df \quad (15)$$

which is equivalent to

$$\Phi_\Omega = \int_{-\infty}^{\infty} f^2 \bar{\Omega}^*(\mathcal{G}(0, 1))(f) df - I_\Omega^2 \quad (16)$$

Note that there is a difference between the selection variance and the loss of diversity. The loss of diversity gives the proportion of individuals that are not selected, regardless of their fitness value. The standardized selection variance is defined as the new variance of the fitness distribution assuming a Gaussian initial fitness distribution. Hence, a selection variance of 1 means that the variance is not changed by selection. A selection variance less than 1 corresponds to a decrease in variance. The lowest possible value of Φ_Ω is 0, which means that the variance of the fitness values of the population after selection is itself 0. Again the term “selection variance” is used as equivalent to “standardized selection variance.”

In the following sections four popular selection schemes are examined in detail: tournament, truncation, and linear and exponential ranking. They all have in common the fact that they are scale and translation invariant with respect to fitness values. This means that all properties defined above can be derived without any assumptions about fitness values. However, fitness proportional selection cannot be examined with this methodology because it is not translation invariant (De la Maza & Tidor, 1993). Furthermore, some properties discussed here, including the loss of diversity, are difficult to investigate for fitness proportional selection. The crucial point is the explicit occurrence of the fitness value in the expected fitness distribution after selection. Hence, an analysis is only possible if some further assumptions on the initial fitness distribution are made.

3. Tournament Selection

In tournament selection a group of t individuals are randomly chosen from the population. They may be drawn from the population with or without replacement. This group takes

part in a “tournament”; that is, a winning individual is determined depending on its fitness value. The best individual having the highest fitness value is usually chosen deterministically, although stochastic selection may occasionally be used. In both cases only the winner is inserted into the next population, and the process is repeated N times to obtain a new population. Often, tournaments are held between two individuals (binary tournament). However, this can be generalized to an arbitrary group of size t called the *tournament size*.

The following description assumes that the individuals are drawn with replacement and that the winning individual is deterministically selected (Algorithm 1). The outline of the algorithm shows that tournament selection can be implemented very efficiently because no sorting of the population is required. Implemented in the above way, it has the time complexity $\mathcal{O}(N)$.

Using the notation introduced in the previous section, the entire fitness distribution after selection can be predicted. The prediction is made for the discrete (exact) fitness distribution as well as for a continuous fitness distribution. The results presented here are published in Blickle and Thiele (1995), where the proofs of the cited theorems can also be found.

ALGORITHM 1 (Tournament Selection):

Input: The population $\vec{f} = \{f_1, \dots, f_N\}$,
 the fitness values of the population $\vec{\phi} = \{\phi_1, \dots, \phi_N\}$.
 the tournament size $t \in \{1, 2, \dots, N\}$
Output: The population after selection $\vec{f}' = \{f'_1, \dots, f'_N\}$

tournament($t, \vec{f}, \vec{\phi}$):
 for $i \leftarrow 1$ **to** N **do**
 $f'_i \leftarrow$ best-fit individual of t uniform randomly
 picked individuals from \vec{f}
 od
 return \vec{f}'

THEOREM 3.1: *The expected fitness distribution after performing tournament selection with tournament size t on the distribution s is*

$$\Omega_T^*(s, t)(f_i) = s^*(f_i) = N \left(\left(\frac{S(f_i)}{N} \right)^t - \left(\frac{S(f_{i-1})}{N} \right)^t \right) \quad (17)$$

Equation (17) shows the strong influence of the tournament size t on the behavior of the selection scheme. Obviously, for $t = 1$ (on average) the unchanged initial distribution is obtained as

$$\Omega_T^*(s, 1)(f_i) = N \left(\frac{S(f_i)}{N} - \frac{S(f_{i-1})}{N} \right) = S(f_i) - S(f_{i-1}) = s(f_i)$$

EXAMPLE 3.1: *Using the discrete fitness distribution from Example 2.1 (Fig. 1), the fitness distribution shown in Figure 2 is obtained after applying tournament selection with a tournament size $t = 3$.*

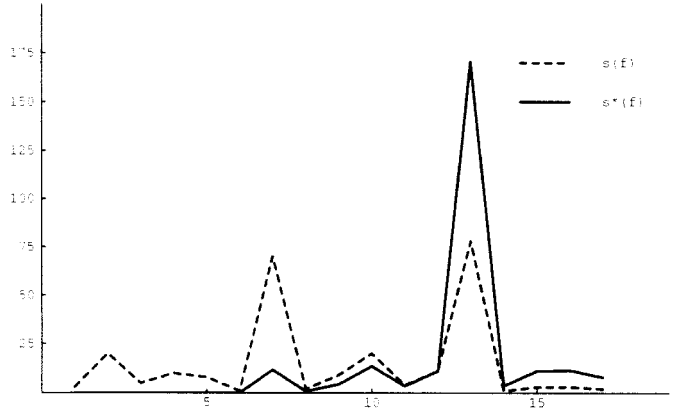


Figure 2. The initial fitness distribution (dashed line) and the resulting expected fitness distribution (solid line) after applying tournament selection with a tournament size of 3.

THEOREM 3.2: *Let $\bar{s}(f)$ be the continuous fitness distribution of the population. Then the expected fitness distribution after performing tournament selection with tournament size t is*

$$\bar{\Omega}_T^*(\bar{s}, t)(f) = \bar{s}^*(f) = t\bar{s}(f) \left(\frac{\bar{S}(f)}{N} \right)^{t-1} \quad (18)$$

3.1 Concatenation of Tournament Selection

An interesting property of tournament selection is the concatenation of several selection phases. Assuming an arbitrary population with the fitness distribution \bar{s} , a tournament selection with tournament size t_1 is applied to this population and then on the resulting population tournament selection with tournament size t_2 . The obtained fitness distribution is the same as if only one tournament selection with tournament size $t_1 \cdot t_2$ is applied to the initial distribution \bar{s} .

THEOREM 3.3: *Let \bar{s} be a continuous fitness distribution and $t_1, t_2 \geq 1$ two tournament sizes. Then, the following equation holds:*

$$\bar{\Omega}_T^*(\bar{\Omega}_T^*(\bar{s}, t_1), t_2)(f) = \bar{\Omega}_T^*(\bar{s}, t_1 \cdot t_2)(f) \quad (19)$$

In Goldberg and Deb (1991), the proportion P_k of best-fit individuals after k selections with tournament size t (without recombination) is given as

$$P_k = 1 - (1 - P_0)^{t^k} \quad (20)$$

This can be obtained as a special case from Theorem 3.3, if only best-fit individuals are considered.

3.2 Reproduction Rate

COROLLARY 3.1: *The reproduction rate of tournament selection is*

$$\bar{R}_T(f) = \frac{\bar{s}^*(f)}{\bar{s}(f)} = t \left(\frac{\bar{S}(f)}{N} \right)^{t-1} \quad (21)$$

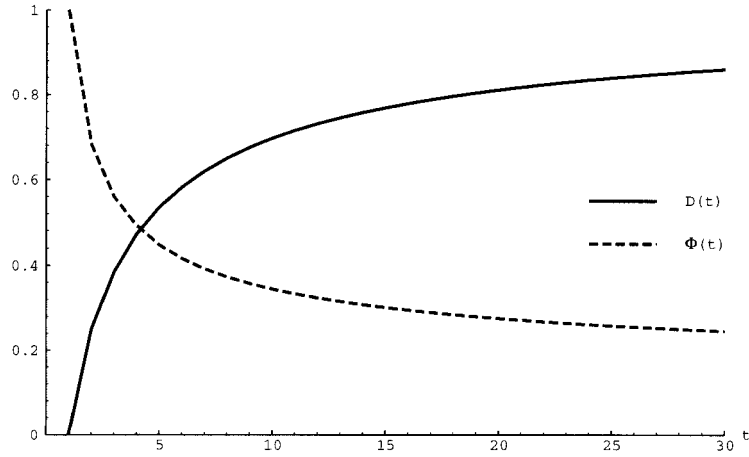


Figure 3. Loss of diversity $D_T(t)$ (solid line) and selection variance $\Phi_T(t)$ (dashed line) of tournament selection.

This is directly obtained by substituting Equation 18 in Equation 10.

Individuals with the lowest fitness have a reproduction rate of almost zero, and the individuals with the highest fitness have a reproduction rate of t .

3.3 Loss of Diversity

THEOREM 3.4: *The loss of diversity D_T of tournament selection is*

$$D_T(t) = t^{-1/(t-1)} - t^{-t/(t-1)} \quad (22)$$

It turns out that the number of individuals lost increases with the tournament size (see Fig. 3). About half of the population is lost at tournament size $t = 5$. Note that the loss of diversity is independent of the initial fitness distribution.

3.4 Selection Intensity

To calculate the selection intensity, the average fitness of the population after applying tournament selection on the normalized Gaussian distribution $\mathcal{G}(0, 1)$ is needed. Using Definition 2.5 leads to

$$I_T(t) = \int_{-\infty}^{\infty} tx \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \left(\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy\right)^{t-1} dx \quad (23)$$

These integral equations can be solved analytically for the cases $t = 1, \dots, 5$ (Blickle & Thiele, 1995; Bäck, 1995; Arnold, Balakrishnan, & Nagaraja, 1992):

$$\begin{aligned} I_T(1) &= 0; & I_T(2) &= \frac{1}{\sqrt{\pi}}; & I_T(3) &= \frac{3}{2\sqrt{\pi}}; & I_T(4) &= \frac{6}{\pi\sqrt{\pi}} \arctan \sqrt{2}; \\ I_T(5) &= \frac{10}{\sqrt{\pi}} \left(\frac{3}{2\pi} \arctan \sqrt{2} - \frac{1}{4} \right) \end{aligned}$$

For a tournament size of $t = 2$, Thierens and Goldberg (1994a) derive the same average fitness value in a completely different manner. However, their formulation cannot be extended to other tournament sizes.

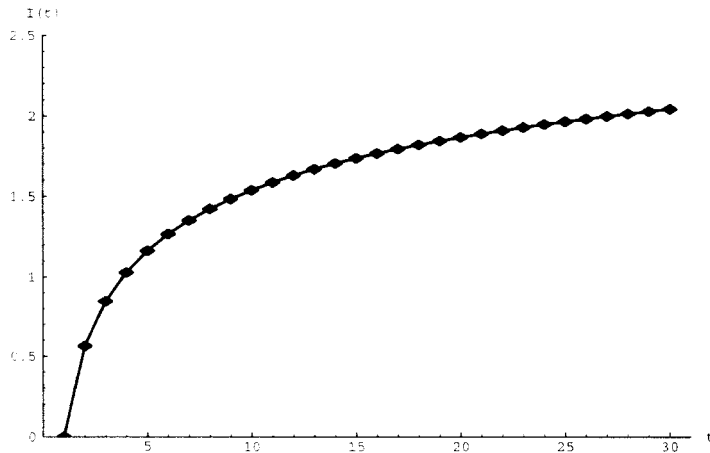


Figure 4. Dependence of the selection intensity $I_T(t)$ on tournament size t .

For larger tournament sizes, Equation 23 can be accurately evaluated by numerical integration. The result is shown in Figure 4 for a tournament size from 1 to 30. Equation 24 approximates Equation 23 with a relative error of less than 2.4% for $t \in [2, 5]$; for tournament sizes $t > 5$, the relative error is less than 1% (Blickle & Thiele, 1995):

$$I_T(t) \approx \sqrt{2(\ln(t) - \ln(\sqrt{4.14 \ln(t)}))} \quad (24)$$

3.5 Selection Variance

To determine the selection variance, the following equation needs to be solved:

$$\Phi_T(t) = \int_{-\infty}^{\infty} t(x - I_T(t))^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \left(\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy\right)^{t-1} dx \quad (25)$$

For a binary tournament, one calculates

$$\Phi_T(2) = 1 - \frac{1}{\pi}$$

Here again, Equation (25) can be solved by numerical integration. The dependence of the selection variance on the tournament size is shown in Figure 3.

To obtain a useful analytic approximation for the selection variance, symbolic regression using the genetic programming optimization method is used. Details on how the data were computed can be found in Blickle (1996). The following expression approximates the selection variance with a relative error of less than 1.6% for $t \in \{1, \dots, 30\}$:

$$\Phi_T(t) \approx \frac{0.918}{\ln(1.186 + 1.328t)}; \quad t \in \{1, \dots, 30\} \quad (26)$$

4. Truncation Selection

In truncation selection with threshold T , only the fraction T best individuals are selected, and they all have the same selection probability. This selection method is often used by

breeders and in population genetics (Bulmer, 1980; Crow & Kimura, 1970). Mühlenbein and Schlierkamp-Voosen (1993) introduced this selection scheme to the domain of genetic algorithms. This method is equivalent to (μ, λ) -selection used in evolution strategies with $T = \mu/\lambda$ (Bäck, 1995).

The outline of the algorithm is given by Algorithm 2.

ALGORITHM 2 (Truncation Selection):

Input: The population $\vec{f} = \{f_1, \dots, f_N\}$,
 the fitness values of the population $\vec{\phi} = (\phi_1, \dots, \phi_N)$,
 the truncation threshold $T \in [0, 1]$

Output: The population after selection $\vec{f}' = \{f'_1, \dots, f'_N\}$

truncation($T, \vec{f}, \vec{\phi}$):

$\vec{f}^* \leftarrow$ sorted population \vec{f} according fitness $\vec{\phi}$
 with least fit individual at the first position

for $i \leftarrow 1$ **to** N **do**

$r \leftarrow$ uniform_random $\{[(1 - T)N], \dots, N\}$

$f'_i \leftarrow f_r^*$

od

return \vec{f}'

Because a sorting of the population is required, truncation selection has a time complexity of $\mathcal{O}(N \log N)$.

Although this selection method has been investigated several times, it will be described using the methods derived here because additional important properties are derived.

THEOREM 4.1: *The expected fitness distribution after performing truncation selection with threshold T on the distribution s is*

$$\Omega_{\Gamma}^*(s, T)(f_i) = \begin{cases} 0 & : S(f_i) \leq (1 - T)N \\ \frac{S(f_i) - (1 - T)N}{T} & : S(f_{i-1}) \leq (1 - T)N < S(f_i) \\ \frac{s(f_i)}{T} & : \text{else} \end{cases} \quad (27)$$

PROOF: The first case in Equation 27 gives zero offspring to individuals with a fitness value below the truncation threshold. The second case reflects the fact that the threshold may lie within $s(f_i)$. Then, only the fraction above the threshold $(S(f_i) - (1 - T)N)$ is selected. This fraction is on average copied $1/T$ times. The last case in Equation 27 gives all individuals above the threshold the multiplication factor $1/T$, necessary to keep the population size constant. \square

EXAMPLE 4.1: *Using the discrete fitness distribution from Example 2.1, the fitness distribution shown in Figure 5 is obtained after applying truncation selection with threshold $T = 0.47$.*

THEOREM 4.2: *Let $\tilde{s}(f)$ be the continuous distribution of the population. Then, the expected fitness distribution after performing truncation selection with threshold T is*

$$\overline{\Omega}_{\Gamma}^*(\tilde{s}, T)(f) = \begin{cases} \frac{\tilde{s}(f)}{T} & : \tilde{S}(f) > (1 - T)N \\ 0 & : \text{else} \end{cases} \quad (28)$$

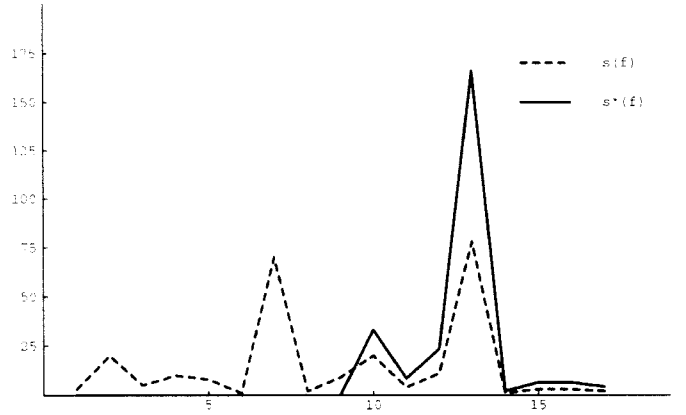


Figure 5. The initial fitness distribution (dashed line) and the resulting expected fitness distribution (solid line) after applying truncation selection with truncation threshold $T = 0.47$.

PROOF: Since $\tilde{S}(f)$ gives the cumulative fitness distribution, it follows from the construction of truncation selection that all individuals with $\tilde{S}(f) < (1 - T)N$ are truncated. Since the population size is kept constant during selection, all other individuals must be copied on average $1/T$ times. \square

4.1 Reproduction Rate

From the construction of the selection method, the reproduction rate can easily be derived.

COROLLARY 4.1: *The reproduction rate of truncation selection is*

$$\bar{R}_T(f) = \begin{cases} \frac{1}{T} & : \tilde{S}(f) > (1 - T)N \\ 0 & : \text{else} \end{cases} \quad (29)$$

This is directly obtained by substituting Equation 28 in Equation 10.

4.2 Loss of Diversity

By construction of the selection method, only the fraction T of the population will be selected; that is, the loss of diversity is

$$D_T(T) = 1 - T \quad (30)$$

The loss of diversity is depicted in Figure 6.

4.3 Selection Intensity

The results presented in this subsection were also derived by Crow and Kimura (1970).

THEOREM 4.3: *The selection intensity of truncation selection is*

$$I_T(T) = \frac{1}{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f_c^2}{2}\right) \quad (31)$$

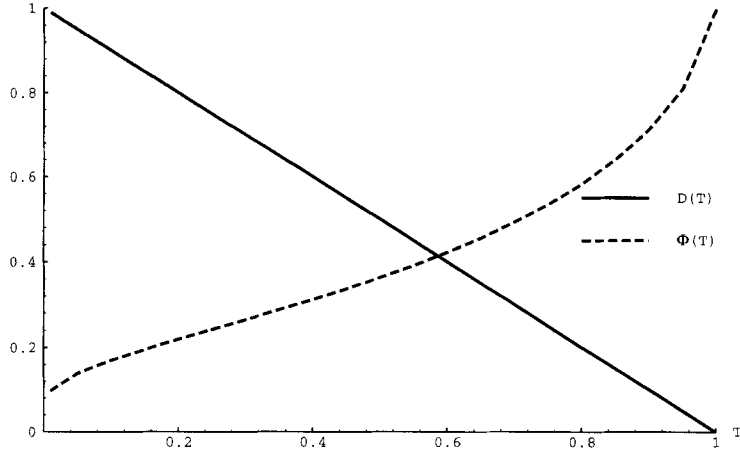


Figure 6. Loss of diversity $D_T(t)$ (solid line) and selection variance $\Phi_T(t)$ (dashed line) of truncation selection.

where f_c is determined by

$$T = \int_{f_c}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f^2}{2}\right) df \quad (32)$$

PROOF: The selection intensity is defined as the average fitness of the population after selection, assuming an initial normalized Gaussian distribution $\mathcal{G}(0, 1)$; hence $I = \int_{-\infty}^{\infty} \tilde{\Omega}(\mathcal{G}(0, 1))(f)f df$. Because no individual with a fitness value worse than f_c will be selected, the lower integration bound can be replaced by f_c . Here, f_c is determined by

$$\tilde{S}(f_c) = (1 - T)N = 1 - T \quad (33)$$

because $N = 1$ for the normalized Gaussian distribution.

Hence, I_T can be computed as

$$\begin{aligned} I_T(T) &= \int_{f_c}^{\infty} \frac{1}{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f^2}{2}\right) f df \\ &= \frac{1}{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f_c^2}{2}\right) \end{aligned}$$

Here, f_c is determined by Equation 33. Solving Equation 33 for T yields

$$\begin{aligned} T &= 1 - \int_{-\infty}^{f_c} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f^2}{2}\right) df \\ &= \int_{f_c}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f^2}{2}\right) df \end{aligned} \quad \square$$

A lower bound for the selection intensity reported by Nagaraja (1982) is $I_T(T) \leq \sqrt{(1 - T)/T}$.

Figure 7 shows the selection intensity in relation to the dependence of parameter T .

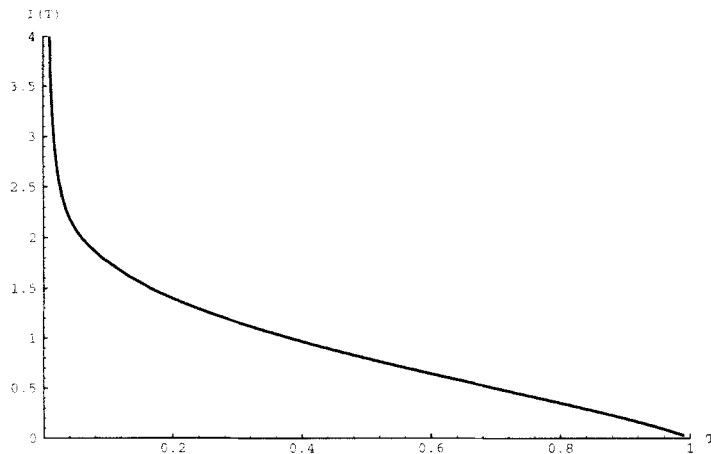


Figure 7. Selection intensity $I_T(T)$ of truncation selection.

4.4 Selection Variance

THEOREM 4.4: *The selection variance of truncation selection is*

$$\Phi_T(T) = 1 - I_T(T)(I_T(T) - f_c) \tag{34}$$

PROOF: The substitution of Equation 28 in the defining Equation 16 gives

$$\Phi_T(T) = \int_{f_c}^{\infty} f^2 \frac{1}{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f^2}{2}\right) df - (I_T(T))^2$$

A partial integration with $u(f) = f$ and $\partial v / \partial f = f \exp(-f^2/2)$ yields

$$\Phi_T(T) = \frac{f_c}{T\sqrt{2\pi}} \exp\left(-\frac{f_c^2}{2}\right) + \frac{1}{T\sqrt{2\pi}} \int_{f_c}^{\infty} \exp\left(-\frac{f^2}{2}\right) df - (I_T(T))^2$$

Substituting Equations 31 and 32 simplifies this equation to Equation 34. □

The selection variance is plotted in Figure 6. Equation 34 has also been derived in Bulmer (1980).

5. Linear Ranking Selection

Linear ranking selection was first suggested in order to eliminate the serious disadvantages of fitness proportional selection (Grefenstette & Baker, 1989; Whitley, 1989). For linear ranking selection, the individuals are sorted according to their fitness values, and the rank N is assigned to the best individual and the rank 1 to the worst. The selection probability is linearly assigned to the individuals according to their rank:

$$p_i = \frac{1}{N} \left(\eta^- + (\eta^+ - \eta^-) \frac{i-1}{N-1} \right); \quad i \in \{1, \dots, N\} \tag{35}$$

Here, η^-/N is the probability of the worst individual to be selected and η^+/N the probability of the best individual to be selected. Because the population size is held constant, the

conditions $\eta^+ = 2 - \eta^-$ and $\eta^- \geq 0$ must be fulfilled. Note that all individuals get a different rank, that is, a different selection probability, even if they have the same fitness value.

Koza (1992) determines the probability by a multiplication factor r_m that determines the gradient of the linear function. A transformation into the form of Equation 35 is possible by $\eta^- = 2/(r_m + 1)$ and $\eta^+ = 2r_m/(r_m + 1)$.

ALGORITHM 3 (Linear Ranking Selection):

Input: The population $\vec{f} = \{f_1, \dots, f_N\}$,
 the fitness values of the population $\vec{\phi} = \{\phi_1, \dots, \phi_N\}$,
 the reproduction rate of the worst individual $\eta^- \in [0, 1]$
Output: The population after selection $\vec{f}' = \{f'_1, \dots, f'_N\}$

linear_ranking(η^- , \vec{f} , $\vec{\phi}$):

$\vec{f}^* \leftarrow$ sorted population \vec{f} according fitness $\vec{\phi}$
 with least fit individual at the first position

sum₀ \leftarrow 0

for $i \leftarrow 1$ **to** N **do**

 sum _{i} \leftarrow sum _{$i-1$} + p_i (Equation 35)

od

for $i \leftarrow 1$ **to** N **do**

$r \leftarrow$ uniform_random[0, 1[

$f'_i \leftarrow f_i^*$ such that sum _{$i-1$} $\leq r <$ sum _{i}

od

return \vec{f}'

Whitley (1989) describes the ranking selection by transforming an equally distributed random variable $\chi \in [0, 1]$ to determine the index of the selected individual

$$j = \left\lfloor \frac{N}{2(c-1)} \left(c - \sqrt{c^2 - 4(c-1)\chi} \right) \right\rfloor \quad (36)$$

where c is a parameter called the “selection bias.” Bäck (1994b) showed that for $1 < c \leq 2$, this method is almost identical to the probabilities in Equation 35, with $\eta^+ = c$.

The pseudocode implementation of linear ranking selection is given by Algorithm 3. The method requires the sorting of the population; hence, the complexity of the algorithm is dominated by the complexity of sorting, that is, $\mathcal{O}(N \log N)$.

THEOREM 5.1: *The expected fitness distribution after performing ranking selection with η^- on the distribution s is*

$$\Omega_R^*(s, \eta^-)(f_i) = s^*(f_i) = s(f_i) \frac{N\eta^- - 1}{N - 1} + \frac{1 - \eta^-}{N - 1} \left(S(f_i)^2 - S(f_{i-1})^2 \right) \quad (37)$$

PROOF: First, the expected number of individuals with fitness f_i or worse is calculated, that is, $S^*(f_i)$. Because the individuals are sorted according to their fitness value, this number is

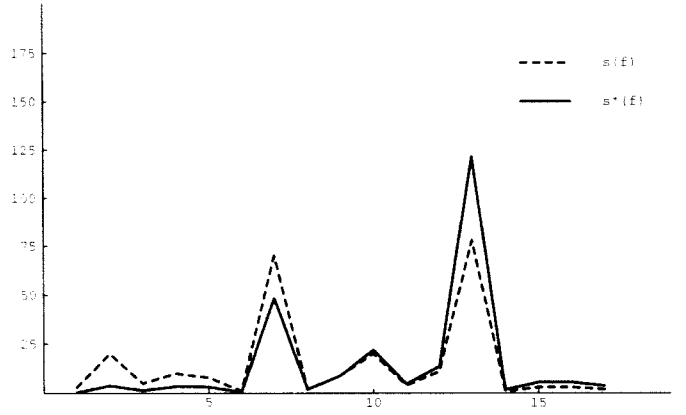


Figure 8. The initial fitness distribution (dashed line) and the resulting expected fitness distribution (solid line) after applying linear ranking selection with $\eta^- = 0.1$.

given by the sum of the probabilities of the $S^*(f_i)$ least fit individuals:

$$\begin{aligned}
 S^*(f_i) &= N \sum_{j=1}^{S(f_i)} p_j \\
 &= \eta^- S(f_i) + \frac{\eta^* - \eta^-}{N-1} \sum_{j=1}^{S(f_i)} j - 1 \\
 &= \eta^- S(f_i) + \frac{\eta^* - \eta^-}{N-1} \frac{1}{2} S(f_i) (S(f_i) - 1)
 \end{aligned}$$

With $\eta^* = 2 - \eta^-$ and $s^*(f_i) = S^*(f_i) - S^*(f_{i-1})$ it follows that

$$\begin{aligned}
 s^*(f_i) &= \eta^- (S(f_i) - S(f_{i-1})) \\
 &\quad + \frac{1 - \eta^-}{N-1} (S(f_i)(S(f_i) - 1) - S(f_{i-1})(S(f_{i-1}) - 1)) \\
 &= \eta^- s(f_i) + \frac{1 - \eta^-}{N-1} (S(f_i)^2 - S(f_{i-1})^2 - s(f_i)) \\
 &= s(f_i) \frac{N\eta^- - 1}{N-1} + \frac{1 - \eta^-}{N-1} (S(f_i)^2 - S(f_{i-1})^2) \quad \square
 \end{aligned}$$

EXAMPLE 5.1: Using the discrete fitness distribution from Example 2.1, the fitness distribution shown in Figure 8 is obtained after applying linear ranking selection with $\eta^- = 0.1$.

THEOREM 5.2: Let $\tilde{s}(f)$ be the continuous fitness distribution of the population. Then the expected fitness distribution after performing ranking selection $\bar{\Omega}_R$ with η^- on the distribution \tilde{s} is

$$\bar{\Omega}_R^*(\tilde{s}, \eta^-)(f) = \tilde{s}^*(f) = \eta^- \tilde{s}(f) + 2 \frac{1 - \eta^-}{N} \tilde{S}(f) \tilde{s}(f) \quad (38)$$

PROOF: Since the continuous form of Equation 35 is given by

$$\bar{p}(x) = \frac{1}{N} \left(\eta^- + \frac{\eta^+ - \eta^-}{N} x \right)$$

one calculates $\bar{S}(f)$ using $\eta^+ = 2 - \eta^-$:

$$\begin{aligned} \bar{S}^*(f) &= N \int_0^{\bar{S}(f)} \bar{p}(x) dx \\ &= \eta^- \int_0^{\bar{S}(f)} dx + 2 \frac{1 - \eta^-}{N} \int_0^{\bar{S}(f)} x dx \\ &= \eta^- \bar{S}(f) + \frac{1 - \eta^-}{N} \bar{S}(f)^2 \end{aligned}$$

Since $\bar{s}^*(f) = d\bar{S}^*(f)/df$, Equation 38 follows. □

5.1 Reproduction Rate

COROLLARY 5.1: *The reproduction rate of ranking selection is*

$$\bar{R}_R(f) = \eta^- + 2 \frac{1 - \eta^-}{N} \bar{S}(f) \quad (39)$$

This is directly obtained by substituting Equation 38 in Equation 10.

Equation 39 shows that the worst individuals have the lowest reproduction rate $\bar{R}(f_0) = \eta^-$, and the best individuals have the highest reproduction rate $\bar{R}(f_n) = 2 - \eta^- = \eta^+$. This can be derived from the construction of the method, since η^-/N is the selection probability of the worst individual and η^+/N that of the best individual.

5.2 Loss of Diversity

THEOREM 5.3: *The loss of diversity $D_R(\eta^-)$ of ranking selection is*

$$D_R(\eta^-) = (1 - \eta^-) \frac{1}{4} \quad (40)$$

PROOF: Using Theorem 2.2 and realizing that $S(f_z) = N/2$, it follows that

$$\begin{aligned} D_R(\eta^-) &= \frac{1}{N} (\bar{S}(f_z) - \bar{S}^*(f_z)) \\ &= \frac{1}{N} \left(\bar{S}(f_z) - \eta^- \bar{S}(f_z) - \frac{1 - \eta^-}{N} \bar{S}(f_z)^2 \right) \\ &= \frac{1}{N} \left(\frac{N}{2} - \eta^- \frac{N}{2} - \frac{1 - \eta^-}{N} \frac{N^2}{4} \right) \\ &= \frac{1}{4} (1 - \eta^-) \end{aligned} \quad \square$$

Baker (1989) has derived this result using his term of “reproduction rate.”

Note that the loss of diversity is again independent of the initial distribution. The loss of diversity is depicted in Figure 9.

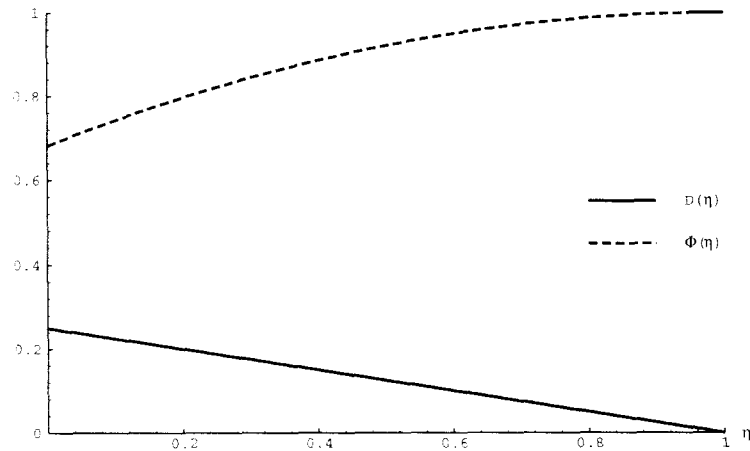


Figure 9. The loss of diversity $D_R(\eta)$ (solid line) and the selection variance $\Phi_R(\eta)$ (dashed line) of linear ranking selection.

5.3 Selection Intensity

THEOREM 5.4: *The selection intensity of ranking selection is*

$$I_R(\eta^-) = (1 - \eta^-) \frac{1}{\sqrt{\pi}} \tag{41}$$

PROOF: Using the definition of the selection intensity (Definition 2.10) and the Gaussian function for the initial fitness distribution, one obtains

$$\begin{aligned} I_R(\eta^-) &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \left(\eta^- + 2(1 - \eta^-) \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \right) dx \\ &= \frac{\eta^-}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{x^2}{2}\right) dx \\ &\quad + \frac{1 - \eta^-}{\pi} \int_{-\infty}^{\infty} x \exp\left(-\frac{x^2}{2}\right) \int_{-\infty}^x \exp\left(-\frac{y^2}{2}\right) dy dx \end{aligned}$$

Using the relations

$$\int_{-\infty}^{\infty} x \exp\left(-\frac{x^2}{2}\right) dx = 0$$

and

$$\int_{-\infty}^{\infty} x \exp\left(-\frac{x^2}{2}\right) \left(\int_{-\infty}^x \exp\left(-\frac{y^2}{2}\right) dy \right)^2 dx = \sqrt{2\pi}$$

Equation 41 follows. □

The selection intensity of ranking selection is shown in Figure 10 in dependence of the parameter η^- .

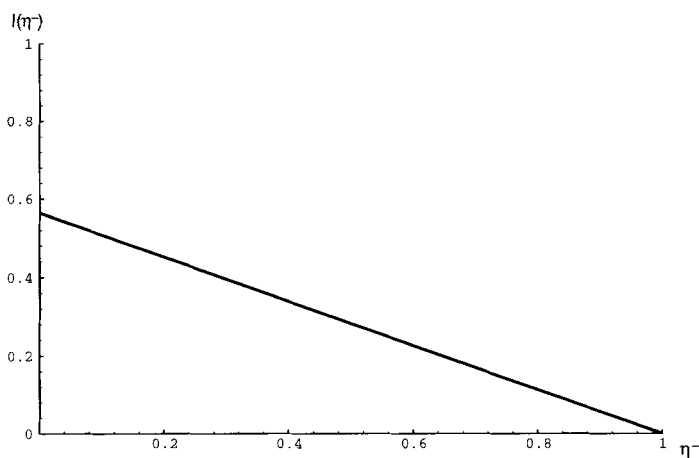


Figure 10. The selection intensity $I_R(\eta^-)$ of ranking selection.

5.4 Selection Variance

THEOREM 5.5: *The selection variance of linear ranking is*

$$\Phi_R(\eta^-) \approx 1 - \frac{(1 - \eta^-)^2}{\pi} = 1 - I_R(\eta^-)^2 \quad (42)$$

PROOF: Substituting Equation 38 in the definition equation (Equation 16) leads to

$$\begin{aligned} \Phi_R(\eta^-) &= \int_{-\infty}^{\infty} f^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f^2}{2}\right) \left(\eta^- + 2(1 - \eta^-) \int_{-\infty}^f \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \right) df \\ &\quad - I_R(\eta^-)^2 \\ &= \frac{\eta^-}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f^2 \exp\left(-\frac{f^2}{2}\right) df \\ &\quad + \frac{1 - \eta^-}{\pi} \int_{-\infty}^{\infty} f^2 \exp\left(-\frac{f^2}{2}\right) \int_{-\infty}^f \exp\left(-\frac{y^2}{2}\right) dy df - I_R(\eta^-)^2 \end{aligned}$$

Using the relations

$$\int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{2\pi}$$

and

$$\int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right) \int_{-\infty}^x \exp\left(-\frac{y^2}{2}\right) dy dx = \pi$$

leads to

$$\Phi_R(\eta^-) \approx \eta^- + (1 - \eta^-) - I_R(\eta^-)^2 = 1 - I_R(\eta^-)^2 \quad \square$$

The selection variance of ranking selection is plotted in Figure 9.

6. Exponential Ranking Selection

Exponential ranking selection differs from linear ranking selection in that the probabilities of the ranked individuals are exponentially weighted. The base of the exponent is the parameter $0 < c < 1$ of the method. The closer c is to 1 the lower is the “exponentiality” of the selection method. The meaning and the influence of this parameter will be discussed in detail. Again, the rank N is assigned to the best individual and the rank 1 to the worst individual. Hence, the probabilities of the individuals are given by

$$p_i = \frac{c^{N-i}}{\sum_{j=1}^N c^{N-j}}; \quad i \in \{1, \dots, N\} \quad (43)$$

The sum $\sum_{j=1}^N c^{N-j}$ normalizes the probabilities to ensure $\sum_{i=1}^N p_i = 1$. Since $\sum_{j=1}^N c^{N-j} = (c^N - 1)/(c - 1)$, Equation 43 can be rewritten:

$$p_i = \frac{c - 1}{c^N - 1} c^{N-i}; \quad i \in \{1, \dots, N\} \quad (44)$$

The algorithm for exponential ranking (Algorithm 4) is similar to that for linear ranking. The only difference lies in the calculation of the selection probabilities.

THEOREM 6.1: *The expected fitness distribution after performing exponential ranking selection with parameter c on the distribution s is*

$$\Omega_F^*(s, c, N)(f_i) = s^*(f_i) = N \frac{c^N}{c^N - 1} c^{-S(f_i)} (c^{S(f_i)} - 1) \quad (45)$$

ALGORITHM 4 (Exponential Ranking Selection):

Input: The population $\vec{f} = \{f_1, \dots, f_N\}$,
the fitness values of the population $\vec{o} = \{o_1, \dots, o_N\}$,
the ranking base $c \in]0, 1]$
Output: The population after selection $\vec{f}' = \{f'_1, \dots, f'_N\}$

exponential_ranking(c, \vec{f}, \vec{o}):

$\vec{f}^* \leftarrow$ sorted population \vec{f} according to fitness
with least fit individual at the first position

sum₀ \leftarrow 0

for $i \leftarrow 1$ **to** N **do**

sum _{i} \leftarrow sum _{$i-1$} + p_i (Equation 44)

od

for $i \leftarrow 1$ **to** N **do**

$r \leftarrow$ uniform_random[0,1[

$f'_i \leftarrow f_i^*$ such that sum _{$i-1$} $\leq r <$ sum _{i}

od

return \vec{f}'

PROOF: First, the expected number of individuals with fitness f_i or worse is calculated; that is, $S^*(f_i)$. Since the individuals are sorted according to their fitness value, this number is

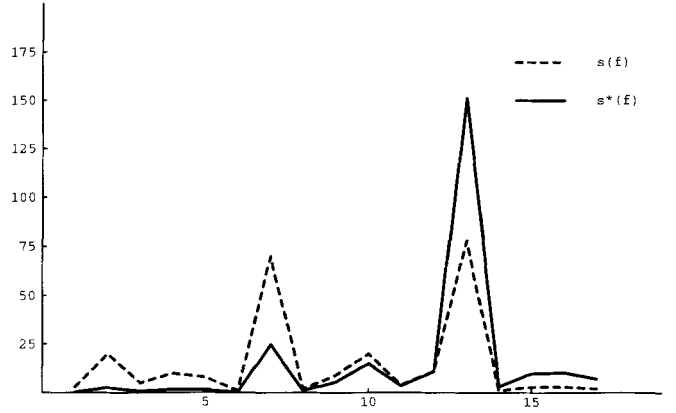


Figure 11. The initial fitness distribution (dashed line) and the resulting expected fitness distribution (solid line) after applying exponential ranking selection with $c = 0.986$ ($N = 250$).

given by the sum of the probabilities of the $S^*(f_i)$ least fit individuals:

$$S^*(f_i) = N \sum_{j=1}^{S(f_i)} p_j = N \frac{c-1}{c^N-1} \sum_{j=1}^{S(f_i)} c^{N-j}$$

and with the substitution $k = N - j$,

$$\begin{aligned} S^*(f_i) &= N \frac{c-1}{c^N-1} \sum_{k=N-S(f_i)}^{N-1} c^k = N \frac{c-1}{c^N-1} \left(\sum_{k=0}^{N-1} c^k - \sum_{k=0}^{N-S(f_i)-1} c^k \right) \\ &= N \frac{c-1}{c^N-1} \left(\frac{c^N-1}{c-1} - \frac{c^{N-S(f_i)}-1}{c-1} \right) = N \left(1 - \frac{c^N}{c^N-1} c^{-S(f_i)} \right) \end{aligned}$$

Since $s^*(f_i) = S^*(f_i) - S^*(f_{i-1})$, it then follows that

$$s^*(f_i) = N \frac{c^N}{c^N-1} \left(c^{-S(f_{i-1})} - c^{-S(f_i)} \right) = N \frac{c^N}{c^N-1} c^{-S(f_i)} \left(c^{S(f_i)} - 1 \right) \quad \square$$

EXAMPLE 6.1: Using the discrete fitness distribution from Example 2.1, the fitness distribution shown in Figure 11 is obtained after applying exponential ranking selection with $c = 0.986$ ($N = 250$).

THEOREM 6.2: Let $\tilde{s}(f)$ be the continuous fitness distribution of the population. Then the expected fitness distribution after performing exponential ranking selection $\bar{\Omega}_E$ with parameter c on the distribution \tilde{s} is

$$\bar{\Omega}_E^*(\tilde{s}, c)(f) = \tilde{s}^*(f) = N \frac{c^N}{c^N-1} \ln c \tilde{s}(f) c^{-\tilde{S}(f)} \quad (46)$$

PROOF: Since the continuous form of Equation 44 is given by

$$\tilde{p}(x) = \frac{c^{N-x}}{\int_0^N c^{N-x}} \quad \text{and} \quad \int c^x = \frac{1}{\ln c} c^x$$

one calculates:

$$\tilde{S}^*(f) = N \frac{c^N \ln c}{c^N - 1} \int_0^{\tilde{S}(f)} c^{-x} dx = -N \frac{c^N}{c^N - 1} [c^{-x}]_0^{\tilde{S}(f)} = -N \frac{c^N}{c^N - 1} (1 + c^{-S(f)})$$

Since $\tilde{s}^*(f) = d\tilde{S}^*(f)/df$, Equation 46 follows. \square

It is useful to introduce a new variable $\kappa = c^N$ to eliminate the explicit dependence on the population size N :

$$\bar{\Omega}_E^*(s, \kappa)(f) = \tilde{s}^*(f) = \frac{\kappa \ln \kappa}{\kappa - 1} \tilde{s}(f) \kappa^{-\tilde{S}(f)/N} \quad (47)$$

The meaning of κ will become apparent in the next section.

6.1 Reproduction Rate

COROLLARY 6.1: *The reproduction rate of exponential ranking selection is*

$$\bar{R}_E(f) = \frac{\kappa \ln \kappa}{\kappa - 1} \kappa^{-\tilde{S}(f)/N} \quad (48)$$

This is directly obtained by substituting Equation 47 in Equation 10.

Equation 48 shows that the worst individual has the lowest reproduction rate $\bar{R}(f_0) = \kappa \ln \kappa / (\kappa - 1)$, and the best individual has the highest reproduction rate $\bar{R}(f_n) = \ln \kappa / (\kappa - 1)$. Hence, a natural explanation of the variable κ is obtained, since $\bar{R}(f_0)/\bar{R}(f_n) = \kappa$: It describes the ratio of the reproduction rate of the worst and the best individual. Note that $c < 1$, and hence $c^N \ll 1$ for large N ; that is, the interesting region of values for κ is in the range $[10^{-20}, 1]$.

6.2 Loss of Diversity

THEOREM 6.3: *The loss of diversity $D_E(\kappa)$ of exponential ranking selection is*

$$D_E(\kappa) = \frac{1 - \ln \frac{\kappa - 1}{\kappa \ln \kappa}}{\ln \kappa} - \frac{\kappa}{\kappa - 1} \quad (49)$$

PROOF: From the demand $R(f_z) = 1$, one calculates

$$\frac{\tilde{S}(f_z)}{N} = -\frac{\ln \frac{\kappa - 1}{\kappa \ln \kappa}}{\ln \kappa} \quad (50)$$

Using Theorem 2.2 leads to

$$\begin{aligned} D_E(\kappa) &= \frac{1}{N} (\tilde{S}(f_z) - \tilde{S}^*(f_z)) \\ &= -\frac{\ln \frac{\kappa - 1}{\kappa \ln \kappa}}{\ln \kappa} - \frac{\kappa}{\kappa - 1} \left(1 - \kappa^{\frac{\ln |(\kappa - 1)/\kappa \ln \kappa|}{\ln \kappa}} \right) \\ &= -\frac{\ln \frac{\kappa - 1}{\kappa \ln \kappa}}{\ln \kappa} - \frac{\kappa}{\kappa - 1} \left(1 - \frac{\kappa - 1}{\kappa \ln \kappa} \right) = \frac{1 - \ln \frac{\kappa - 1}{\kappa \ln \kappa}}{\ln \kappa} - \frac{\kappa}{\kappa - 1} \end{aligned} \quad \square$$

The loss of diversity is shown in Figure 12.

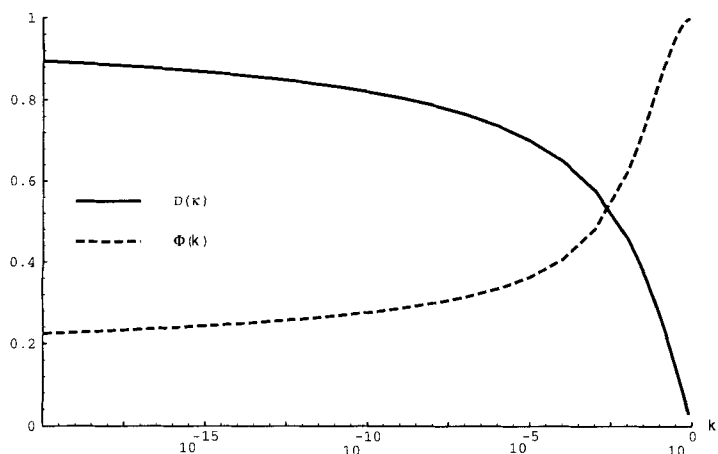


Figure 12. The loss of diversity $D_E(\kappa)$ (solid line) and selection variance $\Phi_E(\kappa)$ (dashed line) of exponential ranking selection. Note the logarithmic scale of the κ -axis.

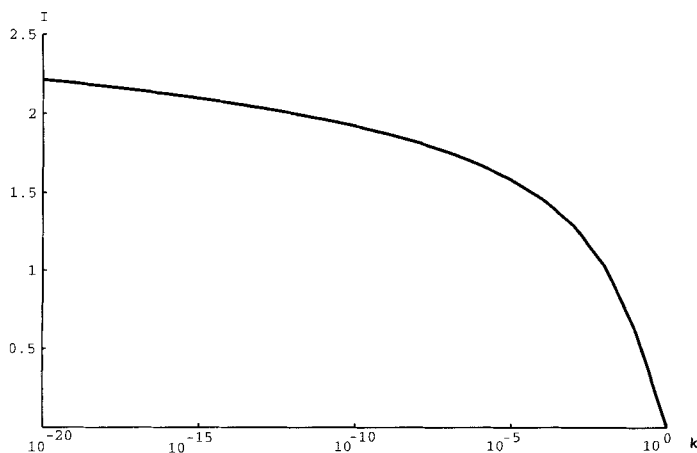


Figure 13. The selection intensity $I_E(\kappa)$ of exponential ranking selection. Note the logarithmic scale of the κ -axis.

6.3 Selection Intensity and Selection Variance

Analytical expressions of the selection intensity and selection variance are difficult to calculate for exponential ranking. Recalling the definition of the selection intensity (Definition 2.10), one sees that the integral of the Gaussian function occurs as the exponent in an indefinite integral. Hence, the selection intensity as well as the selection variance will only be numerically calculated. The selection intensity is shown in Figure 13 and the selection variance in Figure 12.

However, an approximation formula can be derived using the genetic programming optimization method (Blickle, 1996). The selection intensity of exponential ranking selection can be approximated with a relative error of less than 1.81% for $\kappa \in [10^{-20}, 0.8]$ by

$$I_E(\kappa) \approx \frac{\ln \kappa}{-2.548 - 1.086\sqrt{\kappa} + 0.4028 \ln \kappa} \tag{51}$$

Table 1. Comparison of the reproduction rate of the selection methods for discrete distributions.

Selection Method Ω	Reproduction Rate $R_{\Omega}(f_i)$
Tournament	$\frac{N}{s(f_i)} \left(\left(\frac{S(f_i)}{N} \right)^t - \left(\frac{S(f_{i-1})}{N} \right)^t \right)$
Truncation	$\begin{cases} 0 & : S(f_i) \leq (1 - T)N \\ \frac{S(f_i) - (1 - T)N}{s(f_i)T} & : S(f_{i-1}) \leq (1 - T)N < S(f_i) \\ \frac{1}{T} & : \text{else} \end{cases}$
Linear ranking	$\frac{N\eta^- - 1}{N - 1} + \frac{1 - \eta^-}{N - 1} (2S(f_i) - s(f_i))$
Exp. ranking	$\frac{N}{s(f_i)} \frac{\kappa}{\kappa - 1} \kappa^{-S(f_i)/N} (\kappa^{s(f_i)/N} - 1)$

Similarly, an approximation for the selection variance of exponential ranking selection can be found. The following formula approximates the selection variance with an relative error of less than 2.8% for $\kappa \in [10^{-20}, 0.8]$:

$$\Phi_E(\kappa) \approx 0.1 + \kappa^{1/4} (e^{-\kappa/2} + 0.2428\kappa^{-15/64})$$

(52)

The selection variance is shown in Figure 12.

7. Comparison of Selection Schemes

In this section the selection methods are compared according to their properties derived in the preceding sections. First, the reproduction rates of selection methods are compared, and a unified view of selection schemes is then derived. Section 7.3 is devoted to the comparison of the selection intensity and gives a convergence prediction for a simple genetic algorithm optimizing the OneMax function. The selection intensity is also used in the subsequent sections to compare the methods according to their loss of diversity and selection variance.

7.1 Reproduction Rate and Universal Selection

The reproduction rate gives the number of expected offspring of an individual with a certain fitness value after selection. In the preceding sections only the reproduction rate for the continuous case was considered. Table 1 gives the equations for the discrete (exact) case. They have been derived using the exact offspring equations 17, 27, 37, and 45 with some additional algebraic manipulations.

The standard sampling mechanism uses one spin of a roulette wheel (divided into segments for each individual, with the segment size proportional to the reproduction rate) to determine one member of the next generation. Hence, N trials have to be performed to obtain an entire population. Since these trials are independent of each other, a relatively high mean variation in the outcome is observed (see also Section 2 and Theorem 2.1). This is also the case for tournament selection, although there is no explicitly used roulette wheel sampling.

This mean variation can be almost completely eliminated by using the reproduction rate and the so-called “stochastic universal sampling” (SUS) method (Baker, 1987). In this method only a single spin of the wheel is necessary, since the roulette has N markers for the “winning individuals,” and, hence, all individuals are chosen at once.

The SUS algorithm can be regarded as an optimal sampling algorithm. It has zero bias (i.e., no deviation between the expected reproduction rate and the algorithmic sampling frequency). Furthermore, SUS has minimal spread; that is, the range of the possible values for $s'(f_i)$ is

$$s'(f_i) \in \{\lfloor s^*(f_i) \rfloor, \lceil s^*(f_i) \rceil\} \quad (53)$$

The outline of the SUS algorithm is given by Algorithm 5.

ALGORITHM 5 (Stochastic Universal Sampling):

Input: The population \vec{f} and the reproduction rate for each fitness value in $\vec{R} = \{R_1, \dots, R_n\}$

Output: The population after selection \vec{f}'

SUS(\vec{f}, \vec{R}):

$sum \leftarrow 0$

$j \leftarrow 1$

$ptr \leftarrow \text{uniform_random}[0,1)$

for $i \leftarrow 1$ **to** N **do**

$sum \leftarrow sum + R_k$ where R_k is the reproduction rate of individual f_i

while ($sum > ptr$) **do**

$f'_j \leftarrow f_i$

$j \leftarrow j + 1$

$ptr \leftarrow ptr + 1$

od

od

return \vec{f}'

By means of the SUS algorithm, the outcome of a certain run of the selection scheme is as close as possible to the expected behavior; that is, the mean variation is minimal. Even though it is not clear whether there are any performance advantages in using SUS, it makes the run of a selection method more “predictable.”

To be able to apply SUS one has to know the expected number of offspring of each individual. Baker (1987) has applied this sampling method only to linear ranking selection, since here the expected number of offspring is known by construction (see Section 5). Because the offspring values have been derived for the selection methods discussed in the previous sections (see Table 1), it is possible to use stochastic universal sampling for all these selection schemes. This leads to a unified view of selection schemes and allows the construction of a “universal selection method” in the following way: First, the fitness distribution of the population is computed. Next, the expected reproduction rates are calculated using the equations derived in the preceding sections (Table 1). In the last step, the SUS algorithm is used to obtain the new population after selection (Algorithm 5). The time complexity of the universal selection method is $\mathcal{O}(N \ln N)$, because the fitness distribution has to be computed. Hence, if “tournament selection” is performed with this algorithm, the lower mean variation is paid for with a higher computational complexity.

7.2 The Complement Selection Schemes: Tournament and Linear Ranking

If several properties of tournament selection and linear ranking selection are compared, one observes that binary tournament behavior is similar to that of a linear ranking selection with a very small η^- ; and it is indeed possible to prove that binary tournament and linear ranking with $\eta^- = 1/N$ have identical expected behavior. By this the complementary character of the two selection schemes is revealed.

ALGORITHM 6 (Universal Selection Method):

Input: The population $\vec{f} = \{f_1, \dots, f_N\}$,
the fitness values of the population $\vec{o} = \{o_1, \dots, o_N\}$.

Output: The population after selection \vec{f}'

```
universal_selection( $\vec{f}, \vec{o}$ ):
     $\vec{s} \leftarrow \text{fitness\_distribution}(\vec{o})$ 
     $\vec{r} \leftarrow \text{reproduction\_rate}(\vec{s})$ 
     $\vec{f}' \leftarrow \text{SUS}(\vec{f}, \vec{r})$ 
    return  $\vec{f}'$ 
```

THEOREM 7.1: *The expected fitness distributions of linear ranking selection with $\eta^- = 1/N$ and tournament selection with $t = 2$ are identical, that is,*

$$\Omega_R^* \left(s, \frac{1}{N} \right) = \Omega_T^*(s, 2) \quad (54)$$

PROOF:

$$\begin{aligned} \Omega_R^* \left(s, \frac{1}{N} \right) (f_i) &= s(f_i) \frac{N \frac{1}{N} - 1}{N - 1} + \frac{1 - \frac{1}{N}}{N - 1} (S(f_i)^2 - S(f_{i-1})^2) \\ &= \frac{1}{N} (S(f_i)^2 - S(f_{i-1})^2) \\ &= \Omega_T^*(s, 2)(f_i) \end{aligned} \quad \square$$

Goldberg and Deb (1991) have also shown this result, but only for the behavior of the best fit individual.

7.3 Comparison of the Selection Intensity

The selection intensity is a very important property of the selection method; Table 3 gives some settings for the selection methods that yield the same selection intensity.

The importance of the selection intensity is based on the fact that the behavior of a simple genetic algorithm can be predicted if the fitness distribution is normally distributed. In Mühlenbein and Schlierkamp-Voosen (1993), a prediction is made for a genetic algorithm optimizing the OneMax (or bit-counting) function. Here, the fitness is given by the number of 1's in the binary string of length l . Uniform crossover is used and assumed to be a random process that creates a binomial fitness distribution. As a result, after each recombination phase the input of the next selection phase approximates a Gaussian distribution. Hence, a prediction of this optimization using the selection intensity should be possible. For a

Table 2. Comparison of the selection intensity of the selection methods.

Selection Method	Selection Intensity
Tournament	$I_T(t) \approx \sqrt{2(\ln t - \ln(\sqrt{4.14 \ln t}))}$
Truncation	$I_\Gamma(T) = \frac{1}{T} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{f_c^2}{2}\right)$
Linear ranking	$I_R(\eta^-) = (1 - \eta^-) \frac{1}{\sqrt{\pi}}$
Exponential ranking	$I_E(\kappa) \approx \frac{\ln \kappa}{-2.548 - 1.086\sqrt{\kappa} + 0.4028 \ln \kappa}$

Table 3. Parameter settings for truncation selection Ω_Γ , tournament selection Ω_T , linear ranking selection Ω_R , and exponential ranking selection Ω_E to achieve the same selection intensity I .

I	0.34	0.56	0.84	1.03	1.16
$\Omega_T : t$	—	2	3	4	5
$\Omega_R : \eta^-$	0.4	0	—	—	—
$\Omega_\Gamma : T$	0.8	0.66	0.47	0.36	0.30
$\Omega_E : \kappa$	0.29	0.12	0.032	$9.8 \cdot 10^{-3}$	$3.5 \cdot 10^{-3}$
I	1.35	1.54	1.87	2.16	
$\Omega_T : t$	7	10	20	40	
$\Omega_\Gamma : T$	0.22	0.15	0.08	0.04	
$\Omega_E : \kappa$	$4.7 \cdot 10^{-4}$	$2.5 \cdot 10^{-5}$	10^{-9}	$2.4 \cdot 10^{-18}$	

sufficiently large population, the authors calculate

$$p(k) = \frac{1}{2} \left(1 + \sin \left(\frac{I}{\sqrt{l}} k + \arcsin(2p_0 - 1) \right) \right)$$

(55)

where p_0 denotes the fraction of 1's in the initial random population and $p(k)$ the fraction of 1's in generation k . Convergence is characterized by the fact that $p(k_c) = 1$, so the convergence time for the special case of $p_0 = 0.5$ is given by

$$k_c = \frac{\pi}{2} \frac{\sqrt{n}}{I}$$

It is straightforward to give the convergence time for any other selection method, by substituting I with the corresponding terms derived in the preceding sections.

One calculates

$$k_{T,c}(t) \approx \frac{\pi}{2} \sqrt{\frac{l}{2(\ln t - \ln \sqrt{4.14 \ln t})}}$$

(56)

for tournament selection;

$$k_{\Gamma,c}(T) = T \frac{\pi \sqrt{\pi l}}{\sqrt{2}} \exp\left(-\frac{f_c^2}{2}\right)$$

(57)

Table 4. Comparison of the loss of diversity of the selection methods.

Selection Method	Loss of Diversity
Tournament	$D_T(t) = t^{-1/(t-1)} - t^{-t/(t-1)}$
Truncation	$D_T(T) = 1 - T$
Linear ranking	$D_R(\eta^-) = (1 - \eta^-)^{\frac{1}{4}}$
Exponential ranking	$D_E(\kappa) = \frac{1 - \ln \frac{\kappa - 1}{\kappa \ln \kappa}}{\ln \kappa} - \frac{\kappa}{\kappa - 1}$

for truncation selection;

$$k_{R,\kappa}(\eta^-) = \frac{\pi \sqrt{\pi l}}{2(1 - \eta^-)} \tag{58}$$

for linear ranking selection; and

$$k_{E,\kappa}(\kappa) \approx \frac{\pi \sqrt{l}}{2} \frac{-2.548 - 1.086\sqrt{\kappa} + 0.4028 \ln \kappa}{\ln \kappa} \tag{59}$$

for exponential ranking selection.

7.4 Comparison of Loss of Diversity

Table 4 summarizes the loss of diversity for the selection methods. It is difficult to compare these relations directly because they depend on different parameters that are characteristic of the specific selection method (e.g., the tournament size t for tournament selection, the threshold T for truncation selection). Hence, one has to look for an independent measure to eliminate these parameters and to be able to compare the loss of diversity. This measure is chosen to be the selection intensity. The loss of diversity of the selection methods is viewed as a function of the selection intensity. To calculate the corresponding graph, one first computes the value of the parameter of a selection method (i.e., t for tournament selection, T for truncation selection, η^- for linear ranking selection, and κ for exponential ranking selection) that is necessary to achieve a certain selection intensity. With this value, the loss of diversity is then obtained using the corresponding equations (i.e., Equations 22, 30, 40, and 49). Figure 14 shows the result of this comparison: the loss of diversity for the different selection schemes in relation to the dependence of the selection intensity. Tournament and exponential ranking behave almost identically. Ranking selection only covers the lower quarter of the I -axis, since the maximum selection intensity of ranking selection is $I_R(0) = 1/\sqrt{\pi} \approx 0.56$. In this region its behavior is also very similar to that of exponential ranking. To achieve the same selection intensity, more bad individuals are replaced using truncation selection than using tournament selection or one of the ranking selection schemes, respectively. This means that more “genetic material” is lost using truncation selection.

The graph in Figure 14 is based on the standardized selection intensity as given by Definition 2.9; that is, a normal distribution of the initial fitness values was assumed. Furthermore, continuous distributions have been used to derive the above results. To estimate their generality and the generality of the continuous calculations, the following experiment was set up: 800 different fitness distributions were randomly created and used as input to tournament selection and truncation selection. Eight different parameter settings for the selection methods were used (corresponding to the first eight entries in Table 3). The *realized*

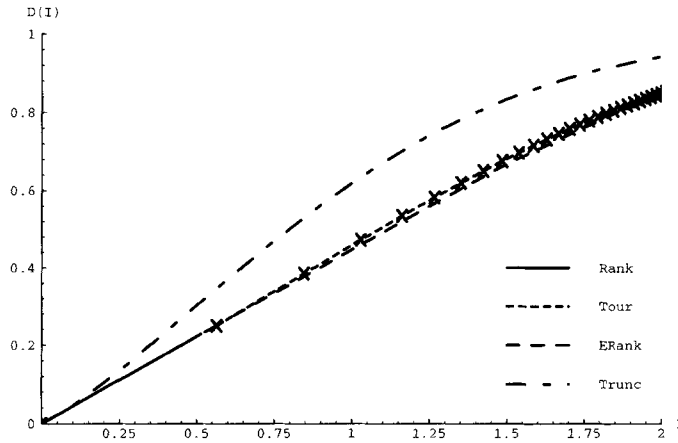


Figure 14. Dependence of the loss of diversity D on selection intensity I for tournament selection (Tour), truncation selection (Trunc), linear ranking selection (Rank), and exponential ranking selection (ERank). Note that for tournament selection only, the points marked as crosses on the graph correspond to valid (integer) tournament sizes.

selection intensity \hat{I} and the loss of diversity were measured for each fitness distribution. The realized selection intensity was computed according to the general definition of selection intensity (Definition 2.9), that is,

$$\hat{I} = \frac{M^* - M}{\sqrt{\sum_{i=1}^n s(f_i)(f_i - M)^2}} \quad (60)$$

The difference between the loss of diversity of both methods for a given selection intensity is shown in Figure 15. The solid line is the theoretical expected difference according to the standardized selection intensity (as in Fig. 14). The dots are the measured differences of the loss of diversity. As can be seen, in almost all cases this difference is positive (i.e., truncation selection has a higher loss of diversity than tournament selection *for arbitrary fitness distributions*). Furthermore, the computed loss of diversity seems to be an accurate approximation and upper bound for this difference.

Suppose that a lower loss of diversity is desirable because it reduces the risk of premature convergence; one then expects that truncation selection should be outperformed by the other selection methods. In general, whether a low loss of diversity is “advantageous” depends on the problem and the representation of the problem to be solved. With Figure 14, however, one has a useful tool at hand for making the right decision with regard to a particular problem.

Another interesting fact can be observed by looking at Table 4: The loss of diversity is independent of the initial fitness distribution. Nowhere in the derivation of these equations is a particular fitness distribution assumed, and nowhere does the fitness distribution $\bar{s}(f)$ occur in the equations. In contrast, the (standardized) selection intensity and the (standardized) selection variance are computed for a certain initial fitness distribution (the normalized Gaussian distribution). Hence, the loss of diversity can be viewed as an inherent property of a selection method.

7.5 Comparison of the Selection Variance

The mechanism used in the preceding subsection is now applied to the selection variance; that is, the selection variance is viewed as a function of the selection intensity.

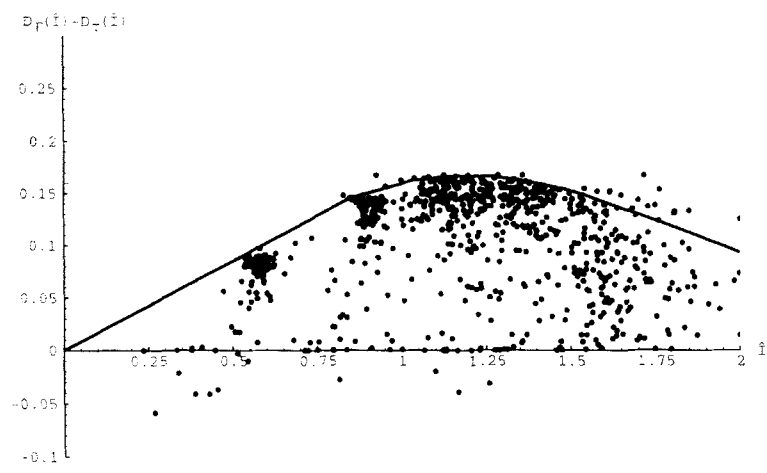


Figure 15. The measured difference of the loss of diversity of truncation selection and tournament selection as a function of the realized selection intensity. The solid line gives the difference according to the standardized selection intensity (as displayed in Fig. 14). The points give the differences measured for 800 randomly created fitness distributions.

Table 5. Comparison of the selection variance of the selection methods.

Selection Method	Selection Variance
Tournament	$\Phi_T(t) \approx \frac{0.918}{\ln(1.186 + 1.328t)}$
Truncation	$\Phi_T(T) = 1 - I_T(T)(I_T(T) - f_c)$
Linear ranking	$\Phi_R(\eta^-) = 1 - I_R^2(\eta^-)$
Exponential ranking	$\Phi_E(\kappa) \approx 0.1 + \kappa^{1/4}(e^{-\kappa/2} + 0.2428\kappa^{-15/64})$

It can be seen in Figure 16 that truncation selection leads to a lower selection variance than tournament selection. The highest selection variance is obtained by exponential ranking.

An interpretation of the results is difficult because whether a high selection variance is advantageous depends on the optimization task and the type of problem to be solved. But, again, Figure 16 may help one decide on the “appropriate” selection method for a particular optimization problem. Under the assumption that a higher variance is advantageous to the optimization process, exponential ranking selection reveals itself to be the best selection scheme.

8. Conclusions

In this paper some common selection schemes were analyzed. The analysis was based on fitness distributions and examined only the local behavior of the selection; that is, the fitness distributions after selection were compared with the fitness distributions before selection. Although this idea is not new, the consequent realization of this idea led to a powerful framework that allowed the following results to be obtained:

- Fitness distributions allow an unified view of the selection schemes and enable several up to now independently and isolated obtained aspects of selection schemes to be derived with one single methodology.

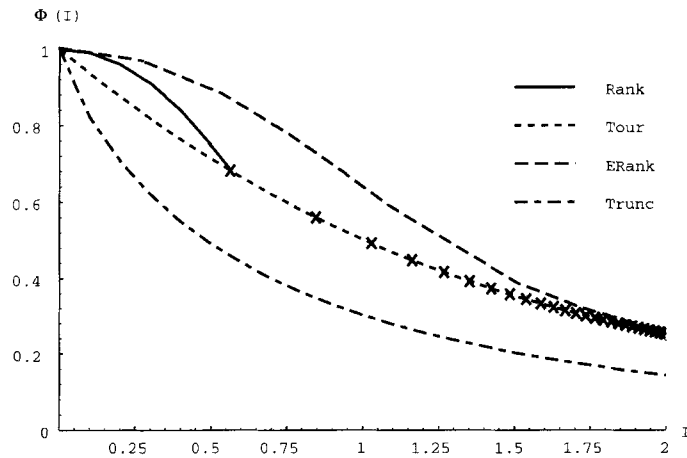


Figure 16. The dependence of the selection variance Φ on the selection intensity I for tournament selection (Tour), truncation selection (Trunc), ranking selection (Rank), and exponential ranking selection (ERank). Note that for tournament selection only the points marked as crosses on the graph correspond to valid (integer) tournament sizes.

- Some interesting features of selection schemes could be proved, for example, the concatenation of several tournament selections (Theorem 3.3) and the equivalence of binary tournament and linear ranking (Theorem 7.1).
- Important characteristics of selection schemes, such as selection intensity, selection variance, and loss of diversity were derived.
- Selection intensity was used to obtain a convergence prediction of the simple genetic algorithm with uniform crossover optimizing the OneMax function.
- Comparison of the loss of diversity and the selection variance based on the selection intensity allowed for the first time “second-order” properties of selection schemes to be compared. This gives a well-grounded basis for deciding which selection scheme is applicable, assuming that for a particular problem the impact of these properties on the optimization process is known.

The analysis in this paper gave a detailed insight into the working mechanism of selection schemes. Although the comparison of the loss of diversity and selection variance is helpful, for a complete understanding of the local behavior of an evolutionary algorithm, the variation operator (crossover and mutation) also has to be considered.

References

- Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A first course in order statistics*. New York: Wiley.
- Bäck, T. (1994a). *Evolutionary algorithms in theory and practice*. Unpublished doctoral dissertation, Fachbereich Informatik, Universität Dortmund, Germany.
- Bäck, T. (1994b). Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. In *Proceedings of the First IEEE Conference on Evolutionary Computation* (pp. 57–62). Piscataway, NJ: IEEE Press.
- Bäck, T. (1995). Generalized convergence models for tournament- and (μ, λ) -selection. In L. Eshelman

- (Ed.), *Proceedings of the Sixth International Conference on Genetic Algorithms* (pp. 2–8). San Francisco, CA: Morgan Kaufmann.
- Baker, J. E. (1987). Reducing bias and inefficiency in the selection algorithm. In J. Grefenstette (Ed.), *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms and Their Applications* (pp. 14–21), Hillsdale, NJ: Lawrence Erlbaum.
- Baker, J. E. (1989). *An analysis of the effects of selection in genetic algorithms*. Unpublished doctoral dissertation, Graduate School of Vanderbilt University, Nashville, Tennessee.
- Blickle, T. (1996). *Theory of evolutionary algorithms and application to system-synthesis* (Doctoral dissertation, Swiss Federal Institute of Technology [ETH] Zurich, Switzerland, 1996, ETH dissertation no. 11894). Zurich: VDF-Verlag Zurich.
- Blickle, T., & Thiele, L. (1995). A mathematical analysis of tournament selection. In L. Eshelman (Ed.), *Proceedings of the Sixth International Conference on Genetic Algorithms* (pp. 9–16). San Francisco, CA: Morgan Kaufmann.
- Bulmer, M. G. (1980). *The mathematical theory of quantitative genetics*. Oxford: Clarendon Press.
- Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. New York: Harper and Row.
- De la Maza, M., & Tidor, B. (1993). An analysis of selection procedures with particular attention paid to proportional and Boltzmann selection. In S. Forrest (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 124–131). San Mateo, CA: Morgan Kaufmann.
- Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. In G. Rawlins (Ed.), *Foundations of genetic algorithms* (pp. 69–93). San Mateo, CA: Morgan Kaufmann.
- Grefenstette, J. J., & Baker, J. E. (1989). How genetic algorithms work: A critical look at implicit parallelism. In J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 20–27). San Mateo, CA: Morgan Kaufmann.
- Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.
- Miller, B. L., & Goldberg, D. E. (1995). *Genetic algorithms, tournament selection, and the effects of noise* (Tech. Rep. No. 95006). Urbana, IL: University of Illinois, The Illinois Genetic Algorithm Laboratory.
- Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm. *Evolutionary Computation*, 1(1), 25–49.
- Nagaraja, H. (1982). Selection differentials. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical science* (pp. 334–336). New York: Wiley.
- Shapiro, J., Prügel-Bennett, A., & Rattray, M. (1994). A statistical mechanical formulation of the dynamics of genetic algorithms. In T. C. Fogarty (Ed.), *Evolutionary computing AISB workshop* (pp. 17–27). Berlin: Springer-Verlag.
- Thierens, D., & Goldberg, D. (1994a). Convergence models of genetic algorithm selection schemes. In Y. Davidor, H.-P. Schwefel, & R. Männer (Eds.), *Parallel Problem Solving from Nature—PPSN III* (pp. 119–129). Berlin: Springer-Verlag.
- Thierens, D., & Goldberg, D. (1994b). Elitist recombination: an integrated selection recombination GA. In *Proceedings of the First IEEE Conference on Evolutionary Computation* (pp. 508–512). Piscataway, NJ: IEEE Press.
- Whitley, D. (1989). The GENITOR algorithm and selection pressure: Why rank-based allocation of reproductive trials is best. In J. D. Schaffer (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 116–121). San Mateo, CA: Morgan Kaufmann.